

Planning, Shortage, and Transformation

edited by **Eric S. Maskin**

András Simonovits

Essays in Honor of János Kornai



Planning, Shortage, and Transformation

Essays in Honor of János Kornai

edited by Eric S. Maskin and András Simonovits

The three major themes of János Kornai's work reflected in the title of this book—planning, shortage, and transition, or transformation—figure prominently in the essays. After a philosophical introduction by Edmond Malinvaud, the book is divided into three sections: Markets and Organizations, Theory of Transition, and The Transitional Experience.

János Kornai is one of the world's leading experts on the economics of socialism and transition. An early advocate of reform in Hungary, he has written penetrating analyses of centrally planned economies and their transformation to market-oriented systems. This volume is being published in celebration of Kornai's seventieth birthday.

Planning, Shortage, and Transformation

Planning, Shortage, and Transformation

Essays in Honor of János Kornai

edited by

Eric Maskin and András Simonovits

The MIT Press

Cambridge, Massachusetts

London, England

©2000 Massachusetts Institute of Technology

All rights reserved. No part of this book may be reproduced in any form by any electronic or mechanical means (including photocopying, recording, or information storage and retrieval) without permission in writing from the publisher.

This book was set in Times Roman by Best-set Typesetter Ltd., Hong Kong.
Printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Planning, shortage, and transformation: essays in honor of János

Kornai / edited by Eric Maskin and András Simonovits.

p. cm.

Includes bibliographical references and index.

ISBN 0-262-13357-1 (hard: alk. paper)

1. Economic policy. 2. Scarcity. 3. Post-communism. 4. Europe.
Eastern—Economic policy—1989- . 5. Kornai, János Biography.

I. Kornai, János. II. Maskin, Eric. III. Simonovits, András.

IV. Title: Essays in honor of János Kornai.

HD87.P57 2000

338.947—dc21

99-37733

CIP

Contents

Preface	vii
Contributors	xiii
A Short Biography of János Kornai	xv
Introduction: Some Notes on Assessments about Economic Systems	1
Edmond Malinvaud	
I MARKETS AND ORGANIZATIONS	15
1 Monopoly under Labor Management	17
Abram Bergson	
2 Walras-Keynes Equilibria Coordination and Macroeconomics	25
Jacques H. Drèze	
3 Viable Domains in the Control Space	47
Béla Martos	
4 Coordinating Activities under Alternative Organizational Forms	57
Yingyi Qian, Gérard Roland, and Chenggang Xu	
5 Democracy, Egalitarian Redistributions, and the Level of Public Bads	81
John E. Roemer	
6 Parable and Realism in Overlapping Cohorts Models	95
András Simonovits	
7 Internal Efficiency and External Conditions	109
Jörgen W. Weibull	
8 On Buyers' and Sellers' Markets under Capitalism and Socialism	127
Martin Weitzman	
II THEORY OF TRANSITION	141
9 Soft Budget Constraints and Transition	143
Mathias Dewatripont, Eric Maskin, and Gérard Roland	
10 Restructuring an Industry during Transition: A Two-Period Model	157
Richard E. Ericson	
11 Shortage and the Soft Budget Constraint: New Interpretations	189
Paul Hare	

12	The Kornai Effect Revisited	209
	Richard E. Quandt	
III	THE TRANSNATIONAL EXPERIENCE	231
13	Determinants of Productivity in Central Europe during Economic Transition	233
	Sharad Bhandari and David M. Kemme	
14	The Evolutionary Path away from Socialism: The Chinese Experience	249
	Bernard Chavance	
15	Perestroika Economics from the Inside	269
	Michael Ellman	
16	The Internal versus Occupational Labor Market—A Neglected Dimension of Hungary's Postsocialist Transformation	283
	István R. Gábor	
17	Transformation before the Transition: Employment and Wage Setting in Hungarian Firms, 1986–1989	307
	János Köllő	
18	How Does the Transitional State Behave? Evidence from Retained State Ownership in Mongolian Privatization	327
	Georges Korsun and Peter Murrell	
19	Do Power Consumption Data Tell the Story? Electricity Intensity and Hidden Economy in Postsocialist Countries	345
	Mária Lackó	
20	Transition to the Market in Asia: The Case of Vietnam	367
	Marie Lavigne	
21	The Debate on Understanding China's Economic Performance	385
	Jeffrey D. Sachs and Wing Thye Woo	
22	Stability and Growth: Commentary on a Commentary	407
	Robert Solow	
23	The Surprising Success of Gradual Price Liberalization in Hungary	413
	Wim Swaan	
	The Publications of János Kornai	435

Preface

Thirty scholars from all over the world have contributed to this volume in celebration of János Kornai's seventieth birthday. The book's title, *Planning, Shortage, and Transformation*, reflects three of the major themes in János's work. And, indeed, all three concepts figure prominently in the contributions assembled here.

The volume is divided into three parts, with a philosophical introduction, "Some Notes on Assessments about Economic Systems," by Edmond Malinvaud. Malinvaud takes up the central "Kornai question": How do alternative socioeconomic systems work? More specifically, he offers an assessment of how well different systems have performed historically, a discussion of the role of planning in market economies, and his forecast for the future of the welfare state. Most of his conclusions should be agreeable to Kornai. The two concur that Soviet-style economies were basically unworkable and unreformable. They also share the view—in fact, they were among the first to emphasize—that a major role of economic planning is to disseminate information. They seem to part company, however, on the welfare state: Malinvaud's verdict on past performance and prognosis for the future seem rather rosier than we suspect Kornai would feel comfortable with.

Part I of the volume comprises eight chapters on markets and organizations, beginning with Abram Bergson's contribution "Monopoly under Labor Management." The labor-managed enterprise has piqued the interest of a good many economic theorists. According to its proponents, such an institution can skirt soft-budget constraints—the investigation of whose harmful effects was pioneered by János Kornai—while improving on capitalism's poor employment record. Bergson demonstrates, however, that, at least in the case of monopoly, a labor-managed firm is likely to employ *less* labor (and therefore produce lower output) than its capitalist counterpart. This is a proposition dating back to Benjamin Ward, but Bergson shows that it holds more generally than was previously thought.

Since the early 1970s, Jacques Drèze has been a pioneer of "disequilibrium" theory, which investigates the nature of market-clearing when prices are rigid, and which complements the literature inspired by Kornai on the theory of shortage. In his contribution "Coordination Failures in Walras-Keynes Macroeconomics," Drèze provides a disequilibrium analysis of unemployment. Specifically, he shows that equilibria in which suppliers of labor are constrained may arise through a combination of coordination failures and inflexible prices.

Béla Martos's contribution, "Viable Domains in the Control Space" is a step toward fulfilling an objective enunciated in Kornai's 1971 book *Anti-Equilibrium*, that is, that neoclassical economics, which is predominantly static, should be replaced by a dynamic theory through the application of control theoretic methods. Martos's essay provides sufficient conditions for a control problem solution to be *viable* in

the sense that it belongs to specified subsets of the state and control spaces. The requirement of viability has clear pertinence to many economic problems, which typically impose some such constraint, for example, that consumption bundles be nonnegative or that production plans be feasible.

The essay by Yingyi Qian, Gérard Roland, and Chenggang Xu, "Coordinating Activities under Alternative Organizational Forms," centers on an issue at the forefront of János Kornai's research agenda: how economic activities are arranged within organizations when they are not determined by markets. Qian, Roland, and Xu introduce the concept of "attribute matching" as a way of modeling coordination within organizations. By doing so, they are able to derive some interesting comparisons between different organizational structures, in particular between the so-called M-form and U-form structures.

Kornai has understandably been intensely interested in the economic consequences of political structure. "Democracy, Egalitarian Redistributions, and the Level of Public Bads" is a contribution to John Roemer's ongoing research on how political competition affects the allocation of goods and bads. In particular, he studies a model of two political parties—one of the poor, one of the rich—in which their electoral competition determines the equilibrium levels of income redistribution and pollution. The principal conclusion is a demonstration that the level of pollution is apt to rise with the degree of egalitarian redistribution.

In his "Parable and Realism in Overlapping Cohorts Models," András Simonovits evaluates the assumptions underlying the popular overlapping-generations model. In doing so, he joins Tjalling Koopmans, Paul Samuelson, and János Kornai to challenge Milton Friedman's famous claim that the descriptive realism of an economic model is less important than its predictive power. More specifically, he argues that standard theorems on the existence, optimality, and stability of steady states with overlapping generations are lost once one modestly enriches the model to make it more realistic.

Jörgen Weibull's essay "Internal Efficiency and External Conditions" investigates another topic of long-standing interest to Kornai: how a firm's external environment affects its efficiency. Weibull supposes that a firm is managed by its owner, who can exert costly effort in order to reduce the firm's costs. He is interested in the question of how the integration of two isolated markets (or, equivalently, the liberalization of trade between those markets) affects managers' effort choices. He provides plausible conditions under which such integration enhances effort and hence efficiency.

"On Buyers' and Sellers' Markets under Capitalism and Socialism," Martin Weitzman's contribution, tackles a classic Kornai subject: explaining why in

capitalist economies sellers are typically on the long side of the market (i.e., they cannot sell all they would wish to at the going prices), whereas it is *consumers* who are constrained in socialist economies (in that they cannot buy all that they want). Weitzman argues that pervasive monopolistic competition under capitalism is responsible for the sellers' constraints (their demand curves are strictly downward sloping). By contrast, the long queues of buyers under socialism can be traced to firm-specific limitations on inputs.

Part II of the volume consists of four chapters on the theory of transition. The first is "Soft Budget Constraints and Transition" by Mathias Dewatripont, Eric Maskin, and Gérard Roland. János Kornai was responsible for relating the soft budget constraint—the term he coined for persistent refinancing of loss-making enterprises—to the recurrent shortages experienced by socialist economies. The Dewatripont-Maskin-Roland essay is a survey of the extensive literature that has grown out of this seminal contribution and of its relevance to the transition from socialism to capitalism. A major point of emphasis is that the hardness or softness of budget constraints is not a direct policy variable but rather the result of institutional design. Thus, it is not enough for those overseeing transition to call for tougher financial discipline; to be credible, they must be backed up by an appropriate set of institutions.

In his chapter "Restructuring an Industry during Transition: A Two-Period Model," Richard Ericson provides a formal framework for the study of what Kornai has called the "transformation of the real structure of the economy" during the transition process. He studies the question of which state-owned firms to shut down in an environment where ultimate profitability is uncertain. Indeed, he obtains a complete characterization of the optimal shut-down policy contingent on the parameters of the model.

Kornai has attributed the recurrent shortages experienced by socialist economics to the soft budget constraint (SBC) syndrome (discussed by Dewatripont, Maskin, and Roland in this volume). This theory, although accepted by many, has attracted some controversy. In particular, Richard Portes and others have laid the blame for shortages largely on incorrect pricing. In "Shortage and the Soft Budget Constraint: New Interpretations," Paul Hare revisits the debate in the light of postsocialist experience. In his view, the rapid dissipation of shortages in the economies of the former Soviet Union once prices were liberalized calls Kornai's explanation into question, particularly because no concomitant elimination of SBCs occurred. Nevertheless, he acknowledges that the rapid inflation that ensued after liberalization is powerful evidence of the importance of the SBC syndrome that Kornai was the first to identify.

Richard Quandt also turns his attention to the SBC syndrome. He reiterates Kornai's observation that if a firm is bailed whenever its profit turns negative, its demand for inputs will be inefficiently high—a phenomenon Quandt dubs the *Kornai effect*. This he formally models in "The Kornai Effect Revisited." The model enables him to perform a number of interesting comparative statics exercises by varying the degree to which bailouts are complete. It also allows him to explore what happens when the firm's manager has the opportunity to lobby the bailout agency.

The eleven chapters of part III chronicle the transitional experience. Since the fall of communism in Eastern Europe and the former Soviet Union, only two economically viable socialist economies remain: China and Vietnam (Cuba and North Korea are not viable, whereas Cambodia and Laos can be neglected). Moreover, as three chapters in this volume make clear, these last two are already well down the road toward capitalism in certain respects.

In their essay "Determinants of Productivity in Central Europe during Economic Transition," Sharad Bhandari and David Kemme study the causes of productivity differences in small- and medium-sized manufacturing firms in the Czech and Slovak Federal Republic (which ceased to exist in 1993), Hungary, and Poland. Factors that appear to enhance significantly productivity include managers' level of education, the extent to which firms export, and the newness of their equipment.

After a philosophical introduction, Bernard Chavance's essay, "The Evolutionary Path away from Socialism: The Chinese Experience," reviews post-1978 developments in China in detail. Like Kornai, Chavance characterizes socialism as the combination of one-party rule and state ownership. Accordingly, as he notes, the first step in dismantling socialism can be either the introduction of democracy—as in Eastern Europe—or the attenuation of economic control by the state, as in China. It is quite possible, however, that either path has the same ultimate destination: full-fledged capitalism.

In "Perestroika Economics from the Inside," Michael Ellman scrutinizes the published memoirs of Soviet leaders—Gorbachev especially—to explain the denouement of the Soviet Union. He is of the opinion—a view shared by Kornai—that the attempts at economic reform actually led only to further disintegration.

Socialism has devastating flaws but at least it ensured full employment. No wonder that the precipitous rise in unemployment in several formerly socialist countries has attracted so much attention. István Gábor's study "Internal versus Occupational Labor Market—A Neglected Dimension of Hungary's Postsocialist Transformation" sheds light on this problem. Under socialism, large enterprises with internal labor markets dominated the economy. Once the socialist system collapsed,

occupational labor markets took over, but the institutional infrastructure needed to assure market-clearing failed to evolve as quickly.

János Köllő's "Transformation before the Transition: Employment and Wage-Setting in Hungarian Firms, 1986–1989" addresses the question of why Hungarian state enterprises reduced employment drastically in the last years of state socialism. Köllő argues that job destruction was the result of changing incentives and greater freedom at the firm level (e.g., harder budget constraints and fewer political constraints on downsizing) rather than of market-related shocks. This hypothesis is tested for a panel of nearly 2,700 firms.

Georges Korsun and Peter Murrell offer a wealth of empirical detail with a theoretical interpretation in "How Does the Transitional State Behave? Evidence from Retained State Ownership in Mongolian Privatization." One of their most intriguing findings is that in the most vulnerable enterprises in Mongolia have tended to have the largest residual shares of their markets.

Kornai has always emphasized the importance of the "underground" economy in the Hungarian reform process. And the share of national income attributable to underground transactions has clearly been rising during the transition. But getting an accurate measurement of the magnitudes is formidably difficult. In her chapter "Do Power Consumption Data Tell the Story?" Mária Lackó criticizes several popular measurement techniques for determining the size of the underground economy in postsocialist countries. She goes on to propose a new and promising methodology.

Vietnam has received relatively little attention from economists. With her study "Transition to the Market in Asia: The Case of Vietnam," Marie Lavigne is filling an important gap. In doing so, she follows the same general methodology laid out by Kornai in his 1992 book *The Socialist System*, and, in particular, she eschews single-factor explanations of any kind. One of the interesting points brought out by the chapter is how similar Vietnam's reform process has been to that of China.

In their essay "The Debate on Understanding China's Economic Performance," Jeffrey Sachs and Wing Thye Woo criticize the "experimentalist school" of economic transition and enunciate four hypotheses about the Chinese experience deriving from their own "convergence school" perspective: (1) China's rapid growth has occurred in spite of, not because of, gradualist reform; (2) China's gradualist strategy is not transferable to Eastern Europe because of important differences in economic structure; (3) China's experiments in noncapitalistic institutions are proving to be unsuccessful; and (4) China is gradually harmonizing its economic institutions with those of other East Asian market economies.

Before 1990 Kornai refrained from giving policy advice to the Hungarian government. Since then, however, he has taken on many public duties and published several policy-oriented articles in newspapers. Robert Solow's "Stability and Growth: Commentary on a Commentary" deals with one of these articles, "Lasting Growth is the Top Priority: Macroeconomic Tensions and Government Economic Policy in Hungary." At the time (1994), Hungary was suffering from a staggering foreign trade deficit, and there was a growing consensus that something serious had to be done about it. Kornai argued that it would be a mistake to curtail growth for the sake of restoring the foreign trade balance. Solow largely applauds this analysis and suggests that the sort of prescriptions that Kornai laid out should apply much more broadly than just to the Hungarian economy of 1994.

In "Price Liberalization in Hungary," Wim Swaan continues his ongoing analysis of the Hungarian price system. The essay is impressive in its detail, particularly of how prices were actually liberalized at the micro level. Interestingly, Swaan notes, the gradualist approach did not lead to intertemporal speculation, a danger often stressed by advocates of "shock therapy."

We have hardly done justice to the remarkably varied and deep contributions in this volume. But our intention was only to whet the reader's appetite. Having served the hors d'oeuvres, we trust he or she is now ready for the main course. Bon appétit!

Contributions

Abram Bergson

Harvard University

Sharad Bhandari

University of Pittsburgh

Bernard Chavance

CEMI (EHESS), Paris

Mathias Dewatripont

ECARE, Université Libre de Bruxelles

Jacques H. Drèze

CORE, Université Catholique de Louvain

Michael Ellman

Universiteit van Amsterdam

Richard E. Ericson

Columbia University

István R. Gábor

Budapest University of Economic Sciences

Paul Hare

Heriot-Watt University, Edinburgh

David Kemme

University of Memphis

János Köllő

Institute of Economics, Hungarian Academy of Sciences

Georges Korsun

University of Maryland, College Park

Maria Lackó

Institute of Economics, Hungarian Academy of Sciences

Marie Lavigne

Institut de Sciences Mathématiques et Économiques Appliquées, Paris

Edmond Malinvaud

Centre de Recherche en Économie et Statistique

Béla Martos

Institute of Economics, Hungarian Academy of Sciences

Eric S. Maskin

Harvard University

Peter Murrell

University of Maryland, College Park

Yingyi Qian

Stanford University

Richard E. Quandt

Princeton University

John E. Roemer

University of California, Davis

Gérard Roland

ECARE, Université Libre de Bruxelles

Jeffrey Sachs

Harvard University

András Simonovits

Institute of Economics, Hungarian Academy of Sciences

Robert Solow

Massachusetts Institute of Technology

Wim Swaan

Economic Center for Research, Hungarian Academy of Sciences, Budapest

Jörgen W. Weibull

Stockholm School of Economics

Martin Weitzman

Harvard University

Wing Thye Woo

University of California, Davis

Chenggang Xu

London School of Economics

A Short Biography of János Kornai

János Kornai was born in Budapest in 1928. He began (but did not complete) his university studies after the Second World War. At the age of seventeen, he joined the Communist Party as a true believer and rapidly advanced in its ranks. At nineteen he was appointed economics editor of the *Szabad Nép*, the daily with the largest circulation in Hungary. This was Kornai's first real "university," and it offered him the chance to observe the socialist economy from within. When the crimes of Stalin's regime came to light, however, he lost his faith in the system irretrievably. He joined the party's anti-Stalinist wing and accordingly was fired from the *Szabad Nép*. In 1956 he left the Party for good, whereupon he was branded a renegade and a traitor. In this respect Kornai's experience resembles that of Koestler, Malraux, or Orwell: a committed Party member sees the light and becomes an implacable critic.

In 1955 Kornai entered the newly founded Institute of Economics (Hungarian Academy of Sciences), and the following year produced the first book (*Overcentralization*) by a citizen of a communist state that described the socialist system realistically and critically, rather than in idealized form. This was Kornai's Ph.D. thesis. In Hungary, Ph.D. defenses are normally conducted in the presence of a small number of experts. However, several hundred people came to hear Kornai defend, and the occasion turned into a mass demonstration against the failures and absurdities of the system. After the book was published (in Hungarian in 1957 and in English in 1959), he was fired from the Institute for initiating "revision threatening the socialist system" and "counter-revolution" and as "a traitor to Marxism-Leninism." Fortunately, he was not banned from scientific work. Between 1958 and 1963 he worked in the less prestigious Institute of Textile Industry, where he started to learn mathematical economics.

In 1963 Kornai entered the newly formed Cybernetics Institute (also part of the Hungarian Academy of Sciences). The main fruit of this stage of his career was his monumental work *Mathematical Planning*, first published in 1967. He also guided a team of several dozen economists and computer experts, charged with making planning more rational (a concise description of the underlying mathematical method of two-level planning was published in *Econometrica*, co-authored by Tamás Lipták).

In 1967 Kornai was reinstated to the Institute of Economics, where he remained until 1992. By this time he was pessimistic about the prospect for rational planning and turned to general equilibrium theory (GET) for the light it might throw on economics. One fruit of this study was *Anti-Equilibrium* (published in 1971), which argues that GET's grip on economic reality would be improved by incorporating elements of bounded rationality, systems theory, and disequilibrium theory.

Kornai began work in the 1970s on several research topics laid out in *Anti-Equilibrium*. This led to a number of journal articles and the book *Nonprice Control*, jointly edited by Béla Martos in 1981. At the same time he produced a monograph, *Economics of Shortage*, published in 1980, perhaps his most influential work. The book has been translated into many languages, including a 100,000-copy printing in Mandarin. Ironically, it was not published in Russian until 1990, although the Soviet Union probably provided the best evidence in support of the thesis that shortage is an inevitable feature of the socialist system (indeed, that is probably why it was not translated earlier).

Kornai's early monograph *Overcentralization* created a stir in the West, and from 1958 onward he received many invitations to visit foreign institutions. However, he was denied a passport by the Hungarian authorities and was not allowed to travel until 1963, after political repression had begun to ease. He had several visiting appointments at universities in the United States and elsewhere. In 1986 he accepted a permanent professorship (on a half-time basis) at Harvard University. In 1992 he left the Institute of Economics for the Collegium of Budapest, which is the Hungarian equivalent of the Institute for Advanced Study at Princeton.

Kornai started work in 1988 on a book summarizing his life's work on socialism (it was published in 1992 as *The Socialist System*). At the same time, he joined in the task of analyzing postsocialist transition and formulating policy recommendations. His book *The Road to a Free Economy* (Hungarian version, 1989; English version, 1990), which has been published in sixteen languages, was a pioneering study of stabilization, liberalization, and privatization. Ever since, he has been working on transition issues. Happily, he has witnessed the virtual disappearance of economic shortage, one of the major features that he attributed to the socialist system. Less fortunately, he has seen one of the predictions of *Anti-Equilibrium* come true: the transition from a "shortage" to a "slack" economy typically gives rise to recession.

Kornai's influence has been profound. Vast numbers of people were first exposed to the true nature of socialism through his books. And his critical writings contributed significantly to early attempts to reform communism. It is remarkable too that the number of Kornai students far outstrips that of any other economics professor in Eastern Europe, although he has never had a regular affiliation with any Hungarian university.

We have no space to list all of Kornai's many honors. Let us just mention that he received the F. E. Seidman Distinguished Award in Political Economy in 1980 and the Alexander von Humboldt Prize (Federal Republic of Germany) in 1983. He was president of the Econometric Society in 1978 and of the European Economic Association in 1987.

Having outlined János's scientific career, we would like to indulge in a few personal remarks.

Simonovits's reflections: I first met János in 1969 and worked under his guidance over two decades. We wrote several papers together, and from this I absorbed the elements of good economic research: the importance of choosing a significant topic, studying the literature on the subject thoroughly, and expressing oneself as simply as possible, always emphasizing the economics, not technique.

I attended several of János's courses at the Budapest University of Economics, where he has been an Honorary Professor since 1968. I was astonished by how much time János devoted to students. If there were too many student papers to read by himself, he would distribute some of them among his colleagues and assistants. I recall that, as one of those unofficial assistants, I spent hours in discussion with him and the students whose papers I refereed.

Maskin's reflections: I first got to know János when he joined the Harvard Economics Department in 1986. We would meet from time to time to discuss how some of his intuitions and observations about the socialist system could be modeled formally. These conversations became more frequent when I had the pleasure of acting as his discussant for his 1991 Tanner Lectures at Stanford. We went through the text of the lectures beforehand almost line by line, arguing every step of the way. It was an exciting and intense experience.

János also had a profound effect on my research agenda. Indeed, his concept of the soft budget constraint syndrome inspired Mathias Dewatripont and me to propose a formal theory for why centralized economies are more vulnerable to this syndrome than are decentralized economies.

But the most rewarding aspect of having János as a colleague—indeed, one of the great highlights of my career—was our joint supervision of a remarkable cohort of Chinese graduate students, all deeply fascinated by the working of the Chinese economic system. These students turned to János for his unsurpassed expertise in the economics of socialism and to me for help with the tools of modern theory. It was an honor to participate with János in their training. Two of those students—Yingyi Qian and Chenggang Xu—are represented in this volume.

Of course, we admire János for his accomplishments as an economist but we love him for his great warmth and friendship. He is devoted both to the well-being of his family—his wife Zsuzsa, his children Gábor, Judit, and András, and a host of grandchildren—and to that of his colleagues. It is a privilege to know him.

Long live János!

Eric Maskin and András Simonovits

Introduction: Some Notes on Assessments about Economic Systems

Edmond Malinvaud

To understand how alternative socioeconomic systems function has been the main concern of János Kornai throughout his life. I would like to present some personal reflections touching on this major concern and on related challenges facing the economic system of Western European democracies.

The two of us did not grow up as economists in the same intellectual environment. Mine was probably more marked by the historical background of the nineteenth-century search for alternative systems to capitalism and by the belief—in my part of the world at least—that the best compromise would rely not only on markets but also on the welfare state, that is, on enlightened state guidance and on the redistribution provided by the institutions of social security (in the broad sense of the word). Our choice of a profession was also somewhat different. I decided to be a socioeconomic statistician as well as an economist, but to concentrate my research on economic problems, excluding sociological and political issues. Notwithstanding these differences there was, and there is, a large degree of mutual understanding between us. This is why, on this occasion, I shall go somewhat beyond the domain I chose.

More precisely, I shall first consider how we can objectively assess the performances of alternative systems; I shall hint that in this respect the verdict of history provides a more powerful test than the laments of sociologists or the logic of economic reflections. I shall then give my assessment about the role of planning in market economies. Finally, I shall join all those who now, like János, wonder about the future of the welfare state; the theme will be: how can we assess the future of economic systems?

Economists, Historians, and Others

For the general public, choosing between economic systems most often appears to be a question of ethics and values. You are for capitalism because it brings free enterprise and pays for saving as well as efforts; or you are for socialism because it cares about the needy and mitigates wealth inequalities. You are now for the U.S. system because it achieves nearly full employment; or you are for the continental Western European system because it has managed to avoid the decrease in the real wages of unskilled labor.

We economists know that the question is much more complex. Capitalism and socialism must not be judged according to what their respective proponents say, nor according to what happens a few years after a change of system, but according to

their long-run economic and social performances. The U.S. and continental Western European systems cannot be evaluated by just a comparison bearing on what happened in the two regions during the past twenty years; we need a framework of analysis in order to disentangle what comes from different economic systems and what is otherwise explainable.

Because economic systems are not meant to change often, the relevant assessments about them concern the long run. Moreover, economists as different as Friedrich Hayek, Gunnar Myrdal, and János Kornai have pointed out that the choice of an economic system has not only social, but also legal and political implications, and it cannot be an independent, purely economic choice; therefore the relevant assessments should not isolate economic performance from other aspects of societies in which they would exist.

Economists, as we know, base their conclusions either on observation or on logical reflection, or still on various mixtures of the two. The methodology works, more or less, on well-defined and limited economic issues. But it is strained when one pretends to use it in order to assess all the long-run implications of the choice of a socioeconomic system. The relevant observed samples are too small; the logical analysis should involve too many components and too many links about which assumptions cannot be firmly founded on undisputable evidence.

However, the stakes are so important that at least some of us cannot refrain from trying to go beyond what a strict application of our methodology would require. Few go as far as to write and publish the results of their reflections, but we must be thankful to those who do. And we must also recognize that many of us often took positions in informal talks; such talks cannot but have some impact on the prevailing views that economists hold, in various parts of the world, on how alternative economic systems work and perform. It is interesting to wonder how views are so formed, either by the few who publish on the issue or by the common wisdom that more collectively and informally emerges in various environments.

It is proper, I believe, to call "interpretation" the main operation by which views are formed. Interpretation is based on facts and experience, but not in the same rigorous way as statistical inference ought to be. Interpretation uses mental models, which are themselves part of a cultural heritage and are continually but slowly revised with new experience. These models accept ordinary logic, but more selectively than the users will admit; in particular they are not demanding as regards confirmation of hypotheses, and so they protect users against doubts.

Interpretation is widely used in history and in social sciences other than economics. One might even say that it is the feature distinguishing social sciences from hard sciences, and that its use in economics shows how our discipline is softer than

natural sciences. But history and, I suppose, each one of the other social sciences developed its own standards about what may be a legitimate interpretation. For instance, good historians have to document events that they are reporting and on which they build generalizations, which have to be carefully worded, even though they come from interpretations rather than from statistical testing. I should not be surprised if it were shown that, in their interpretations for assessments about socio-economic systems, economists were often more sloppy than historians. In other words, it seems to me that the frequent use of interpretations in economics should be better recognized, so as to lead us to reflect on the norms we ought to follow when using them.

Considering the subject of this chapter, the main interpretations are those calling on “the verdict of history.” The fall of communism comes to mind. But there are, of course, many other cases. I shall consider in the next part the institutions of planning in Western Europe and recognize that history rejected at least some of the forms of economic planning whose implementation was entertained one or two decades after the war. Another case in point concerns an economic system organized around labor-managed cooperative firms. From the middle of the nineteenth century many politicians and economists tried, within what may be called cooperative movements, to promote such a system. Still, it was recently argued that the system could lead to an efficient allocation of resources. Cooperative firms also exist within market economies, but they are everywhere a minor form of production units. Most economists of today, even among those sensitive to social evils resulting from capitalism, are skeptical about the viability of the cooperative system—mainly because the verdict of history seems to have shown that such a system appears unable to develop in modern societies.

In his writings, János Kornai goes much beyond such simple-minded interpretations of history and raises the level of objectivity of our understanding. For instance, one of his 1990 articles shows that the form of business ownership and the form of coordination mechanism are strongly linked. The article explains why the third ways—attempting to combine the market with other forms of ownership than a predominantly private one—is “weak.” Similarly, cooperative ownership linked with “associative coordination,” an experiment in Yugoslavia, was weak, and in fact bureaucratic coordination prevailed. He then concludes: “History is unpredictable. But *as long as no contrary evidence is provided by experience* (my emphasis), it is worth keeping in mind these preliminary observations.” (See also chapters 20 and 21 in Kornai 1992.)

His book *The Socialist System: Political Economy of Communism* proposes a full explanation on how that socialist system worked. From the political project of the

Communist Party resulted its need to exercise sole political power, to repress all signs of challenge to this power, to nationalize firms, and to organize central quantity planning. Hence the shortage economy with its great lack of efficiency. Trying to reform the system so as to reach a better efficiency actually undermined its consistency; sooner or later the system had to be replaced by a wholly different one.

The argument, supported by detailed analysis and much evidence, is fully convincing today—so convincing that someone of my generation has to wonder why it was not made and widely accepted thirty years ago, half a century after the Soviet revolution.¹ At that time, I believe, the argument that the Soviet system was, and had to be, inefficient would not have looked persuasive to many and would not have been more widely accepted than was the more general abstract argument on the inefficiency of socialism, as it was formulated by von Mises in 1922, by Hayek in 1935, and by quite a few others later.

Of course, the empirical evidence at our disposal on the Soviet economic and social performance misled us. But this is precisely part of the problem. Why would have we needed that evidence? Why did not we receive more correct comparative evidence? Let us consider the two questions in turn.

If many economists, not to speak of others at this point, were not convinced by just the abstract argument, it is because its “proof” that a planned economy could not work efficiently was partly based on interpretations, mostly about what such an alternative system would have to be—mostly, but not only, because the argument had to be comparative. One had also to assess how the market economy of the actual capitalist system worked and performed, both with respect to its efficiency in the allocation of resources and to its equity in the distribution of welfare among people. Interpretations in this second respect varied considerably (and still do). Even on the narrow question of something like Pareto efficiency, many in the generation who were twenty years old in 1930 had very negative assessments about the actual market economies.

Misleading information about the state of Soviet society and about the progress it was experiencing was of course in part due to the Soviet government propaganda and to the distorted statistics it was publishing.² But this is not the whole story, because the existence of such a propaganda and distortion was well realized after the war. According to my tentative assessment, however, we overestimated the real performance in the USSR in 1960, and we also had for long a too favorable image of the changes since then. It is a tentative conclusion because I did not make a serious study on the point, neither on the actual state and evolution of Soviet society nor on the average image of the Soviets we had through the years in the West. This

image of course directly depended on the assessments conveyed by Western intelligentsia.

If my conclusion holds, why were our intellectuals wrong? I shall venture into giving two reasons explaining why propaganda achieved its aim to some extent. First, we were too much impressed by some Soviet successes: in World War II, hence, we thought, by its war economy; and on the technological front, particularly in the nuclear and aerospace fields. Those successes were too easily accepted as such (we did not know their cost) and as more widely significant.

Second, looking at the social problems in our part of the world we fell too much into the mistake of believing that “the grass was greener” in the other part. In this respect information was not explicitly biased; the social problems of Eastern Europe were simply not considered. Unconsciously, our sociologist colleagues gave a hand to Soviet propaganda by just doing their job, namely to detect where our society was deviating from the ideals we liked to assert. Without accepting the Communist political project, whose political features were found by most to be too undemocratic, many were ready to concede that the Soviet system was performing better on the social front.

The Role of Planning

Let us turn our attention to another related but less extreme case in which prevailing views changed over the last five decades, namely the case concerning the best mix of government and markets for economic development, with the role of government planning in particular. My own experience concerns mainly my country and, more indirectly, neighboring European countries. But the change of views in France and in Europe is somewhat similar to the one that occurred in developing countries about at the same issue. So, I believe reflection here may bear on a wider geographical coverage than the one with which I am most familiar.

Assistance is provided in this respect by the recent work in which I was involved with colleagues much more knowledgeable than I am about development economics. Under UN auspices, we prepared a book precisely intended to explain what ought to be the present doctrine about government management of the economy in a development context (Malinvaud, Sen, et al., 1997). The core group for this work was made up of J.-C. Milleron, M. Nabli, A. Sen, A. Sengupta, N. Stern, J. Stiglitz, K. Suzumura, and me. Our message does not claim to be original³; it is summarized by three carefully worded long paragraphs, which I am grossly simplifying here as follows.

1. Many of the mistakes in earlier development policies arose from an inadequate appreciation of the role of markets. However, in some cases markets either do not exist or fail to operate satisfactorily. Unfortunately development experience shows that government intervention also is often exposed to failures and cannot mitigate all market failures.
2. The enhancement of human capabilities is the key ingredient of successful development policies. This does not impose a specific kind of government involvement. Nevertheless, some well-designed planning may be required for ensuring that development objectives are met.
3. Macroeconomic instability is a serious danger to the sustainability of any development process. This implies that development strategies should incorporate, rather than violate, conditions of macroeconomic stability.

I need not insist that such an assessment differs in significant respects from the ones which were current during the first two decades after World War II. The two main differences concern on the one hand judgment about the role of planning, and on the other hand appreciation of the role of markets, both for domestic allocation of resources and for insertion in the world economy, with the opportunities this insertion offers and the disciplines it imposes. The text is written mainly for developing countries and economies in transition. But the same two differences exist between old and new views in Western Europe of our economies.

I will not discuss in general why in market economies there remains an important need for good public management. The idea is widely accepted and the available economic literature fairly well presents the context within which this management takes place and should be geared. The pendulum of fashion has often, during the last two decades, led some economists to see the markets through rosy glasses and government through dark glasses, but that will change again.

On the contrary, the literature now ignores the part played by planning in good government, to the point that many may think that no sensible economist should now refer to the concept. The fact is, however, that some form of planning by public authorities exists in all countries and that, in many places where this activity is on the downgrade, we witness a deterioration in the quality of public services and an excessive intrusion of political considerations. This is why I present here some reflections on the role of planning, reflections inspired by my past participation in French planning and in the planning literature.

Looking back on the postwar West European planning movement, which was marked by such figures as J. Tinbergen, R. Frisch, and R. Stone, we must certainly recognize that it too often thought in engineering terms: the economy was a system;

government had to act on instruments in order to reach national targets; knowing the economic system and the objective function ranking the goals of the country, represented by political authorities, planners would determine the best choice of targets among the feasible choices, and the corresponding optimal set of actions on instruments. This vision was most systematically held by R. Frisch throughout the second part of his life and then motivated all his research. At the end of his life, Frisch particularly emphasized the difficulty of the last phase: how to determine the set of government actions that would implement the optimal plan? But difficulties also concerned the proper representation of the economic system, as well as the interactions between planners, politicians, and the people in democratic countries.

The fact that this engineering conception was somewhat inappropriate was, however, realized by many of those who had actual planning responsibilities, even in France, precisely where most of the latter were so-called engineer-economists. Under the intellectual leadership of a few high civil servants, particularly P. Massé, we had reached a different conception of our role, a conception for which I may quote J. J. Carré, P. Dubois, and E. Malinvaud (1975, p. 493):

The influence of national planning has been exerted chiefly through the changes it led to in the attitude and behavior of decision makers, public and private. It has not, in the main, stemmed from the mere existence of official reference documents that the text of the Plan represents, but far more from the concerted action, the efforts of the Plan's working groups and commissions, the studies stimulated by planning within government agencies and firms, and finally the actions of the Plan's Commissariat while each Plan was being put into effect. The role of the exchange of information and ideas, of increased awareness of the problems to be resolved, and of the need for research to provide solutions stemmed from the process of seeking concerted action, which led decision makers themselves to define the outlook for development and its necessary conditions.

Clearly, we were far from the engineering conception. We were also far from the practice of planning by direct bureaucratic control, which, as explained by J. Kornai (1992, chapter 7), turned out to be "a curious combination of order and disorder, foresight and haste, service of great common interest and uninhibited assertion of partial interests" (p. 130). We were looking for order, foresight, and common interest, but we did not pretend to control the whole process that might lead to them.

The conception recalled here emphasizes the role of the planning institutions and the planning process rather than that of the planning documents. It perfectly fits within the institutional approach advocated by D. North (1994), according to whom we cannot understand economic development if we "ignore the incentive structure

embodied in institutions that determine the extent of societal investment" in technology, human capital, and the like, institutions that structure human interactions, and individual beliefs and that are essential parts in a process of collective learning. The role of institutions in economic development is indeed widely recognized now. For instance, a World Bank book (1993) is largely devoted to the institutions that improved economic performance of the markets and of government interventions, so contributing to the success of East Asian countries. (It is, however, revealing to note that the book hardly mentions the role of planning, despite the fact that most of the countries concerned have planning institutions.)

Because planning must be viewed within the wider institutional framework and economic system of the country, its nature and design may vary a lot from one country to another. There is no reason to believe in particular that the pattern existing in any one of the most advanced countries will fit the needs of less developed ones. Before deciding about planning institutions and processes, one ought to consider a number of questions. Existing planning institutions are often justified as substitutes in charge of functions that are elsewhere performed otherwise. This may be seen by consideration of the traditional economic functions of government, as well as by discussion of its part in the information gathered and processed for, and by, all economic agents, private or public.

The theories of bureaucracy and public choice often argue that long-term concerns and economic efficiency criteria are not the main determinants of decisions within government. If so, there are, on constitutional grounds, good reasons for the existence of institutional structures and processes concerned with efficiency. Those may be found within technical ministries, budget administrations, parliamentary committees, houses of commissioners of audit, and the like; but where such structures are weak or not economically minded, a planning institution may be essential for exhibiting the long-term policies, watching the sustainability of the public debt and the foreign debt, looking at the coherency, coordination, and rigor in the allocation of resources to the various parts of the public sector, which is everywhere sizable, and to the various public investment projects, among which cost-benefit analyses often have to bar extravagant political requests. Once this planning institution has become credible, it has the ability to insulate government from excessive interest-group influence.

Economic agents get information more or less directly from many sources: the markets, the economic and technical press, their business or trade associations, government agencies, and so on. For long-term decisions the information conveyed by markets is not transparent; it has to be complemented by information from other

sources and it has to be competently processed. The productivity of an economy depends on the adequacy of these diversified sources. Clearly, the range is wide: at one extreme stands the United States, with its developed market system, its large service industry, its rich set of private institutions, and its widespread technical and economic education; at the opposite extreme we find a large number of developing countries. Building and maintaining planning institutions in charge of gathering, processing, and diffusing information is needed particularly where the education of producers and traders is low, where private sources of information are either lacking or conveying inadequate messages about technical knowledge and socioeconomic trends.

A particular information-processing function concerns the set of prices to be used in cost-benefit analyses of public projects. For reasons that were widely discussed and will not be repeated here,⁴ some market prices have to be corrected. Moreover, there are no market prices for most inputs and outputs implied by the projects at future dates except the most proximate. It is natural to give to a planning office the function of finding what ought to be the main price indicators to be used in the decentralized evaluation of public projects, precisely because the values of these indicators depend on a general equilibrium analysis.

Such were my ideas when, in July 1963 at Cambridge, I had the pleasure of first meeting János Kornai. I had already seen his article, written with his mathematician colleague T. Lipták, on "two-level planning," and the article happened to be closely connected with my own paper on decentralized procedures for planning.

The latter had been written in order to provide some foundation to the analysis that was performed in French planning, precisely in order to determine the price indicators to be used in project evaluations; it followed the stimulation given by a working party organized by P. Massé, to which M. Boiteux had also been taking part. I still believe this "application" was more relevant than more ambitious ones entertained at the time and inspired by the engineering conception of planning. In a very persuasive section of his 1992 book (section 17.4) János Kornai refers to such broad applications of "mathematical methods" in planning. He shows how they were not really used in centrally planned economies and why they could not be.

Indeed, if viewed as providing methods to be directly implemented on a large scale, the mathematical theory of economic planning, as surveyed by G. Heal (1973), cannot be taken very seriously. It should rather be seen as aiming at conveying some initial understanding of the processes through which relevant information is collected, for instance for decisions about the provision of public goods, to quote another relevant "application."

Challenges to the Social Protection Provided by the Welfare State

The most difficult question about the future of the economic systems of developed countries concerns the institutions of their so-called welfare state, from pure income redistribution to employment policies. Most economists agree that there are two serious challenges: (1) to discern what will be the evolution, starting from what seems to be in many countries an unsustainable system, (2) to recommend reforms to our fellow citizens. Indeed, János Kornai recently turned his attention to the latter problem, thinking more particularly of the case of Hungary (Kornai 1996). The following tentative notes will present how for my part I perceive the difficulties (see also Malinvaud 1994).

Most of us, as economists, tend to look at the problems in terms of feasibility and efficiency. Moralists, on the contrary, see them in terms of ethical principles. As soon as we want to make recommendations, we also have to be explicit about the principles inspiring us. Even pure predictions rely on some hypotheses concerning the predominant ethics in the society and its interaction with the evolution of the performance of the system. Thus it is natural to begin with ethical principles.

In his 1996 article János Kornai indeed does so and states in this respect two principles, which will serve also as backbone for a brief discussion here. His second principle is "solidarity," specified as, "Those who are suffering, in trouble or disadvantaged must be helped." As a principle this is widely accepted, although some moralists might phrase it a little differently and refer to the rights of those in need of help (they might also object to Kornai's expression "compassionate solidarity"). The main question is to know the degree and form of solidarity, and it is where the other main principle comes in.

Some would simply call it "individualism." Kornai uses the definitely better short expression "human dignity." This is often specified in terms of guarantee of basic rights of the person, as an individual whose freedom is respected and as a participating member of society. But these basic rights sometimes happen to conflict with one another and to conflict with the basic needs, to which the above second principle refers. This is why the interpretation given to human dignity matters.

Having suffered from living in a totalitarian regime, Kornai, quite naturally, argues that "the autonomy of the individual needs strengthening against the sprawling Leviathan of the state." Others, living in a different environment, might stress different aspects of human dignity, for instance the right to work and the frustrations when it is denied by discrimination or lack of jobs. Where priority is placed between solidarity and human dignity, as well as within each, has great importance, of course, for decisions concerning reform of the welfare state.

When ethical dilemmas appear, the question of knowing what people prefer comes to mind and the proposal of leaving decision to the democratic process is most often made. The third "transparency" principle in Kornai (1996) seems to so conclude. But Kornai expresses, in this respect, a distrust of politicians and a concern about the fact that people show a lack of tax awareness. He then calls on the intelligentsia to have the courage to exhibit the facts and to take into account the dishonesty of representatives and officials. My own assessment about the same issue would be a little different, partly because I lived beside honest civil servants but not only because of that.

First, I am impressed by the contradictions in modern attitudes. The prevailing moral philosophy asserts both solidarity and individual autonomy, but it does not recognize the difficulty in reconciling the two, particularly on questions concerning the sustainability of a reformed but still important welfare state. This philosophy indeed tends to refuse the means required: social compulsion, redistribution of wealth, recognition of individual duties toward others, starting with family dependents and close neighbors. The common ethic is benevolent in principle but cynical in individual and political behavior; verbal compassion on misfortunes often coexists with eagerness to collect the full returns of good fortunes. I anticipate the resulting tensions to keep increasing; I see accumulating ingredients for heated conflicts.

Second, I remember that in the past social policies were often promoted by leaders who felt a duty to protect individuals against shortsightedness and other weaknesses of their own spontaneous choices: public pensions, mandatory insurances, compulsory schooling, drug legislation, and so on., were so motivated. I do not think similar considerations are obsolete. Public economics ought to still now recognize the relevance of so-called merit wants. In other words, now as it was the case in the past, the intelligentsia has a major role to play in helping our societies both in the selection of just and consistent ethical principles, and in their implementation.

Indeed, the main problem with present welfare states does not come from a wrong choice of objectives but rather from an incorrect assessment of the range of feasible alternatives. The main problem concerns long-run implementation. This was particularly well argued in the recent writings of A. Lindbeck: in the past an overly optimistic vision was entertained about microeconomic behavior and about macroeconomic regulation.

Lindbeck (1995) discusses the microeconomic aspects, particularly in efficiency terms, to which I now limit attention. He first points to the fact that welfare arrangements are indeed more efficient in some respects than what would happen in free markets on a voluntary basis: because of the difficulty of borrowing with human capital as collateral, of the high administrative costs of voluntary insurance, and of

other considerations, compulsory social security is comparatively efficient in dealing with redistribution of income over the life cycle, reduction of income risk, and stimulation of investment in human capital. But opposite such advantages there are static efficiency costs due to distortions resulting from tax arrangements financing the welfare state and from benefit arrangements. Lindbeck puts still more emphasis on a thus far neglected dimension of the efficiency problem, namely what will happen in the long run, as some incentive effects gradually accumulate and interact with other factors, possibly in the form of virtuous circles, more probably in that of vicious circles, as people learn how to adapt their behavior and as social norms evolve.

Lindbeck (1996) looks at the relations between the welfare state and full-employment policies. He argues that the relations are much more complex than was earlier recognized, when it was simply said that development of public social security was strengthening automatic stabilizers. Actually, it now seems that the automatic budget responses will stabilize aggregate output and employment only up to a point relative to the sizes of the budget deficit and of the public debt, a point beyond which private agents perversely react. Moreover, generous welfare benefits, stricter job-security legislation and higher minimum wages raise the so-called equilibrium unemployment rate. Lindbeck even argues that the long-run incentive effects are likely to increase unemployment persistence.

The various investigations on which these conclusions are based often reveal a high degree of uncertainty about the strength of the effects envisaged. Among the various challenges that reform of the welfare state puts to the profession of economists, the difficulty of reliable assessment of relevant effects ranks high, because no other profession can say anything objective about them. Meeting this challenge will be particularly important to us under the optimistic scenario in which the nexus of problems with demography, medicine, natural resources, economy, and society would progressively turn out to be less acute than expected by some. Then reformers will need to be precisely informed about the consequences of alternative programs, which will differ in the extent of public controls and regulations, in the degree and form of public compensation for disadvantages and bad fortunes, and in the rules defining taxes and social security contributions.

Notes

1. For whatever it may be worth, I confess here the personal conclusion I had reached in the late 1940s and held throughout the following decades: the repression of public and private liberties was, according to me, inconsistent with human aspirations and therefore unsustainable in the long run; sooner

or later it would have to give way, quite independently of whatever economic performance could have been achieved. While agreeing with János's book about the starting point, my argument was clearly different.

2. As my level of responsibility in French official statistics was rising, I felt increasingly uneasy in this respect, particularly in international meetings of official statisticians, where the problem was never faced. My Western colleagues and I thought silence was imposed on us by the requirements of high international politics.

3. Indeed, similar messages can be found elsewhere, as we may see when we read D. Rodrik (1996).

4. See for instance J. Drèze and N. Stern (1987).

References

- Carré, J.-J., P. Dubois, and E. Malinvaud (1975). *French Economic Growth*. Stanford: Stanford University Press.
- Drèze, J., and N. Stern (1987). The theory of cost-benefit analysis. In A. Auerbach and M. Feldstein, eds., *Handbook of Public Economics*. Amsterdam: North Holland.
- Hayek, F. (1935). The present state of the debate. In F. Hayek, ed., *Collectivist Economic Planning*. London: Routledge.
- Heal, G. (1973). *The Theory of Economic Planning*. Amsterdam: North Holland.
- Kornai, J. (1990). The affinity between ownership forms and coordination mechanisms: the common experience of reform in socialist countries. *Journal of Economic Perspectives* 4: 131–147.
- Kornai, J. (1992). *The Socialist System: The Political Economy of Communism*. Oxford: Clarendon Press.
- Kornai, J. (1996). The citizen and the state: reform of the welfare system. Collegium Budapest, Discussion Paper 32.
- Kornai, J., and T. Liptak (1963). Two-level planning. *Econometrica* 33: 141–169.
- Lindbeck, A. (1995). Incentives in the Welfare State: Lessons for Would-Be Welfare States. Austin Robinson Lecture, Congress of the International Economic Association.
- Lindbeck, A. (1996). *Full Employment and the Welfare State*. Memphis, Tenn.: P. K. Seidman Foundation.
- Malinvaud, E. (1967). Decentralized procedures for planning. In E. Malinvaud and M. O. L. Bacharach, eds., *Activity Analysis in the Theory of Growth and Planning*. London: Macmillan.
- Malinvaud, E. (1994). Is the European welfare state unsustainable. *Rivista di Politica Economica* (December): 237–256.
- Malinvaud, E., A. Sen, et al. (1997). *Development Strategy and Management of the Market Economy*. Clarendon Press, Oxford.
- Mises, L. von (1922). *Socialism: An Economic and Sociological Analysis*. London: Jonathan Cape.
- North, D. (1994). Economic performance through time. *American Economic Review* (June): 359–368.
- Rodrik, D. (1996). Understanding economic policy reform. *Journal of Economic Literature* 34: 9–41.
- World Bank (1993). *The East Asian Miracle: Economic Growth and Public Policy*. New York: Oxford University Press.

I **MARKETS AND ORGANIZATIONS**

1 Monopoly under Labor Management

Abram Bergson

A curious feature of the labor-managed enterprise (LME) is its response to monopoly power. As Benjamin Ward (1958), Jaroslav Vanek (1970, pp. 58ff), and James E. Meade (1974) have shown, a LME enjoying a pure monopoly can be expected theoretically to be even more restrictive than a capitalist firm (CF) in that position. With demand, available technologies, and the capital stock the same in one case as in the other, the LME predictably will produce a smaller output than its capitalist counterpart. That is so provided only that the latter is operating profitably.

Insofar as capital is given, the analysis relates to the short run, but as Vanek has pointed out, the same result obtains in the long run, where capital is variable. Indeed, should the LME have a linear homogeneous production function, the restriction must take an extreme form such as a CF could be expected to impose only if demand elasticity were less than unity. The LME might be comparably restrictive even should the elasticity of demand for its product exceed unity.

That might still be the result, moreover, should the LME experience economies of scale. Or so Vanek affirms, and, although no formal proof is given, the proposition is too transparently plausible to dispute. Nevertheless, his brief comment leaves unanswered the further question as to precisely when such restrictive behavior is indicated. While addressing that interesting aspect, I may also be able briefly to provide further perspective on the behavior more generally of LME under pure monopoly.

The Short Run

I refer to a LME whose output, X , is described by a function:

$$X = F(L, K), \quad (1.1)$$

where L is the number of workers employed, and K , the firm's capital stock. Although the analysis is readily formulated for the case where there are different kinds of workers, nothing is gained by doing so here, so L represents a homogeneous labor force. Similarly, it suffices to refer to a firm utilizing but one kind of capital goods.

The objective of the enterprise is supposedly to maximize net income per worker (NIW), or

$$Y = (pX - rK)/L, \quad (1.2)$$

where p is the price of X . The firm pays for use of its capital stock a rental rate, r , that is taken to be competitively determined. The firm confronts a demand schedule,

$$X = D(p), \quad (1.3)$$

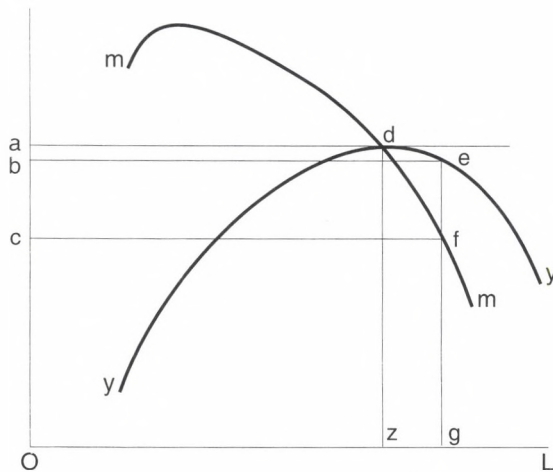
so p is in effect a function of the firm's output.

Our concern is with the long run, but it facilitates inquiry into that if we first review essentials concerning the short run. Because K is then fixed, output becomes a function of L alone,

$$X = \bar{F}(L), \quad (1.4)$$

and Y similarly depends only on that variable. In figure 1.1, adapted from Vanek (1970, p. 99), Y varies with L as in yy , and reaches a maximum, as theory requires, at d , where it intersects mm , representing a corresponding schedule of net income added by a marginal worker (NIA), that is,

$$\text{NIA} = d(pX - rK)/dL. \quad (1.5)$$



- yy : net income per worker (NIW)
- mm : net income added by marginal worker (NIA)
- f: CF maximum
- d: LME maximum

Figure 1.1
Different optima

With K given, this comes to the same thing as the marginal revenue product of labor, M_l :

$$M_l \equiv R\bar{F}'(L), \quad (1.6)$$

where R is marginal revenue, that is,

$$R = p(1 - 1/N_{xp}). \quad (1.7)$$

The elasticity of demand for X , N_{xp} is understood to be scored positively.

Because demand, the technologies, and the capital stock are the same for our CF as for the LME, the yy and mm schedules of one are the same as those of the other. Thus, at the wage rate Oc , the CF employs Og workers. As represented by the rectangle $Ocfg$, its wage bill is less than its total net income, allowing it a profit $cbef$, which is also a maximum one. At d , then, total employment, and hence output, of the LME are less than the corresponding items for the CF.

In the figure, NIA is shown, as it is customary, as initially exceeding and then falling short of NIW. As a result, yy is markedly concave from below, and d does indeed represent a maximum of NIW. Symbolically, we see by differentiating Y with respect to L that

$$dY/dL = (M_l - Y)/L, \quad (1.8)$$

so

$$dY/dL \cong 0 \quad (1.9)$$

according to whether

$$M_l \cong Y, \quad (1.10)$$

or,

$$(R/p)\bar{N}_{xl} \cong 1 - (rK/pX). \quad (1.11)$$

Here $\bar{N}_{xl}$ is the output elasticity of labor.

The assumed configuration of yy thus follows if, as L increases, the sign in (1.11) is initially $>$ and finally < 0 . Equality of the two would then materialize at an intervening point. Reasons to suppose that that sequence does indeed prevail are familiar. Although in principle there is no bar to an N_{xp} that is ≤ 1 , the analysis only becomes interesting when $N_{xp} > 1$. Supposedly it is also $< \infty$. With that

$$0 < R/p < 1. \quad (1.12)$$

Also, $\bar{N}_{xl}$ must be positive, and initially could even exceed unity; in any event, the left side of (1.11) is positive, and at low levels of employment should exceed the term on the right. Unless fixed charges per unit of sales (rK/pX) are quite low even at low employment levels, the right hand term should initially be very small, if not negative.

Under the impact of diminishing returns, however, $\bar{N}_{xl}$, and presumably the left-hand side of (1.11) should become indefinitely small as employment expands. The right side, though, should approach unity, so the relationship between the two sides is reversed, as required.

The Long Run

Turning to the long run, the production function in (1.1) is now fully operative, but although K is variable, its volume always is such that its marginal revenue product equals the capital rental rate:

$$M_k \equiv RF_k = r \quad (1.13)$$

where F_k is the marginal product of capital, labor given. Hence K is implicitly a function of L , and Y again depends on the latter alone. How might it be expected to vary with L ?

In question again is the relation dY/dL . As given by (1.5), NIA again represents the net income added by a marginal worker. With capital now variable, however, the differentiation called for is of a total character. But even so, we have

$$\text{NIA} = M_l \equiv RF_l. \quad (1.14)$$

Here R is, as before, marginal revenue, and, though with somewhat different notation, F_l is much as before the marginal product of labor, with capital given.

Although capital is variable, this odd result follows easily, and becomes understandable when we consider that, as indicated, capital is always employed to the point where its marginal revenue product equals the capital rental rate. As capital varies marginally with labor, it adds to net output value an amount that is equal to the rental charges concomitantly being deducted from that value. The change in net output value, thus, reduces simply to the marginal revenue product of labor, capital given.¹

It also follows that

$$dY/dL \geq 0, \quad (1.15)$$

depending on whether

$$M_l \cong Y, \quad (1.16)$$

or, much as before,

$$(R/p)N_{xl} \cong 1 - (rK/pX). \quad (1.17)$$

Here, N_{xl} is the output elasticity of labor, K given.

Should $F(L, K)$, as often assumed, be homogeneous, however, (1.17) translates at once into another, more interesting relation:

$$(R/p)h \cong 1. \quad (1.18)$$

Here h is the degree of homogeneity, and thus represents scale economies or diseconomies.² In brief, dY/dL is $\cong 0$ depending on the extent to which such scale effects offset or more than offset the shortfall of marginal revenue from price.

In trying to gauge what this means, we may usefully refer to these hypothetical demand elasticities, corresponding values of R/p and magnitude of h that just offset R/p , and so assure a null dY/dL :

N_{xp}	R/p	h
1.5	.333	3.000
2.0	.500	2.000
2.5	.600	1.667

Note that h represents the effect of scale exponentially; for example, for $h = 2$, a doubling of inputs means a quadrupling of output. Should demand elasticity be constant, and h be of the magnitude required for a null dY/dL , as might be suspected, the resultant constancy of Y and the appropriate r are achieved when L and K expand at the same rate.³

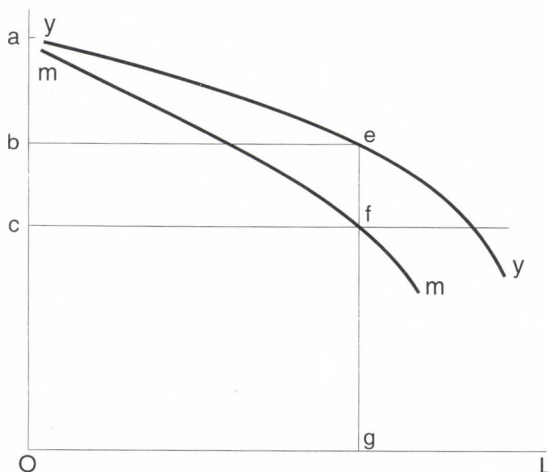
To come finally to the question posed at the outset, implications as to LME behavior are more or less evident, but it is well to be explicit that, whatever the magnitude of h , Y could not be expected to vary with L , in the long run, in conformity with a schedule as markedly concave from below as that obtaining in the short run. With both factors varying, the long-run yy could hardly replicate that for the short run in that regard.⁴

Vanek nevertheless considers that the long-run yy might still be concave from below so that Y initially increases and then declines. From (1.18) we see that that would be so if the exponential scale factor, h , should at first be large enough to more than offset numerically the shortfall of marginal revenue from price, but should subsequently fall below that magnitude.

And should that be so, of course, the LME again attains a maximum at a point such as d in figure 1.1, where dY/dL vanishes and where, much as before, the long-run yy is intersected by a corresponding mm , representing the variation, in the long run, of the marginal revenue product of L with L . Similarly, the resultant employment and output of the LME again fall short of the corresponding items for a CF. The latter would reach equilibrium at a point such as f , where profits are such as represented by $cbef$ and again a maximum.

But circumstances could easily be otherwise. Among other things, Y in the long run, rather than increasing initially, might decline throughout. That must be so if diseconomies of scale ($h < 1$) should be experienced from the outset, or if the production function should throughout be linear homogeneous ($h = 1$), so that there are neither economies nor diseconomies of scale. It could also be so even if there are economies of scale ($h > 1$), provided only that they are insufficiently potent to make good the deficit of marginal revenue relatively to price. Depending on the demand elasticity, admissible scale economies evidently could be sizable.

And the result in all cases is the anomaly that Vanek noted but only partially delineated. Graphically, the schedule for Y might still be concave from below as shown (figure 1.2). But that need not be so, and whether it is so or not, with $(R/p)h$



yy: net income per worker (NIW)

mm: net income added by marginal worker (NIA)

Figure 1.2
Schedules

< 1 , Y already declines with L at low levels of L and continues to do so as L increases. While M_l is initially equal to Y , mm necessarily declines more sharply than yy . For the LME, then, equilibrium is attained paradoxically with employment at a negligible level. It is at that level that Y is a maximum. On the other hand, the CF attains an equilibrium of a usual kind at f , where the wage rate Oc equals the marginal revenue product of labor, and profits, $cbef$, are a maximum (see figure 1.2). Thus, while generally producing a smaller output than the CF, the LME might well eschew sizable economies of scale that the CF would find it profitable to exploit.

I have tacitly assumed throughout that labor, capital, and output are infinitesimally divisible. Caveats needed to allow for indivisibilities are generally evident, but it may well be explicit that for the last case considered, where long run yy declines throughout, the equilibrium level of LME employment that is indicated in the presence of indivisibility is either one worker, or, more realistically, the staff appropriate to a capital stock of technically minimal scale.⁵

Under pure monopoly, as seen here, circumstances where such ultrarestrictive behavior is indicated are probably more usual than often supposed.⁶ The analysis thus underscores the problematic nature of the LME in the context of pure monopoly. Such a market structure is apt to be a rare bird under labor-managed socialism, as it is in fact under capitalism. But the behavior of a pure monopolist under capitalism is to a degree often paradigmatic of that of monopolistic enterprise more generally. That should also be so under socialism. The LME could be expected often to be oddly hyperrestrictive even where monopoly is impure.

Notes

1. With use made of (1.13), (1.14) follows readily from the total differentiation indicated in (1.5), that is with capital as well as labor variable.

2. Formula (1.18) follows immediately from (1.17), (1.13) and the relation:

$$h = N_{xl} + H_{xk}. \quad (1.19)$$

3. To see that it is so under the assumed condition, note that in (1.2) if L and K increase by a factor λ , X increases by a factor λ^h , while p is discounted by a factor $\lambda^{-h/N_{sp}}$. Also, with L , K , and p varying in the required way, (1.13) continues to hold, for F_k increases by a factor λ^{h-1} and p , as above, is discounted by a factor $\lambda^{-h/N_{sp}}$.

4. The two schedules must also differ in another interesting way. If capital charges are at all consequential, Y must be negative in the short run for very small L , and at least for the first unit of L , that must also be true of M_l . Barring consequential indivisibilities (see below) negative values of Y and M_l are not to be expected in the long run as L becomes very small.

5. A curious corollary: if the capital stock should be minimal, there is not so much overhead to spread. Hence, the case for supposing that yy and mm are generally of the figure 1.1, rather than figure 1.2, type in the short run is no longer as compelling as was argued above.

6. Vanek speculates (1970, p. 104) that "even with mildly increasing returns . . . and comparatively inelastic functions, the locus of monopolistic long run equilibrium will be likely to degenerate into, or somewhere near, the origin." With the aid of (1.18), one sees that the degenerate case is much more inclusive than this suggests. For example, consulting the illustrative tabulation given above, we see that the labor-managed firm is still ultrarestrictive even if the output should as much as quadruple with a doubling of inputs, provided only that the demand elasticity is at all short of two.

References

Meade, James E. Labor managed firms in conditions of imperfect competition. *Economic Journal* 84 (336), December 1974: 817–824.

Vanek, Jaroslav. *The General Theory of Labor-Managed Economies*. Ithaca, N.Y.: Cornell University Press, 1970.

Ward, Benjamin. The firm in Illyria: market syndicalism. *American Economic Review* 48(4), September 1958: 566–589.

2 Walras-Keynes Equilibria Coordination and Macroeconomics

Jacques H. Drèze

1 Motivation

1.1

This chapter, a progress report, grew out of two preoccupations, one empirical and the other theoretical.

On the empirical side, I had followed with awesome interest the rise and persistence of unemployment in Europe over the past twenty years (Drèze 1991, 1995; Drèze and Bean 1990). The empirical record reveals clearly that excess supply of both labor and productive capacities has prevailed systematically over that period. Business surveys and econometric investigations confirm that a growing proportion of business firms in Europe (though not in the United States) experienced insufficient demand, and viewed it as the proximate reason for excess capacities. This, of course, does not rule out a deeper explanation in terms of supply shocks or insufficient competitiveness, a theme explored further here. It is also part of the record that real wages did not respond significantly to the rise and persistence of unemployment in Europe (though they did to a large extent in the United States). In short, we witness excess supply of productive capacities and labor, with downward rigidity of real wages.¹

It is also noteworthy that growth rates of output have displayed substantial volatility. Table 2.1 splits the past twenty years into half-decades and reveals differentials of 1.5 percent among successive averages (i.e., of 7.5 percent among overall growth rates for successive subperiods). Yet it seems entirely plausible that 3 percent growth was well within reach throughout.

In assessing that evidence, I was led to wonder (Drèze 1996, section 3.1) about the extent to which demand deficiencies—whatever their origin—might become self-perpetuating due to the well-known coordination failure, or aggregate demand externality: firms do not hire for lack of demand, depressed incomes generate weak demand, and the situation repeats itself until an outside stimulus breaks the circle.

This is where the theoretical preoccupation comes in. Recently, a number of authors have proposed models endowed with multiple equilibria, reflecting coordination failures (information externalities). The nature of the externalities, and the theoretical frameworks, vary from model to model. The list includes Arrow-Debreu economies (Roberts 1987, 1989), overlapping generations models (Geanakoplos and Polemarchakis 1991; Benhabib and Rustichini 1994); market games² (Mas-Colell 1991; Peck, Shell and Spear 1992), as well as a number of partial equilibrium or

Table 2.1
GNP growth (real) average rates

	1976–80	1981–85	1986–90	1991–95
EC12	3	1.5	2.9	1.3
United States	3	2.7	2.5	2.3
Japan	4.4	3.8	4.5	1.4
Belgium	3	0.8	3	1.4
Denmark	2.5	2.7	1.4	2.2
France	3.1	1.5	3.2	1.1
Greece	4.4	1.3	1.7	1.1
Ireland	4.6	2.6	5.2	4.8
Italy	4.8	1.6	3	2.5
Luxembourg	2.3	2.5	4.7	2.7
Netherlands	2.6	1	3.1	1.8
Portugal	1.8	1.5	4.5	1.3
Spain	2.6	1	3.1	1.8
United Kingdom	1.8	2.5	3.4	1.2
West Germany	3.3	1.2	3.4	1.9

Source: *European Economy*, 60.

specific models (Bryant 1983; Cooper and John 1989; Hahn and Solow 1995). Sunspot equilibria in all guises are the subject of a growing literature (Chiappori and Guesnerie 1991).

I was thus led to wonder about the *pervasiveness*, in standard general equilibrium models, of underutilization of resources reflecting pure coordination failures rather than price distortions. The above mentioned work of Roberts, some intriguing unpublished papers by Giraud (1993) and by Citanna, Crès, and Villanacci (1995), and the comprehensive book of Herings (1996), called to mind earlier work on “supply-constrained equilibria”—see van der Laan (1980, 1982, 1984), Dehez-Drèze (1984), and the additional references in Herings (1996, section 4.8). A simple reinterpretation looks at supply-constrained equilibria as possibly reflecting coordination failures rather than the price distortions that motivated the earlier work. I was thus led to reconsider the existence of supply-constrained equilibria also at prices that are either Walrasian or flexible,³ as a way to distinguish pure coordination failures from consequences of price distortions.

A first exploration of that simple idea is presented here. I begin with a transparent example (section 1.2), then introduce a general theorem (section 3), followed by a general discussion (sections 4–5). The model (section 2) is that of a real Arrow-

Debreu economy, hence it remains essentially static at this stage. The theorem establishes the following: if some relative prices are downward rigid, *unlimited* underutilization of the associated resources is possible, that is, it may persist once established, *even when the downward rigid prices are compatible with competitive equilibrium*. A discussion of the price rigidities is postponed until section 5. The emphasis in this essay is on interpretation and implications for macroeconomics. No doubt most of these implications have been noted somewhere, and I apologize for my credit failures.

1.2

To set the stage, I present now (with some simplification) the model of Roberts (1987, 1989) which is a good illustration of the theorem to follow.⁴ The theorem shows that the main conclusion of Roberts does not rest on the specificities of his model, although ancillary conclusions do.

In a notation extended later, there are G goods g , H households h , and F firms f . The production set of each firm is a cone (constant returns), so that competitive profits are zero. The initial endowment of each household consists exclusively of productive inputs supplied to the firms (not consumed). The consumption preferences of each household are homothetic (linear homogeneous Engel curves). Otherwise, assumptions are standard.⁵ Such an economy has a competitive equilibrium, say (x^*, y^*, p^*) , where $x^* = (\dots x^{i*} \dots)$ in R^{HG} , $y^* = (\dots y^{j*} \dots)$ in R^{FG} and $p^* \in R_+^G$. For all λ in $[0, 1]$, $(\lambda x^*, \lambda y^*, p^*)$ is a feasible allocation, also yielding zero profits to all firms, with the consumption of each household optimal given its (labor) income. If real wages, *here set at their competitive levels*, are downward rigid in the face of unemployment, there is a continuum of equilibria with excess supply, which may be viewed as reflecting coordination failures.

Of course the question immediately comes to mind why wages remain downward rigid in the face of arbitrarily severe un(der)employment. Roberts is primarily interested in defining a game for which $(\lambda x^*, \lambda y^*, p^*)$ is a (subgame perfect) Nash equilibrium.⁶ In that framework, he ultimately answers the above question as follows: Bertrand (price) competition among employed and unemployed workers would drive the wages down to the reservation level of the unemployed, who accordingly have nothing to gain (under rational expectations) from trying to undercut the employed. I return to that issue in section 5.⁷

The interpretation of the Roberts underemployment equilibria as pure coordination failures is natural enough. There is no auctioneer. Firms set competitive output prices given factor costs,⁸ then meet demand within the limits of capacity

and profitability. Workers offer their labor services at going wages and refrain, or are discouraged, from undercutting these wages.⁹ Adverse shocks, whether to supply, demand or expectations, easily result in underutilization of resources, possibly viewed as temporary. But output and input levels adjust to each other, with multiplier effects. Once attained, reduced activity levels may persist, because it would take coordinated adjustments of quantities and/or expectations on many markets to restore equilibrium with fuller use of resources. But such coordination lies beyond the capabilities and motivations of individual agents.

There is an undeniable flavor of descriptive realism about that interpretation. There is also a connotation that the prevailing equilibrium at a point in time is a historical datum, being neither uniquely determined by "fundamentals" nor stable. Such equilibria are multiple and path-dependent, even though they are by no means arbitrary. I argue here that these features have important implications for the methodology of macroeconomics. First, however, we wish to check that the existence of such equilibria does not hinge on the special assumptions of constant returns and homotheticity.

2 Preliminaries

2.1

Before looking into the definition, and then the existence, of non-Walrasian allocations at Walrasian prices, I must introduce two precisions.

First, I shall be concerned exclusively with equilibria where, possibly, agents are quantity-constrained in their supplies. Similar results hold for equilibria where agents are only constrained in their demands. The formal symmetry is complete. The reason for my choice is current relevance to Western economies. Whereas excess capacities and unemployment are commonplace, rationing of demand is seldom observed in market economies, and then for narrowly defined commodities over short time spans. Had I been concerned with shortages in communist countries, as was once the case for János Kornai (1980), or perhaps in transition economies today, I would have stressed demand rationing. The nature of coordination failures under communist planning was no doubt different, and I have no competence to discuss it. My tribute to Kornai is on another level. I know his interest for comparative economic systems (Kornai 1990) and for equilibrium concepts (Kornai 1971). The possibility of underemployment equilibria reflecting pure coordination failures in decentralized market economies, also at Walrasian prices, is bound to intrigue Kornai. I hope that we have a chance to discuss the matter.

Second, rationing the supply of some commodities, like money or close real substitutes (gold, assets), would be tantamount to rationing indirectly demand for other commodities; besides, such rationing is not observed. It is accordingly desirable to introduce *ex ante* a distinction between those commodities, say group I, that are never rationed; and those commodities, say group II, that might be subject to quantity rationing of supply. The first group automatically includes all goods whose prices are flexible, including among others real commodities whose markets clear continuously (like perishable commodities sold through auctions) or real assets traded on organised exchanges (like shares of equity or oil futures); see Morishima (1992) for elaboration. This natural distinction is introduced formally in section 3.1 and discussed in section 3.3.

2.2

Consider an economy $\mathcal{E} = (X^h, \succsim_h, \omega^h)_{h=1 \dots H}, (Y^f)_{f=1 \dots F}, \Theta = [\theta_{hf}])$ where each household h is defined by its consumption set $X^h \subset R^G$, preferences $\succsim_h$, initial endowment $\omega^h \in R^G$ and ownership shares in the F firms $\theta_{hf}, f = 1 \dots F$; and where each firm f is defined by its production set $Y^f \subset R^G$. A *feasible allocation* is a tuple,

$$(x^h)_{h=1 \dots H}, (y^f)_{f=1 \dots F},$$

such that:

1. for each $h, x^h \in X^h$; for each $f, y^f \in Y^f$;
2. $\sum_h (x^h - \omega^h) - \sum_f y^f = 0$.

A *competitive equilibrium* is a pair consisting of a feasible allocation (x^*, y^*) and a price vector $p^* \in R^G$ such that¹⁰:

3. for each h, x^{h*} is $\succsim_h$ -maximal on the budget set

$$B^h(p^*, y^*) = \{x^h \in X^h \mid p^* \cdot (x^h - \omega^h) - \sum_f \theta_{hf} p^* \cdot y^{f*} \leq 0\};$$

4. for each f, y^{f*} maximizes $p^* \cdot y^f$ on Y^f .

In the standard specifications, every vector $p \in R^G$ is an *admissible price vector*. When that is not the case, the set of admissible price vectors $P \subset R^G$ must be introduced in the definition of the economy. In this paper, I shall be concerned exclusively with admissible price sets reducible to the form

$$P(\underline{p}, \bar{p}) = \{p \in R^G \mid 0 \leq \underline{p} \leq p \leq \bar{p}\}$$

(where possibly $\bar{p}_g = +\infty$ for some $g's$). I shall denote by $\mathcal{E}(P)$ the economy defined by the pair $(\mathcal{E}, P)$ —occasionally making explicit the definition of P as $P(\underline{p}, \bar{p})$.

A *supply-constrained equilibrium* for $\mathcal{E}(P)$ is a triple consisting of a feasible allocation, a price vector $p \in P$ and a vector of supply constraints $s \in R_+^{G(H+F)}$, $s = (s^1 \dots s^h \dots s^H, s^{H+1}, \dots, s^{H+f} \dots s^{H+F})$, with $s^j \in R_+^G \forall j = 1 \dots H + F$, such that:

3'. for each h , x^h is $\succeq_h$ -maximal on the *constrained* budget set

$$B^h(p, y, s^h) = \{x^h \in X^h | p \cdot (x^h - \omega^h) - \sum_f \theta_{hf} p \cdot y^f \leq 0, \omega^h - x^h \leq s^h\};$$

4'. for each f , y^f maximises $p \cdot y^f$ on the *constrained* production set

$$Y^f(s^f) = \{y^f \in Y^f | y^f \leq s^f\};$$

5. for each g , if $p_g > \underline{p}_g$, then $\omega_g^h - x_g^h < s_g^h \forall h$ and $y_g^f < s_g^f \forall f$.

That is, quantity constraints affect supply only (3', 4') and are binding only when prices are downward rigid (5).¹¹

3 Walras-Keynes Equilibria

3.1

We are now ready to verify the pervasiveness of supply-constrained equilibria, also at prices compatible with competitive equilibrium.

The set of G goods is partitioned into two subsets, labeled I and II; and all vectors in R^G , like x^h , y^f , or s^j , are partitioned into corresponding subvectors (x_I^h, x_{II}^h) , (y_I^f, y_{II}^f) , (s_I^j, s_{II}^j) . By a slight abuse of notation, the symbols I and II are used to denote both the relevant sets of commodities and their cardinality, so that, for instance, $s_{II}^j \in R^{II}$. *Set II owns at least two elements.*¹² Set I could be empty.

Let the economy $\mathcal{E} = ((X^h, \succeq_h, \omega^h)_{h=1 \dots H}, (y^f)_{f=1 \dots F}, \Theta)$ satisfy the following assumptions, for all h and f :

A.1 $X^h \subset R_+^G$ is non-empty, closed, convex; $X^h + R_+^G \subset X^h$;

A.2 $\succeq_h$ is complete, transitive, continuous, weakly convex; if $x' \geq x$, then $x' \succeq_h x$; if in addition either $x'_I > x_I$ or $x'_{II} > x_{II}$, then $x' \succ_h x$;

A.3 $\omega^h \in \text{int} X^h$;

A.4 $Y^f \subset R^G$ is closed and convex; $0 \in Y^f$;

A.5 $Y := \sum_f Y^f$ is closed; $Y \cap (-Y) = \{0\}$; $R^G \subset Y$.

Define $\bar{P} = \{p \in R_+^G | p_{II} = \bar{p}_{II}\}$ and consider the economy $\mathcal{E}(\bar{P})$. The fixed prices $\bar{p}_{II}$ induce a scale normalization. A supply-constrained equilibrium for $\mathcal{E}(\bar{P})$ is a tuple (x, y, p, s) with (x, y) feasible and $p \in \bar{P}$, such that (3'), (4'), and (5) are satisfied. Note that, for $\bar{P}$ as defined, conditions (5) are automatically satisfied if, $\forall j = 1 \dots H + F$, $s^j = (s_I^j, s_{II}^j)$ with $s_g^j = +\infty \forall g \in I$, so that s is fully defined by $s_{II} = (\dots s_{II}^j \dots)_{j=1 \dots H+F} \in R_+^{II(H+F)}$.

3.2

THEOREM: Let $\mathcal{E}(\bar{P})$ be as defined in section 3.1, that is $\bar{P} = \{p \in R_+^G | p_{II} = \bar{p}_{II}\}$; assume A.1–A.5; let $\bar{s}_{II}$ be an arbitrary vector in $R_+^{II(H+F)}$; there exists a supply-constrained equilibrium $(\hat{x}, \hat{y}, \hat{p}, \hat{s})$ for $\mathcal{E}(\bar{P})$ with $\hat{s}_{II} \leq \bar{s}_{II}$.

Proof The proof is an extension and application of the theorem in Dehez-Drèze (1984), hereafter DD. In that paper, the set of commodities is assumed to include: (1) a specific, always desired commodity 0 with price p_0 constrained to satisfy $0 = p_0 \leq p_0 \leq \bar{p}_0 = \frac{1}{b}$, with $b > 0$; (2) a set of commodities J such that $p_g = p = \bar{p}_g \forall g \in J$, with $\sum_{g \in J} \bar{p}_g = 1$. I will use group II as corresponding to the set J , satisfying $p_g = \bar{p}_g \forall g \in II$, and thus normalize prices by $\sum_{g \in II} \bar{p}_g = 1$. I will apply the DD theorem with $0 = p_g \leq p_g \leq \bar{p}_g = \frac{1}{b} \forall g \in I$ and then let b tend to zero in step 3. The sets $\bar{P}_b = \{p \in R_+^G | 0 \leq p_g \leq \frac{1}{b} \forall g \in I, p_{II} = \bar{p}_{II}\}$ are then special cases of the set of admissible prices in DD. The assumptions A.1–A.5 here are equivalent to assumptions one to nine in DD, except for the strict desirability of good 0 there, replaced here the desirability assumption on groups I and II. In step 2, I shall extend the DD theorem to that different assumption. Prior to that, I show in step 1 that the DD theorem holds with $\hat{s}_{II} \leq \bar{s}_{II}$ for arbitrary $\bar{s}_{II} \in R_+^{II(H+F)}$. These three steps assume that the set of goods I is non empty and includes good 1. When that set is empty, the proof is trivial: $(\hat{x} = \omega, \hat{y} = 0, \hat{p} = \hat{p}_{II}, \hat{s} = 0)$ always verifies the conclusion.

Step 1. Following a remark in Drèze (1975, last paragraph of section 2), I define (for $b > 0$) the *truncated economy* $\mathcal{E}(\bar{P}_b; \bar{s}_{II})$ by the following straightforward truncation of the sets X^h, Y^h :

$$X^h(\bar{s}_{II}) = \{x \in X^h | \omega_{II}^h - x_{II}^h \leq \bar{s}_{II}^h\}, Y^f(\bar{s}_{II}) = \{y \in Y^f | y_{II}^f \leq \bar{s}_{II}^{H+f}\}.$$

Because $\bar{s}_{II}^h > 0 \forall h$, the economy $\mathcal{E}(\bar{P}_b; \bar{s}_{II})$ satisfies assumptions A.1–A.5 above. Assume in addition that:

A.6 $x' \geq x$ with $x'_1 > x_1$ implies $x' \succ_h x^h, h = 1 \dots H$.

All the conditions of the DD theorem are satisfied and there exists a supply-constrained equilibrium for $\mathcal{E}(\bar{P}_b; \bar{s}_{II})$, hence for $\mathcal{E}(\bar{P}_b)$ with $s_{II} \leq \bar{s}_{II}$.

Step 2. Assumption A.6 is used in the proof of the DD theorem only once, to show that $p_1 > 0$; otherwise, unbounded demand for good 1 would prevail. And $p_1 > 0$ implies $p_1 \bar{s}_1^h > 0 \forall h$, hence continuity of the budget set correspondences; see the lemma in Drèze (1975). Using A.3, the same argument implies $\sum_{g \in I} p_g > 0$; otherwise, unbounded demand for some good in I would prevail. Also, $\sum_{g \in I} p_g > 0$ implies $p_i \bar{s}_i^h > 0 \forall h$. Accordingly, under A.1–A.5, there exists a supply-constrained equilibrium for $\mathcal{E}(\bar{P}_b)$ with $s_{II} \leq \bar{s}_{II}$.

Step 3. Consider finally a sequence $b^v \rightarrow 0$. For each v there exists an equilibrium $(\hat{x}^v, \hat{y}^v, \hat{p}^v, \hat{s}^v)$ with $\hat{s}_{II}^v \leq \hat{s}_{II}$. The whole argument in DD is conducted on a compact set of allocations independent of b , $(\hat{x}^v, \hat{y}^v, \hat{s}^v)$ which converge, possibly on a subsequence, to $(\hat{x}, \hat{y}, \hat{s})$. To show that the sequence $\hat{p}^v$ is bounded, assume on the contrary that

$$\|\hat{p}^v\| \rightarrow +\infty; \text{ let } \pi^v := \frac{\hat{p}^v}{\|\hat{p}^v\|} \|\pi^v\| = 1.$$

π^v converges, possibly on a subsequence, to $\hat{\pi}$ with $\|\hat{\pi}\| = 1$, $\hat{\pi}_{II} = 0$. But A.3 then implies that demand for some good in II is unbounded, a contradiction. Hence, $\|\hat{p}^v\|$ is bounded and there exists a sequence converging to $(\hat{x}, \hat{y}, \hat{p}, \hat{s})$, a supply-constrained equilibrium for $\mathcal{E}(\bar{P})$ with $\hat{s}_{II} \leq \bar{s}_{II}$.

3.3

I present my comments on the interpretation of the theorem as seven remarks.

Remark 1 The theorem asserts, as announced, that if some relative prices—those of goods II —are rigid, hence downward rigid, then unlimited underutilization of the associated resources ($\hat{s}_{II} \leq \bar{s}_{II}$ arbitrary) is possible. This property holds even when the rigid prices are compatible with a competitive equilibrium: let p^* be a competitive price for $\mathcal{E}$, and set $\bar{p}_{II} = p_{II}^*$. In that case, I refer to the equilibrium $(\hat{x}, \hat{y}, \hat{p}, \hat{s})$ as a “Walras-Keynes equilibrium.” The allocation $(\hat{x}, \hat{y})$ is sustained by the prices $\hat{p}$ and the quantity constraints $\hat{s}_{II}$. If all the quantity constraints $\hat{s}_{II}$ were simultaneously lifted, the competitive equilibrium (x^*, y^*, p^*) could be attained (say through a suitable adjustment process in quantities and the flexible prices p_I). In that sense, the “unnecessarily inefficient” allocation $(\hat{x}, \hat{y})$ is interpretable as reflecting a pure coordination failure, sustained but not caused by the price rigidities.

Remark 2 The theorem reveals that the interaction of coordination failures and price rigidities has more dramatic consequences than either of these features considered in isolation. It is not unusual that the interaction of distortions deepens their

consequences. The theorem is illustrative of that possibility, but does not exhaust it, because it applies only to a specific form of price rigidities.¹³ An interesting variation, not developed here for lack of space, establishes a comparable result for the case where $P = \{p \in R^G | \sum_{g \in I} p_g = 1, 0 < \bar{p}_{II} \leq p_{II}\}$.

Remark 3 The conclusion of the theorem holds as a special case in the simplified Roberts economy of section 1.2. Just let goods II correspond to inputs supplied by the households and goods I correspond to the consumer goods (and intermediate inputs); set $\bar{p}_{II} = p_{II}^*$.¹⁴ The special assumptions (constant returns, homotheticity) are thus unnecessary for the existence of Walras-Keynes equilibria, which holds generally under assumptions A.1–A.5. The special assumptions do, however, entail two special features, which lack generality: $\hat{p}_I = p_I^*$ (so that $\hat{p} = p^*$); and $\forall h, \hat{s}_{II}^h = \lambda \omega_{II}^h \leq \bar{s}_{II}^h$. I comment on these features in remarks 5 and 6. They illustrate the general property that Walras-Keynes equilibria have specific properties, to be spelled out under specific assumptions. It is not so that *any* allocation may obtain.

Remark 4 Readers not familiar with the earlier work on supply-constrained equilibria might find it counterintuitive that equilibria with arbitrarily severe underutilization of rigidly priced resources can be sustained without demand rationing. The intuitive reason is that demand for specific goods can always be curtailed by reducing purchasing power, namely the supply of other goods (in type II) or their value (in type I). After all, that is also the rationale for combatting demand-pull inflation through contractionary policies.

Remark 5 It is important to understand precisely the implications of fixing the prices $\hat{p}_{II} = \bar{p}_{II}$, while keeping $\hat{p}_I$ flexible. It means that *relative prices within group II are fixed*, while the *overall level of prices $\hat{p}_{II}$ relative to prices $\hat{p}_I$ is flexible*. In particular if a numeraire commodity 1 is chosen in group I, the prices for group II in terms of the numeraire, $\hat{p}_{II}/\hat{p}_I$, are scale-free (yet relatively fixed inside group II).¹⁵ In particular, let wages belong to group II but a composite CPI-commodity belong to group I; the wage hierarchy is then fixed, but real consumption-wages are flexible. It may then (but need not) be the case that unemployment equilibria set in only for high-enough real consumption wages. This also illustrates the point that only in special circumstances (like the Roberts economy) will $\hat{p} = p^*$ (given $\bar{p}_{II} = p_{II}^*$). In general, $\hat{p}_I$ adjusts to clear the markets for goods I, given the constraints $\hat{s}_{II} \leq \bar{s}_{II}$ on the supply of goods II.¹⁶

Remark 6 In the Roberts economy of section 1.2, there exists a *continuum* of supply-constrained allocations $(\lambda x^*, \lambda y^*)$, $\lambda \in [0, 1]$. These can be used to verify the conclusion of the theorem, by selecting

$$\hat{\lambda} \leq \bar{\lambda} = \min_{h, \ell \in I} \frac{\hat{s}_I^h}{\omega_I^h}.$$

Note already the existence of distinct Walras-Keynes equilibria for distinct $\hat{\lambda}s \in [0, \bar{\lambda}]$. But there may exist many more, as already noted in the early literature on fixed-price equilibria. For instance, let h and i be two identical households, and let $\hat{s}_{II}^h = \lambda^h \omega_{II}^h$, $\hat{s}_{II}^i = \lambda^i \omega_{II}^i$; setting $\lambda^h \neq \lambda^i$ with $\lambda^h + \lambda^i = 2\lambda$ will generate (trivially) a continuum of distinct equilibria. An interesting question for macroeconomics is whether there always exists a connected set of Walras-Keynes equilibria including a competitive equilibrium and underemployment equilibria (as in the Roberts economy of section 1.2).¹⁷ Assumptions A.1–A.5 do not imply that property. The appendix of Drèze (1997) exhibits a simple example of an exchange economy where constraining *however slightly* the supply of one particular good reduces the set of supply-constrained equilibria to no trade at all. I shall refer to that extreme example as “the dismal example.” Additional assumptions are needed to prove existence of a connected set of distinct equilibria, including a competitive one. Herings and Drèze (1997) investigate that topic.

Remark 7 The assumptions and wording of the theorem have been chosen with the objective of keeping the proof simple—while leaving technical refinements open for further work, in particular Herings and Drèze (1997). The more restrictive assumption is A.3, for which alternatives like “resource relatedness” exist. The detour did not seem justified here. In the wording of the theorem, I restrict $\bar{s}_{II}$ to lie in $R_{++}^{II(H+F)}$, whereas $R_+^{II(H+F)}$ would suffice; and I do not state the additional property that $\hat{s}_{II} \prec \bar{s}_{II}$. These two extensions require a different method of proof that did not seem to pass a cost-benefit test.

3.4

I turn now to an important remark about the domain of application of the theorem. The economy $\mathcal{E}$ is a standard Arrow-Debreu economy. It is thus susceptible to the same interpretations, covering in particular *economies extending over time under uncertainty*. It is well known that such an interpretation is obtained by specifying, for each commodity, not only a time and place of delivery, but a “state of the environment” on which the delivery is contingent; see Arrow (1953), Debreu (1959, chapter 7) or Radner (1982). That interpretation covers, for instance, the “real business cycles” model as a special case. My theorem thus reveals that coordination failures are a pervasive possibility also in that framework, once the multiplicity of agents is recognized.¹⁸

For my purposes here, the more relevant interpretation of the Arrow-Debreu model with time and uncertainty is that proposed by Radner (1972), namely “equilibrium of plans, prices and price expectations in a sequence of markets.” Under that interpretation, one considers an economy starting at some date in real time, with spot markets for commodities and assets and with some (a few) financial markets (i.e., markets for futures, options and insurance contracts). It is common knowledge that other markets will be opened sequentially at future dates. The agents transacting at the starting date form expectations about the prices that will prevail on these markets, and plan to transact on them when they open. The simplest equilibrium concept used by Radner assumes “perfect foresight,” namely: (1) that price expectations be *common* to all agents; (2) that planned transactions be *consistent*, in the sense of equality of (planned) supply and demand on the prospective market for each commodity or asset at each date under each state of the environment at that date (i.e., for each “date-event” pair).¹⁹ If, at each date-event, the asset markets and financial markets to be opened then would enable agents to transfer wealth across all the events at the following date, one says that markets are *complete*. Under perfect foresight and complete markets, a competitive equilibrium for the Arrow-Debreu economy can be implemented as an equilibrium of plans, prices and price expectations.

My theorem is directly applicable to an economy with perfect foresight and complete markets. It then asserts that “*under rigidity of some prices and price expectations, an unlimited underutilization, current or anticipated, of the associated resources is possible, even when the rigid prices and price expectations are compatible with a competitive equilibrium.*” In particular, the coordination failure could take the form of *supply constraints expected in the future and spilling over to current markets*. Or it could take the form of supply constraints experienced today, as a legacy of the past or as a consequence of a current surprise (like an oil shock or the Gulf War), and accompanied by expectations of future similar constraints, resulting together in current and anticipated underutilization of resources.

This application to time and uncertainty calls for two observations. First, the possibility of supply constraints is relevant to the formation of expectations. Expectations of agents no longer bear on prices alone but on quantity constraints as well.²⁰ Their domain has to be extended. And the possibility of multiple equilibria, or even of a continuum of equilibria, poses a serious problem of forming expectations and especially of achieving common expectations. That difficulty is well known. Second, both the requirement of perfect foresight and the requirement of complete markets are problematic in their own rights. It would be of obvious desirability to extend my

theorem to a more realistic framework, like “general equilibrium with incomplete markets” (GEI) or “temporary general equilibrium” (TGE).²¹ For instance, I have verified on a standard GEI model that market incompleteness affects the dimensionality of the set of Walras-Keynes equilibria.²²

4 Speculative Implications

The theoretical results of section 3 invite the conclusion that coordination-failure unemployment is a logical possibility, under downward-rigid (though not necessarily “wrong”) prices or wages. That possibility carries some implications with which macroeconomists should be prepared to live, even if it complicates their task some more.

1. *The very possibility of coordination failures precludes drawing simple inferences from the observation of excess supply to price distortions.* Indeed, quantities (current and expected) may be far away from those associated with a competitive equilibrium, with prices (current and expected) either grossly distorted, or again compatible with competitive equilibrium. It would take modeling and computations to tell. In some cases, like for unskilled wages in Europe, that is possible. In other cases, like mid-scale wages in Europe, the task is much more difficult; the answer may depend upon modeling options, or remain uncertain due to econometric as well as modelling uncertainties.

2. *Coordination problems may magnify the consequences of price rigidities.* We tend to think about real rigidities as keeping prices and wages at levels inconsistent with attainment of a first-best allocation, mostly due to market power. The consequences of these rigidities (distortions) are then viewed as proportional to the distance at which prices stand from market clearing levels, and to elasticities of supply and/or demand. Under coordination failures, the price distortion may be modest—in the limit nonexistent—yet the inefficiencies sustained by the rigid prices may be substantial, arbitrarily severe according to the theorem. This point is well illustrated by the insider-outsider theory of unemployment; see Lindbeck and Snower (1988), Lindbeck (1993). Insider power, rooted in hiring and firing costs or threats of non-cooperation, can keep wages above the reservation level of outsiders, by a bounded amount (say the sum of hiring and firing costs). Offhand, one might think that the resulting damage will be bounded as well, by the relevant area under the demand curve for labor. But downward wage rigidity rooted in insider power can also *sustain* (not cause) severe damage, if the employment level is depressed by coordination failures. The consequences are thus magnified.

3. *The possibility of coordination failures places the issue of price and wage flexibility in an altogether different setting.* We think naturally of flexibility as the means to avoid or eliminate wasteful price distortions. We should in addition ask whether and how it contributes to overcoming coordination failures. That is a different question altogether. And it is not obvious at all that price or wage adjustments susceptible of removing inefficiencies caused by price distortions would also operate in the right direction, or with any effectiveness, to circumvent coordination failures.

4. *Coordination-failure equilibria are inherently multiple and unstable.* In the statement of the theorem, the reference levels $\bar{s}_H^h$ have no intrinsic significance; they are simply a way of formulating the result that arbitrary degrees of resource underutilization are compatible with arbitrary prices. The theorem says nothing about how Walras-Keynes equilibria come about. It implicitly announces that they are unstable. If they were displaced, a new equilibrium with similar features could come about, sustained by the same prices $\hat{p}_H = \bar{p}_H$ and a new set of supply constraints $\hat{s}'_H$. The movement from one supply-constrained equilibrium to another is a topic in dynamics, inviting the study of adjustment processes defined over prices (definitely $\hat{p}_I$, but perhaps also $\hat{p}_H$), quantities, price expectations, and plans.²³

5 Wage and Price Heuristics

5.1

Recognizing the possibility that unemployment may reflect, at least in part, a coordination failure places the issue of wage flexibility in a new setting. The starting point is that, *irrespective of the level of wages*, their very rigidity may sustain (the persistence of) arbitrarily severe rates of unemployment. This is a strong argument, to my knowledge a new one, in favor of wage flexibility, both downward and upward. The upward aspect is important. It means in particular that when flexible wages rise together with employment, the associated rise in the price level should not automatically trigger restrictive anti-inflation policies. For if that were the case, downward wage flexibility would become a one-way process, ill suited to overcome coordination failures.

There are, however, conflicting considerations. Wage flexibility need not be a suitable instrument to overcome coordination failures: first, it may not be effective; second, it may not be efficient; third, it may not be perceived as fair.

1. *Effectiveness.* In simple illustrations (like section 4.3 of Drèze 1997), lower product wages induce only a transfer of income from wage earners to profit earners,

possibly compounded by lower employment. More generally, the theorem asserts that a given level of underutilisation of resources (for the commodities in group II) may be reproduced, or even aggravated, at alternative prices $p_{II} \leq \bar{p}_{II}$. (Of course, one should realize that the prices $\hat{p}_I$ will also change, so that $p_{II}/\hat{p}_I$ need not fall.) The only safe conclusion is that effectiveness of wage flexibility remains to be demonstrated, in a suitable dynamic formulation.

We know from the literature on stability²⁴ that special conditions are required to obtain convergence of price-guided adjustment processes, especially of the more realistic non-tâtonnement processes. In that literature, however, there is no doubt that some kind of price adjustments is desirable, if the process starts at prices incompatible with an efficient equilibrium. The situation is altogether different at a Walras-Keynes equilibrium. Here one knows that price adjustments could possibly be dispensed with, if another approach toward overcoming the coordination failure were accessible; one knows that price adjustments may have to "cycle," returning eventually to the starting point; and one knows that intermediate steps in the wrong direction could do persistent harm, if the economy settled halfway at a supply-constrained equilibrium inferior to the starting point.

2. *Efficiency.* This consideration is related to risk sharing. When output and labor demand are subject to random fluctuations, and when not all workers are engaged in long-term labor contracts (in particular because some will enter or reenter the labor force at a future date), it is not *ex ante* efficient to let the wages applicable to new hirings in the future adjust flexibly to clear the labor market at each date-event; this would lead to inefficient risk sharing between those hired in the future and the other agents (workers under contract and equity-owners); a second-best efficient solution calls for keeping the compensation of newly hired workers below the market clearing wages, by means of labor taxes, in states that are exceedingly favorable to workers; and for keeping that compensation above the market clearing wages, by means of unemployment benefits, in states exceedingly unfavorable to workers; that second-best solution strikes a proper balance between the loss of output due to unemployment and the gains in risk sharing due to reduced wage flexibility; see Drèze-Gollier (1993).

Although that analysis is conducted in a framework that does not explicitly recognize coordination failures, its logic remains valid, as it does not hinge on specificities of the output and employment fluctuations; and *its relevance is magnified by the possibility of coordination failures and the associated volatility.* However, the second-best analysis assigns an important role to the wage-elasticity of labor demand in defining the desirable level of unemployment benefits and hence of downward wage rigidity. In the presence of coordination failures, that elasticity

remains to be defined. It is not a property of the static (unstable) equilibrium; again, we must wait for a dynamic formulation.

3. *Fairness.* Downward wage flexibility has an immediate impact on the incomes of employed workers. By their very nature, coordination failures are a recurrent threat. If overcome today, they may reappear tomorrow, as a consequence of the next shock (to expectations?). Relying recurrently on wage flexibility to overcome coordination failures might in bad times bring wages all the way down to reservation levels, at uncertain gains in employment.

I thus see three cogent reasons why organized labor would resist the idea of fighting coordination-failure unemployment through wage flexibility:

1. uncertain effectiveness, *specific* to the context of coordination failures;
2. inefficient risk sharing, *compounded* by the volatility of employment associated with coordination failures;
3. adverse redistributive transfers, *compounded* by the potential recurrence of coordination failures.

These three reasons fit naturally into the objective function of a union, concerned with the expected utility of its members (present and perhaps also potential or future).²⁵ The insider-outsider theory explains the source and limits of the market power of insiders or unions. This closes the circle: recognition of coordination failures as a source of persistence of unemployment should also lead us to expect persistent resistance of real wages, where and when organized labor exerts a significant influence on wage formation. The three reasons above also suggest why such an influence may be needed to cope with macroeconomic volatility. And yet, coordination failures magnify the consequences of wage rigidities. Food for thought.

5.2

A related question—why the unemployed do not undercut the employed—will not detain me because I do not see how its terms are affected by the *specific* dimension of coordination failure. The fact that undercutting rarely happens is well documented. One explanation has been alluded to above, because it is advocated in Roberts (1987); namely that price competition between insiders and outsiders could only drive the wage down to the reservation level of outsiders, making the attempt at undercutting worthless. Hahn and Solow (1995, section 5.3) take that argument one step further by embedding the situation in a repeated game and obtaining absence of undercutting as a subgame perfect equilibrium.²⁶ It is thus not surpris-

ing that unions penalize such behavior, as predicted by the “noncooperation” version of the insider-outsider theory; see Lindbeck and Snower (1988). And it is instructive to read the survey evidence in Bewley and Brainard (1993) to the effect that firms shun cutting wages, out of concern for long-run labor relations and productivity.

5.3

Turning to prices, I would be generally happy to leave them alone and to assume clearing output markets. But downward rigidity of output prices in the presence of excess capacities invites a brief comment. Excess capacity means both capacity, hence fixed factors (in the short run), and prices above marginal costs. Constant returns and competitive pricing, as in the Roberts and RBC models, sidestep the issue. Imperfect competition offers one explanation.²⁷ It may be useful to point out that perfect competition and incomplete markets provide a parallel explanation. Investment in productive capacity generates fixed costs. Under incomplete financial markets, these must be covered by surviving firms at all date-event pairs; this will typically require competitive firms to practice average cost pricing in states of depressed demand, thus explaining the downward rigidity of prices in the presence of excess capacity. (The feasibility of average cost pricing would be built into the investment decisions, under correct expectations.²⁸)

In a complex economy with many firms, some of these may operate with locally diminishing returns, others with locally increasing returns. The prices which they set may react in opposite directions to changes in capacity utilization. Over a range, the aggregate price level may then be quite insensitive to excess capacities, a feature not devoid of econometric validation.²⁹

One set of prices is apt to play a very important role in overcoming coordination failures, namely interest rates. Indeed, encouraging (the advancement of) investment, including public investment and direct investment by households, and encouraging (the advancement of) consumption seem natural to overcome an aggregate demand externality. Low real interest rates should help. It is beyond the scope of this progress report to discuss interest rates.

6 Summary

This chapter can be summarized as follows:

1. In market economies with decentralized price formation, supply-constrained equilibria may persist as coordination failures, sustained but not caused by price

rigidities. This feature may be arbitrarily severe, even at prices compatible with a competitive equilibrium (Walrasian fixed relative prices for goods subject to quantity constraints, flexible prices for the remaining goods).

2. In an economy extending over time with uncertainty, under downward rigidity of some relative prices and price expectations, unlimited underutilization, current or anticipated, of the associated resources is possible, even when the rigid prices and price expectations are compatible with a competitive equilibrium. This property holds under perfect foresight and complete markets; it is complementary to the additional coordination problems raised by the formation of expectations and the incompleteness of markets.

3. Consequently:

- one cannot infer the existence or nature of price distortions simply from the observation of excess supply;
- the consequences of price and wage rigidities are magnified by coordination failures;
- the merits of price flexibility must be assessed also in terms of ability to overcome coordination failures; this calls for a specification of dynamic adjustment;
- equilibria are multiple, unstable and path-dependent—though far from arbitrary.³⁰

4. (Heuristics.) Fighting coordination-failure unemployment through real wage flexibility has three specific drawbacks for wage earners:

- uncertain effectiveness, specific to the context of coordination failures;
- inefficient risk sharing, compounded by the volatility of employment associated with coordination failures;
- adverse redistributive transfers, compounded by the potential recurrence of coordination failures.

It is therefore likely that organized labor, endowed with insider power, would resist wage flexibility even more strongly in the presence of coordination failures, in spite of the fact that the wage rigidity sustains these failures.

An inescapable (though not new) consequence of these propositions is that we should brace ourselves to deal with multiple, unstable, path-dependent equilibria. This is not a pleasant prospect—whether for theory, econometrics, or policy. Explicit dynamics may brighten the prospect; they are needed in their own right, also with regard to expectations.

Notes

This is an abridged version of a progress report on my current research in theoretical macroeconomics. The report was undertaken as a tribute to János Kornai's long-standing interest in economic systems. As the report took form, it exceeded the format of this volume. The full version, referred to as Drèze (1997), is appearing in *European Economic Review*, which kindly granted permission for the present publication. Drèze (1997) contains a section devoted to applications to a "real business cycle" model and to the three-good model of Barro-Grossman-Malinvand. I am grateful to Pierre Dehez, Frank Hahn, Jean-Jacques Herings, Assar Lindbeck, Edmond Malinvaud, Erik Maskin, Enrico Minelli, Heracles Polemar-chakis, Robert Solow, and two anonymous referees for helpful discussions or suggestions.

1. See, for example, Drèze and Bean (1990) for supportive evidence.
2. In market games, multiplicity is pervasive when labor services are agentspecific.
3. The possibility of such equilibria was first brought out by Hahn (1977, 1978).
4. My presentation does not do justice to the extensive game theoretic analysis of Roberts, to whose papers I refer readers; see also Jones and Manuelli (1992) for an alternative formulation, with significantly different implications.
5. Roberts does not assume homothetic preferences, but generates homotheticity by replicating the same basic economy several times.
6. A key assumption to that effect is that workers do not consume goods produced by the firm that employs them.
7. About the Roberts model, it should also be noted that given constant returns and competitive output markets, the only consequence of lowering real wages *across the board* is to induce substitution of labor for other inputs, and of more labor-intensive consumer goods for less intensive ones. In the absence of other inputs, lower nominal wages lead to lower prices, at unchanged real wages and employment: the overall level of real wages is then set by technology, not by market clearing.
8. Competitive price-setting by firms is captured by the simple condition that prices belong to the marginal cost correspondence (the normal cone to the production set); see Dehez-Drèze (1988a), also for an extension to nonconvex technologies. Noncompetitive, for instance monopolistic or oligopolistic, prices carry similar implications, as will become clear later; see also Roberts (1988).
9. I discuss the link with union wages and insider-outsider theories in section 6.
10. Vector inequalities obey: $x \leq x'$ if $x_g \leq x'_g \forall g$; $x \leq x'$, if $x \leq x'$ and $x \neq x'$; $x < x'$ if $x_g < x'_g \forall g$. $R_+^G = \{x \in R^G \mid x \geq 0\}$, $R_{++}^G = \{x \in R^G \mid x > 0\}$.
11. A supply-constrained equilibrium for $\mathcal{E}(P)$ with at least one good unconstrained may be referred to as a "van der Laan equilibrium," because the concept was introduced and a general existence result was proven, for pure exchange economies under standard assumptions, in van der Laan (1982, 1984). The extension of the existence results to economies with production is straightforward, as can be seen in Dehez and Drèze (1984).
12. Otherwise fixing p_H would boil down to a scale normalization.
13. See Drèze (1997, section 4.5) for an exemple with a form of price rigidities not covered by the theorem.
14. This Roberts economy does not satisfy A.3, but the conclusion is verified constructively. The same remark applies to the economy studied by Jones and Manuelli (1992). The reason why these authors conclude to absence of rationing is different—namely, that prices and wages are set without constraints by firms that implicitly coordinate activities.
15. See Drèze (1997, section 4.4) for illustration.
16. Thus $(\hat{x}, \hat{y}, \hat{p})$ is a competitive equilibrium for the truncated economy $\mathcal{E}(\hat{s}_H)$ defined by $X^h(\hat{s}_H) = \{x \in X^h \mid \omega_H^h - x_H^h \leq \hat{s}_H^h\}$, $Y^f(\hat{s}_H) = \{y \in Y^f \mid y_H^f \leq \hat{s}_H^{H,f}\}$; but not necessarily for $\mathcal{E}(\bar{s}_H)$ similarly defined.
17. Silvestre (1980) provides further examples with that property.

18. See Drèze (1997, section 4.1) for illustration.
19. "An equilibrium of plans, prices and price expectations is a set of prices on the current markets, a set of common expectations for the future, and a consistent set of individual plans, one for each agent, such that, given the current prices and price expectations, each individual agent's plan is optimal for him, subject to an appropriate sequence of budget constraints" (Radner 1982, p. 932).
20. Expectations about future constraints might also take the form of expectations about aggregate demand, as in chapter 6 of Hahn-Solow (1995).
21. See, for example, Geanakoplos (1990) or Magill and Shafer (1991) for surveys of GEI, Grandmont (1982, 1988) for surveys of TGE, and Drèze (1993) for a comparison of the two approaches.
22. The example concerns a two-period economy with S future states and J assets. The dimensionality of the set of equilibria is then $S + 1 - J$, and increases with the degree of market incompleteness $S - J$.
23. In Drèze (1990) I present a simple non-tâtonnement process in quantities and upward flexible prices leading in a *bounded number of steps* to a neighborhood of a supply-constrained equilibrium, under admittedly restrictive assumptions: constant returns, no joint production, and no inferior goods. (These assumptions are automatically satisfied in aggregate single output models with constant returns.)
24. As surveyed, for instance, by Hahn (1982).
25. See Oswald (1985) or Pencavel (1985).
26. A lively illustration is reported by Jean Drèze for the Indian village of Palanpur, where he observed the following: during a year of substantial unemployment among agricultural day laborers, a hard-pressed unemployed laborer offered his services for half the customary wage and was hired. The very next day, the wage of all day laborers was cut by half, at unchanged employment, and the undercutter became very unpopular. The following year, the customary wage was restored, there was substantial unemployment again, but none of the unemployed tried undercutting.
27. See Bénassy (1995) for a collection of relevant papers, with a useful introduction.
28. The general equilibrium consistency of average cost pricing by competitive firms is established in Dehez-Drèze (1988a,b).
29. See, for instance, Drèze and Bean (1990, p. 20).

References

- Arrow, K. J. (1953). Le rôle des valeurs boursières pour la répartition la meilleure des risques. *Econométrie* 41–47, CNRS, Paris. Translated as The Role of Securities in the Optimal Allocation of Risk-Bearing. *Review of Economic Studies* 31: 91–96.
- Bénassy, J. P. (1995). *Macroeconomics and Imperfect Competition*. Aldershot: Elgar.
- Benhabib, J., and A. Rustichini (1994). Introduction to the symposium on growth, fluctuations and sunspots: confronting the data. *Journal of Economic Theory* 63: 1–18.
- Bewley, T., and W. Brainard (1993). *A Depressed Labour Market, as Explained by Participants*. Mimeo, Yale University, New Haven.
- Bryant, J. (1983). A simple rational-expectations Keynes-type model. *Quarterly Journal of Economics* XCVIII: 525–529.
- Citanna, A., H. Crès, and A. Villanacci (1995). Underemployment of resources and self-confirming beliefs. CARESS Working Paper 95-20, University of Pennsylvania, Philadelphia.
- Chiappori, P. A., and R. Guesnerie (1991). Sunspot equilibria in sequential markets models. In W. Hildenbrand and H. Sonnenschein, eds., *Handbook of Mathematical Economics*, Vol. 4. Amsterdam: Elsevier, 1683–1762.
- Cooper, R., and A. John (1988). Coordinating coordination failures in Keynesian models. *The Quarterly Journal of Economics* CIII: 441–463.

- Debreu, G. (1959). *Theory of Value*. New York: Wiley.
- Dehez, P., and J. H. Drèze (1984). On supply-constrained equilibria. *Journal of Economic Theory* 33(1): 172–182. Reprinted in Drèze (1991).
- Dehez, P., and J. H. Drèze (1988a). Competitive equilibria with quantity-taking producers and increasing returns to scale. *Journal of Mathematical Economics* 17: 209–230. Reprinted in Drèze (1991).
- Dehez, P., and J. H. Drèze (1988b). Distributive production sets and equilibria with increasing returns. *Journal of Mathematical Economics* 17: 231–248.
- Drèze, J. H. (1975). Existence of an exchange equilibrium under price rigidities. *International Economic Review* 16(2): 301–320. Reprinted in Drèze (1991).
- Drèze, J. H. (1990). Stability of a Keynesian adjustment process. In W. A. Barnett, B. Cornet, C. d'Aspremont, and A. Mas-Colell, eds., *Equilibrium Theory and Applications*. Cambridge: Cambridge University Press, 197–231. Reprinted in Drèze (1991).
- Drèze, J. H. (1991). *Underemployment Equilibria: Essays in Theory, Econometrics and Policy*. Cambridge: Cambridge University Press.
- Drèze, J. H. (1993). The formulation of uncertainty: prices and states. CORE Discussion Paper 9347, Louvain-la-Neuve.
- Drèze, J. H. (1995). *Pour l'emploi, la croissance et l'Europe*. Louvain-la-Neuve: De Boeck Université.
- Drèze, J. H. (1996). Employment in Europe. In J. Gual, ed., *The Social Challenge of Job Creation: Combating Unemployment in Europe*. Aldershot: Elgar.
- Drèze, J. H. (1997). Walras-Keynes equilibria, coordination and macroeconomics. *European Economic Review* 41: 1732–1762.
- Drèze, J. H., and C. R. Bean, with J. P. Lambert, F. Mehta, and H. Sneessens, eds. (1990). *Europe's Unemployment Problem*. Cambridge, Mass.: MIT Press.
- Drèze, J. H., and C. Gollier (1993). Risk sharing on the labour market and second-best wage rigidities. *European Economic Review* 37: 1457–1482.
- Geanakoplos, J. (1990). An introduction to general equilibrium with incomplete markets. *Journal of Mathematical Economics* 19: 1–38.
- Geanakoplos, J., and H. Polemarchakis (1991). Overlapping generations. In W. Hildenbrand and H. Sonnenschen, eds., *Handbook of Mathematical Economics* Vol. IV. Amsterdam: North-Holland, 1899–1960.
- Giraud, G. (1993). Existence in a Non-Cooperative Market Game. Mimeo, Université Paris I, Paris.
- Grandmont, J. M. (1982). Temporary general equilibrium theory. In K. J. Arrow and M. D. Intriligator, eds., *Handbook of Mathematical Economics*, Vol. 2. Amsterdam: North-Holland, 779–922.
- Grandmont, J. M. (1988). *Temporary Equilibrium: Selected Readings*. San Diego: Academic Press.
- Hahn, F. H. (1977). Unsatisfactory Equilibria. Technical Report 247, IMSSS, Stanford.
- Hahn, F. H. (1978). On non-Walrasian equilibria. *Review of Economic Studies* XLV: 1–17.
- Hahn, F. H. (1982). Stability. In K. J. Arrow and M. D. Intriligator, eds., *Handbook of Mathematical Economics*. Amsterdam: North-Holland, 746–793.
- Hahn, F., and R. Solow (1995). *A Critical Essay on Modern Macroeconomic Theory*. Cambridge, Mass.: MIT Press.
- Herings, J. J. (1996). *Static and Dynamic Aspects of General Disequilibrium Theory*. Boston: Kluwer.
- Herings, J. J., and J. H. Drèze (1997). Continua of Underemployment Equilibria Reflecting Coordination Failures also at Walrasian Prices. Working Paper, Tilburg and Louvain-la-Neuve.
- Jones, L. E., and R. E. Manuelli (1992). The coordination problem and equilibrium theories of recessions. *American Economic Review* 82(3): 451–471.
- Kornai, J. (1971). *Anti-Equilibrium*. Amsterdam: North-Holland.
- Kornai, J. (1980). *Economics of Shortage*. Amsterdam: North-Holland.

- Kornai, J. (1990). The affinity between ownership forms and coordination mechanisms. *Journal of Economic Perspectives* 4(3): 131–147.
- Lindbeck, A. (1993). *Unemployment and Macroeconomics*. Cambridge, Mass.: MIT Press.
- Lindbeck, A., and D. Snower (1988). *The Insider-Outsider Theory of Employment and Unemployment*. Cambridge, Mass.: MIT Press.
- Magill, M., and W. Shafer (1991). Incomplete markets. In W. Hildenbrand and H. Sonnenschen, eds., *Handbook of Mathematical Economics*. Amsterdam: Elsevier, 1523–1614.
- Mas-Colell, A. (1991). Cournotian foundations of Walrasian equilibrium theory. In W. Hildenbrand, ed., *Advances in Economic Theory*. Cambridge: Cambridge University Press.
- Morishima, M. (1992). *Capital and Credit*. Cambridge: Cambridge University Press.
- Oswald, A. (1985). The economic theory of trade unions: an introductory survey. *Scandinavian Journal of Economics* 82(2): 160–193.
- Peck, J., K. Shell, and S. E. Spear (1992). The market game: existence and structure of equilibrium. *Journal of Mathematical Economics* 21: 271–299.
- Pencavel, J. (1985). Wages and employment and trade unionism: microeconomic models and macroeconomic applications. *Scandinavian Journal of Economics* 82(2): 197–225.
- Radner, R. (1972). Existence of equilibrium of plans, prices and price expectations in a sequence of markets. *Econometrica* 40: 289–303.
- Radner, R. (1982). Equilibrium under uncertainty. In K. J. Arrow and M. D. Intriligator, eds., *Handbook of Mathematical Economics*. Amsterdam: North-Holland, 923–1006.
- Roberts, J. (1987). An equilibrium model with involuntary unemployment at flexible, competitive prices and wages. *American Economic Review* 77: 856–874.
- Roberts, J. (1988). Involuntary unemployment and imperfect competition: a game theoretic macro model. In G. Feiwel, ed., *The Economics of Imperfect Competition and Employment: Joan Robinson and Beyond*, London: Macmillan.
- Roberts, J. (1989). Equilibrium without market clearing. In B. Cornet and H. Tulkens, eds., *Contributions to Operations Research and Economics*. Cambridge, Mass.: MIT Press, 145–158.
- Silvestre, J. (1980). Continua of Hahn-unsatisfactory equilibria. *Economic Letters* 5(3): 201–208.
- van der Laan, G. (1980). Equilibrium under rigid prices with compensation for the consumers. *International Economic Review* 21: 63–73.
- van der Laan, G. (1982). Simplicial approximation of unemployment equilibria. *Journal of Mathematical Economics* 9: 83–97.
- van der Laan, G. (1984). Supply-constrained fixed price equilibria in monetary economies. *Journal of Mathematical Economics* 13(2): 171–187.

3 Viable Domains in the Control Space

Béla Martos

Introduction

In a joint *Econometrica* paper with János (Kornai and Martos 1973), we used the term *viability of an economic control system* for a property stronger than stability, but looser than optimality. This concept retained its central role also in our jointly edited volume *Non-Price Control*. In the course of my subsequent mathematical research on this line, the need for some constructive approach to viable control systems emerged beyond the mere existence results of viability theory. This is how Kornai's economic ideas inspired a pure mathematical study whose latest results are presented here.

Motivation

In controlled systems it is a frequent occurrence that the state trajectory is restricted to a subset of the state space, the viability set (Nagumo 1942; Aubin and Cellina 1983). The known results of viability theory refer to conditions under which the existence of viable trajectories can be decided. In the analysis of linear controlled economic systems (Kornai and Martos 1981, Martos 1990), the necessity emerged to go beyond existence and determine a subset of control trajectories that generate viable state trajectories. The state constraints defining the viability set were given usually in the form of inequalities (e.g., nonnegativity constraints, budget constraints), which implied nonlinearity of the system and prevented the use of standard (e.g., Laplace-transform) techniques. Although the motivation came from economics, it is expected that the present viability results prove to be applicable to physical or biological systems, too.

The Problem Setting

Consider the time-invariant linear inhomogeneous differential equation system:

$$\dot{x}(t) = Ax(t) + u(t), \quad x(0+) = x_0, \quad (3.1)$$

where

$t \in \mathbb{R}_+ := (0, \infty)$, (positive) time

$x : \mathbb{R}_+ \rightarrow \mathbb{R}^n$, state vector

$u : \mathbb{R}_+ \rightarrow \mathbb{R}^n$, control vector

$A \in \mathbb{R}^{n \times n}$, system matrix.

The time functions $t \rightarrow x(t)$, $t \rightarrow u(t)$ will be called *state trajectory* and *control trajectory* and denoted as $x(\cdot)$ and $u(\cdot)$, respectively. A trajectory pair $u(\cdot), x(\cdot)$ is called a *solution* if they jointly satisfy (1) for all $t \in \mathbb{R}_+$, and then we also say that the control trajectory $u(\cdot)$ *generates* the state trajectory $x(\cdot)$.

Let us denote by

$$\begin{aligned}\tilde{x} &:= \{x(t) : t \in \mathbb{R}_+\} \\ \tilde{u} &:= \{u(t) : t \in \mathbb{R}_+\}\end{aligned}\tag{3.2}$$

the *range* of a particular state (control) trajectory, these are subsets of the state (control) space, respectively.

Viability Concepts

Let $X \subset \mathbb{R}^n$ denote a given *proper* subset of the state space, the *viability set*. A state trajectory is said to be *viable* if $x(t) \in X$, $\forall t$, or more concisely: $\tilde{x} \subset X$. A control trajectory $u(\cdot)$ is said to be *viable* if it generates a viable state trajectory. Then we also say that the solution is *viable*.

We give sufficient but not necessary conditions for a solution to be viable. More precisely, we determine $\mathcal{U} \subset \mathbb{R}^n$, a subset of the control space, called *viable control set*, such that any control trajectory whose range is contained in it, is a *viable control trajectory*, that is, it generates viable state trajectory. We are not able to determine all viable solutions, but we try to define a possibly large viable control set.

Notations and Assumptions

Vector and Matrix Norms

For any (real or complex) n -vector $g \in \mathbb{C}^n$ and square matrix $G \in \mathbb{C}^{n \times n}$ let us denote by g^+ and G^+ , respectively, the nonnegative vector and matrix consisting of the absolute values (moduli) of the components or entries of g and G

$$\begin{aligned}g^+ &:= [|g_j|], \quad \forall g = [g_j] \in \mathbb{R}^n \\ G^+ &:= [|G_{i,j}|], \quad \forall G = [G_{i,j}] \in \mathbb{C}^{n \times n}.\end{aligned}\tag{3.3}$$

Let us denote by $\|g\|$ and $\|G\|$ a vector norm and a matrix norm, respectively (see Lancaster 1969). Throughout the paper the following properties will be assumed of these norms:

1. The vector norm is an *absolute* norm, that is:

$$\|g\| = \|g^+\|, \quad \forall g \in \mathbb{C}^n. \quad (3.4)$$

2. The matrix norm is *induced* by the vector norm:

$$\|G\| := \max\{\|Gg\| : \|g\| = 1\}. \quad (3.5)$$

(3.5) implies (see Lancaster 1969, theorem 6.3.1.)

$$\|Gg\| \leq \|G\| \cdot \|g\|, \quad \forall g \in \mathbb{C}^n, \quad \forall G \in \mathbb{C}^{n \times n}. \quad (3.6)$$

For any *diagonal matrix* $H := \text{diag}\{\chi_1, \chi_2, \dots, \chi_n\}$ (3.4) and (3.5) simply ([Lancaster 1969], Th. 6.4.1.):

$$\|H\| = \max_i |\chi_i|. \quad (3.7)$$

The Interval Norm

The *interval norm* or *q-norm* is special. Let $q \in \mathbb{R}_+^n$ be a positive vector and define for any $g \in \mathbb{C}^n$, $G \in \mathbb{C}^{n \times n}$:

$$\|g\|_q := \max_i \frac{|g_i|}{q_i}, \quad (3.8)$$

$$\|G\|_q := \max_i \sum_j \frac{q_j}{q_i} |G_{ij}|. \quad (3.9)$$

It is easy to see that $\|g\|_q$ satisfies (3.4) and $\|G\|_q$ satisfies (3.5). (Moreover, $\|G\|_q$ is an absolute matrix norm since $\|G\| = \|G^+\|$.) The interval norm is nothing else than a scale-transformed l_∞ norm.

The Spectrum of the System Matrix and the Stability of the System

Let us denote by $\lambda_1, \dots, \lambda_n \in \mathbb{C}$ the *eigenvalues* of A , the set of the eigen-values is called the *spectrum* of A and any diagonal matrix $\Lambda \in \mathbb{C}^{n \times n}$ formed from the eigen-values (in different scissions) is called a *spectral matrix* of A : $\Lambda = \text{diag}\{\lambda_1, \dots, \lambda_n\}$. In connection with the spectrum of the system matrix the following two parameters shall be needed in the sequel:

$$\rho := \max_j |\lambda_j| \quad (3.10)$$

the *spectral radius* of A , and

$$\mu := \min_j (-\text{Re} \lambda_j) \quad (3.11)$$

the *degree of stability* of A ($\text{Re} x$ = real part of x).

ASSUMPTION A: The system (1) is *asymptotically stable*, that is,

$$\mu > 0. \quad (3.12)$$

This stability assumption seems to be essential for the subsequent analysis, because it guarantees that the state range is bounded whenever the control range is such (see Zadeh and Desoer 1963).

The Set of Modal Matrices and the Simplicity of A

Let us denote the set of diagonal matrices by $\mathcal{H}$. Let us consider now the (possibly empty) set of such nonsingular matrices (the *modal matrices* of A) that applied as similarity transformation on A result in a diagonal matrix

$$\mathcal{L} := \{L \in C^{n \times n} : L^{-1}AL \in \mathcal{H}\}. \quad (3.13)$$

ASSUMPTION B: The system matrix A is *simple* (i.e., diagonalizable by a similarity transformation)

$$\mathcal{L} \neq \emptyset. \quad (3.14)$$

This simplicity assumption is rather technical. I guess that it can be dispensed with on the account of more complicated formulae.

Under Assumption B any modal matrix $L \in \mathcal{L}$ produces a *spectral decomposition* of A :

$$A = L\Lambda L^{-1}, \quad (3.15)$$

where Λ is a spectral matrix of A .

The Modal Condition Number

Besides the two parameters ρ and μ characterizing A a third, perhaps less known parameter, the *modal condition number* κ , plays important role in the sequel (Lancaster 1969, pp. 222, 232).

$$\kappa := \inf\{\|L\| \cdot \|L^{-1}\| : L \in \mathcal{L}\} \quad (3.16)$$

Because $\|L\| \cdot \|L^{-1}\| \geq \|LL^{-1}\| = \|E\| = 1$, we have

$$\kappa \geq 1 \quad (3.17)$$

under Assumption B.

The numerical value of κ depends also on the choice of the matrix norm. Although the parameters ρ and μ are easy to calculate, it is hard to solve the minimum problem

(3.16) in this generality. But for the interval norm (3.9) and for system matrices of distinct eigenvalues we have the following simple formula:

$$\kappa = \|L^+ K^+\|_q, \quad (3.18)$$

for any $L \in \mathcal{L}$, and $K = L^{-1}$. (This formula is a direct consequence of [Bauer, 1963, p. 80, Theorem II. a.].)

Assumptions on the Viability Set

ASSUMPTION C: The interior X^0 of the viability set is non-empty: $X^0 \neq \emptyset$.

ASSUMPTION D: The initial point x_0 belongs to the interior of X : $x_0 \in X^0$.

It is to be noted that no closedness assumption is made on X at this point, in contrast to the common practice in viability theory.

Clearance, Nest, Nestpoints

For any $d \in \mathbb{R}^n$, $\delta > 0$ let us denote by $\mathcal{B}(d, \delta)$ the *open ball* and by $\overline{\mathcal{B}}(d, \delta)$ the *closed ball* with *center* d and *radius* δ . (The shape of the ball depends on the chosen vector norm.)

$$\mathcal{B}(d, \delta) := \{y \in \mathbb{R}^n : \|y - d\| < \delta\} \quad (3.19)$$

$$\overline{\mathcal{B}}(d, \delta) := \{y \in \mathbb{R}^n : \|y - d\| \leq \delta\}. \quad (3.20)$$

If $d \in X^0$ we can take the supremum of such radii δ that $\mathcal{B}(d, \delta)$ is contained in X , denote it by $\theta(d)$ and call it the *clearance* of d :

$$\theta(d) := \sup\{\delta : \mathcal{B}(d, \delta) \subset X\}. \quad (3.21)$$

If we additionally assume that the viability set is *closed*, then we also use the following equivalent definition:

$$\theta(d) := \max\{\delta : \overline{\mathcal{B}}(d, \delta) \subset X\}. \quad (3.22)$$

The clearance is thus the distance of d from the closest boundary point of X . The larger is $\theta(d)$ the “more interior” is d to X .

The following subset $\mathcal{D}$ of X^0 will be called the *nest*

$$\mathcal{D} := \{d \in X^0 : \kappa \|x_0 - d\| \leq \theta(d)\}, \quad (3.23)$$

that is, d is a *nestpoint* ($d \in \mathcal{D}$) if its κ -fold distance from the initial point x_0 does not exceed its clearance. The nest is non-empty because $x_0 \in \mathcal{D}$.

Sufficient Viability Conditions

A Lemma

Before wording the basic results of the present paper we establish this lemma.

LEMMA 1. For any $d \in \mathbb{R}^n$

$$\|x(t) - d\| < \theta(d), \quad \forall t \Rightarrow \tilde{x} \subset \mathcal{B}(d, \theta(d)) \Rightarrow \tilde{x} \subset \mathcal{X}, \quad (3.24)$$

and if $\mathcal{X}$ is closed, then

$$\|x(t) - d\| \leq \theta(d), \quad \forall t \Rightarrow \tilde{x} \subset \overline{\mathcal{B}}(d, \theta(d)) \Rightarrow \tilde{x} \subset \mathcal{X}. \quad (3.25)$$

This lemma bears the promise that if we can construct control trajectories such that the generated state trajectory satisfies the premissa for some d , then a sufficient viability condition is established. As we will see soon the nest is just the set of such points d .

LEMMA 2. Omitted.

THEOREM 1. VIABLE CONTROL TRAJECTORIES. Let assumptions “A” to “D” hold and let $d \in \mathcal{D}$ be any nestpoint (3.23). Then $u(\cdot)$ is a viable control trajectory (i.e. it generates a viable state trajectory) if its range $\tilde{u}$ satisfies $\tilde{u} \subset \mathcal{U}_1(d)$ or $\tilde{u} \subset \mathcal{U}_2(d)$, where

$$\mathcal{U}_1(d) := \mathcal{B}\left(-Ad, \frac{\mu}{\kappa}\theta(d)\right) \quad (3.26)$$

$$\mathcal{U}_2(d) := \left\{u : -A^{-1}u \in \mathcal{B}\left(d, \frac{\mu}{\kappa p}\theta(d)\right)\right\}. \quad (3.27)$$

If the viability set $\mathcal{X}$ is closed, then $\mathcal{B}(\cdot, \cdot)$ can be replaced by $\overline{\mathcal{B}}(\cdot, \cdot)$ in (3.26) and (3.27).

Interpretation of Theorem 1. The set $\mathcal{U}_1(d)$ is constructed in the following way. We select a nestpoint d in the state space and via premultiplication by $(-A)$ transfer it to the control space. Furthermore, we determine the clearance $\theta(d)$ of d and multiply it by μ/κ . The former will be the center, the latter the radius of a ball in the control space, this ball is $\mathcal{U}_1(d)$. For forming the set $\mathcal{U}_2(d)$ we also take a nestpoint, and form a ball around it in the state space, whose radius is its clearance reduced by the factor $\mu/\kappa p$. This ball is then mapped into the control space, the matrix of the mapping is $(-A)$. In this way we obtain $\mathcal{U}_2(d)$ which need not be a ball.

The Problem of the Best Nestpoint

As seen from theorem 1, the larger is $\theta(d)$ the larger will be both $\mathcal{U}_1(d)$ and $\mathcal{U}_2(d)$. Thus, in order to enlarge the available set of viable controls we may want to find the best nestpoint, i.e. the one with the largest clearance. Hence, the following optimization problem is to be solved:

$$\begin{aligned} &\text{"Find } d^* \in \mathcal{D} \text{ and } \theta^* := \theta(d^*) \text{ such that} \\ &\theta^* = \max\{\theta(d) : d \in \mathcal{D}\}." \end{aligned} \quad (3.28)$$

This problem, of course, can not be solved without the specification of the viability set.

Having finished the discussion of theorem 1 we have to emphasize that all the parameters which occur in it are calculable relatively simply. One exception is the clearance $\theta(d)$ whose value depends on the unspecified shape of the viability set. Furthermore, we still have a freedom in the choice of the norm. By an appropriate choice which takes the shape of $\mathcal{X}$ into account we can highly improve upon the "roughness" of our estimations, which is caused just by the estimation in norm. In the sequel we deal with a specification of the viability set, where not only $\theta(d)$ will be easily calculable, but a best nestpoint, an explicit solution of problem (3.28) will also be found.

Interval Shaped Viability Sets

Interval

We discuss the case when the viability set is a closed interval (parallelepiped).

Let b be any point in $\mathbb{R}^n$: $b \in \mathbb{R}^n$, and p a nonnegative vector in $\mathbb{R}^n$: $p \geq 0$. We define the closed *interval* (with center b and half-diagonal p) as

$$I(b, p) := \{y \in \mathbb{R}^n : b - p \leq y \leq b + p\}. \quad (3.29)$$

The Viability Set

ASSUMPTION E: The viability set is an interval:

$$\mathcal{X} = I(c, q), \quad (3.30)$$

where $q > 0$.

$q > 0$ implies that $I(c, q)$ has a non-empty interior; hence Assumption E implies Assumption C.

In the sequel we will use the interval vector-norm $\|\cdot\|_q$ as defined in (3.8) where q will be the same positive n -vector that appears in the definition (3.30) of $\mathcal{X}$. The following lemma clarifies the connection between intervals and the closed balls in interval norm.

LEMMA 3. In the interval norm $\|\cdot\|_q$ the closed balls of center y and radius δ are intervals with center y and half-diagonal δq :

$$\overline{\mathcal{B}}(y, \delta) = I(y, \delta q). \quad (3.31)$$

Substituting $y = c$, $\delta = 1$ into (3.31) from (3.30) we get

$$\mathcal{X} = \overline{\mathcal{B}}(c, 1) \quad (3.32)$$

that is, the viability set is represented as a *unit ball* with center c in the q -norm.

The Calculation of the Clearance

LEMMA 4. The clearance of a point $y \in I(c, q)$ is

$$\theta(y) = 1 - \|c - y\|_q. \quad (3.33)$$

The Calculation of the Maximum Clearance of Nestpoints

For the sake of brevity, let us introduce the following notation for the initial clearance

$$\theta_0 := \theta(x_0) = 1 - \|c - x_0\|_q, \quad (3.34)$$

which can be calculated from the given data of $\mathcal{X}$ and x_0 .

We now turn to the solution of the problem of the best nestpoint (3.28) for the case of interval shaped viability set.

LEMMA 5. For the viability set $\mathcal{X} = I(c, q)$ the maximum clearance of a nestpoint is

$$\theta^* = \min \left\{ 1, \frac{\kappa \theta_0}{\kappa - 1} \right\} \quad (3.35)$$

which is taken at a best nestpoint

$$d^* = x_0 + \frac{\theta^*}{\kappa - (\kappa - 1)\theta^*} (c - x_0). \quad (3.36)$$

Interpretation of Lemma 5. Two cases are to be distinguished.

(1) If

$$\theta_0 \geq 1 - \frac{1}{\kappa}, \quad (3.37)$$

then $\theta^* = 1$ from (3.36) and $d^* = c$ from (3.37). Thus, in this case the center of the viability set is the (unique) best nestpoint. This case occurs if $\|c - x_0\| \leq 1/\kappa$, i.e. if the initial point is close enough to the center of $\mathcal{X}$.

(2) If, on the contrary, $\theta_0 < 1 - 1/\kappa$, then

$$\theta^* = \frac{\kappa\theta_0}{\kappa-1} \quad \text{and} \quad d^* = x_0 + \frac{\theta_0}{(\kappa-1)(1-\theta_0)}(c - x_0). \quad (3.38)$$

In this case an optimal nestpoint occurs in the interior of the line segment connecting the initial point x_0 with the center c . d^* of (3.38) will usually not be a unique best nestpoint because the interval norm is not strictly monotonic in the modulus of the components.

Combining Lemma 3 and Lemma 5 with Theorem 1, we get the following viability conditions for $\mathcal{X} = I(c, q)$.

THEOREM 2. VIABLE CONTROL TRAJECTORIES FOR INTERVAL SHAPED VIABILITY SETS. Under assumptions “A”, “B”, “D” and “E” $u(\cdot)$ is a viable control trajectory if its range satisfies $\tilde{u} \subset \mathcal{U}_1^*$ or $\tilde{u} \subset \mathcal{U}_2^*$, where

$$\mathcal{U}_1^* = \begin{cases} I\left(-Ac, \frac{\mu}{\kappa}q\right), & \text{if } \theta_0 \geq 1 - 1/\kappa \\ I\left(-Ad^*, \frac{\mu\theta_0}{\kappa-1}q\right), & \text{otherwise,} \end{cases} \quad (3.39)$$

$$\mathcal{U}_2^* = \begin{cases} \left\{u : -A^{-1}u \in I\left(c, \frac{\mu}{\kappa\rho}q\right)\right\}, & \text{if } \theta_0 \geq 1 - 1/\kappa \\ \left\{u : -A^{-1}u \in I\left(d^*, \frac{\mu\theta_0}{\rho(\kappa-1)}q\right)\right\}, & \text{otherwise,} \end{cases} \quad (3.40)$$

and

$$\theta_0 = 1 - \|c - x_0\|_q, \quad (3.41)$$

$$d^* = x_0 + \frac{\theta_0}{(\kappa-1)(1-\theta_0)}(c - x_0). \quad (3.42)$$

Note

Published first in a slightly different form in *Proceedings of the IFAC Symposium on Dynamic Modeling and Control of National Economies, Edinburgh, 1989*, by Pergamon Press, Oxford. © IFAC, 1990. For a more complete presentation (with proofs), see *Problems of Control and Information Theory*, Vol. 20 (4). Useful comments by V. Kertész, A. Lee, and A. Simonovits are gratefully acknowledged.

References

- Aubin, J. P., and A. Cellina (1983) *Differential Inclusions*. Berlin: Springer.
- Bauer, F. L. (1963). Optimally scaled matrices. *Numerische Mathematik* 5: 73–87.
- Kornai, J., and B. Martos (1973). “Autonomous Control of the Economic System.” *Econometrica* 43: 509–528.
- Kornai, J., and B. Martos, eds., (1981). *Non-Price Control*. Amsterdam: North-Holland.
- Lancaster, P. (1969). *Theory of Matrices*. New York: Academic Press.
- Martos, B. (1990). *Economic Control Structures*. Amsterdam: North-Holland.
- Nagumo, M. (1942). Über die lage der integralkurven gewöhnlicher differentialgleichungen. *Proc. Phys. Math. Soc. Japan* 24: 551–559.
- Zadeh, L. A., and C. A. Desoer (1963). *Linear System Theory*. New York: McGraw-Hill.

4 Coordinating Activities under Alternative Organizational Forms

Yingyi Qian, Gérard Roland, and Chenggang Xu

1 Introduction

Understanding how economic activity is coordinated inside organizations has always been one of the most fascinating questions in economics. This question has also been at the forefront of János Kornai's research preoccupations. He already analyzed problems with coordination in the centralized system in his early work *Overcentralization in Economic Administration* (1959). In *Anti-equilibrium* (1971) he strongly stressed the importance of nonmarket coordination by pointing to what he called the "control sphere" of economic processes, including the exchange of complex information messages (price and nonprice) between economic agents and their reaction to the information received. In *Economics of Shortage* (1980) he studied coordination processes in a socialist economy by extensively analyzing mutual adjustment behavior of economic agents in response to shortage signals. In *The Socialist System* (1992) he further compared the bureaucratic coordination mechanism with the market coordination mechanism and considered the bureaucratic coordination as the root of the many problems that troubled the centralized system. As his former students, we have been deeply influenced by his thoughts. Like many of our other works, it is easy to find his influence on us in this chapter.

Important themes that emerge in the study of coordination in organization are how different organizational forms affect their coordination abilities, and thus their performances. Chandler (1962, 1977, 1990) and Williamson (1975, 1985) analyzed the emergence of the modern multidivisional corporations (the M-form firm) in the 1920s, which became the prevailing organizational form of large businesses. According to them, M-form firms were more efficient than firms organized along functional lines (unitary or U-form firm) because daily operations were decentralized to divisions, which freed the time of the central office for strategic planning. The management-science literature has distinguished between the product focused and process focused organizations, making the argument that alternative organizational forms are the result of minimization of coordination costs in varying unstable environments (Henderson and Clark 1990; Hayes, Wheelwright, and Clark 1988; Stinchcombe 1990; Athey and Schmutzler 1994). Aoki (1986) noted that in Japanese firms, decision-making is more decentralized to workers who are less specialized and are more able to make frequent adjustments and use on-site information than their American counterparts.

The organizational structures of centrally planned economies also varied along similar patterns to those observed in large business firms. The stylized feature of the

Soviet planning system, which characterizes Eastern Europe and the former Soviet Union, has a unitary hierarchical organization divided by ministries based on functional and specialization principles (Nove 1980; Gregory and Stuart 1989; and Ericson 1991). Recently, Qian and Xu (1993) made an observation that, in contrast to the U-form organization, the planning structure of China was mainly along regional lines which can be characterized as a multilayer-multiregional form of organization (the deep M-form, or in brief, the M-form). Conventional wisdom among scholars of central planning was that the Soviet-type U-form organization was better able to exploit economies of scale through functional ministries whereas the M-form organization in China, also implemented in the brief *sovnarkhoz* period under Khrushchev in the Soviet Union, suffered from inefficient duplication of industries. However, Qian and Xu (1993) points out that even though the Chinese-type structure suffers from the problems of lack of specialization and scale economies, it features more decentralized coordination, which has many significant implications for reforms and transition to markets. For example, this organizational form prevents severe disruption of the operation when the coordination role of the central government is dramatically reduced, it also provides important flexibility and allows for regional experiments without interfering with the rest of the economy. That is, this organizational form provides an institutional foundation for a bottom-up approach of an incremental reform process. In contrast, we observe more frequently a top-down approach with radical programs in Eastern Europe and the former Soviet Union.

In order to formalize these ideas, in this chapter, we introduce a new concept to model coordination inside organizations. Specifically, we define coordination as "attribute matching." With this concept, we study the performance of the M-form and U-form organizations in coordinating routine operation and coordinating changes (such as innovation and reform). Our work is in the team theory tradition (Marschak and Radner 1972; Arrow 1974; Weitzman 1974; Crémer 1980; Aoki 1986, 1995) by assuming away the problem of incentives (the problem of incentives under alternative organizational forms is studied in a companion paper by Maskin, Qian, and Xu [1997]). Whereas the previous literature on coordination focuses on resource allocation, we emphasize that attribute matching is an indispensable part of coordination and deserves attention. In particular, the attribute-matching problem differs from the resource allocation problem because the latter almost always involve some substitutions of resources among tasks when an organization adjusts to exogenous shocks, but the former does not. Moreover, the former is about the quality of tasks whereas the later is more about quantities produced. Our approach allows us to open up the black box of decision-making inside an organization and

to concentrate on the interaction between coordination and alternative organizational forms.

This chapter is organized as follows. We introduce the concept of coordination as attribute matching in section 2 and organizational forms in section 3. Section 4 discusses application of this framework to coordination of routine operation inside organizations. Section 5 discusses the application to change (such as innovation and reform). Section 6 concludes.

2 Coordination as an Attribute-Matching Problem

Our basic concept of coordination within organizations is that of “attribute matching.” The notion of “a design attribute” and “innovation attribute” was first described by Milgrom and Roberts (1992). Any product or service can be viewed as the result of assembling of complementary parts: assembling of the parts of a car or of a computer; synchronizing travel, accommodation, and logistics for a conference or a business meeting; assembling subroutines for a software package; and so on. Each part is characterized by its attributes: time; location; technical specifications such as size, weight and bits; and so on. And these complementary parts must be made to fit together. A product or a service is completed successfully only if characteristics of each attribute of the various parts are matched successfully. To take a simple example, the diameter of a screw must match that of a bolt; they must both meet certain standards of material resistance. They must be transported to a given location at a given time in order to be matched. Most products and services require much more sophisticated assembling of parts, each of which has numerous attributes which are relevant in this matching process. Failure of matching attributes implies most often a drastic production failure. The engine of a Rolls-Royce cannot be made to fit in the body of a mini-Morris; subroutines of programs will not perform if they are not linked adequately to the main program; and so forth.

One of the central tasks of modern manufacturing is to organize economic activities to minimize failures of attribute matching. Indeed, the evolution of manufacturing shows that the way of organizing production goes hand by hand with improvements in technology which allowed to dramatically increase the precision of attributes of parts. In the case of the early car production in Europe, such as in Panhard and Levasseur in France, generalist craftsmanship was necessary to solve the attribute matching problem due to the lack of precision of the delivered parts: “P&L could never have made two—much less 200,000 identical cars, even if these were built to the same blueprints. The reasons? P&L contractors did not use a

standard gauging system, and the machine tools of the 1890s couldn't cut hardened steel. . . . When these parts eventually arrived at P&L's final assembly hall, their specifications could best be described as approximate. The job of skilled fitters in the hall was to take the first two parts and file them down until they fit perfectly together. Then they filed the third part until it fit the first two, and so on until the whole vehicle—with its hundreds of parts—was complete. . . . So, by the time the fitters reached the last part, the total vehicle could differ significantly in dimensions from the next stand that was being built to the same blueprint" (Womack et al. 1990). By contrast, mass production in Ford Motors benefited from specialized plants (U-form), which was based on new technologies that reduced drastically attribute shocks by increasing immensely the precision of parts.

Performances of organizations thus depend to a great deal on their ability at coordination in terms of avoiding attribute matching failures in production. Moreover, changes in the ability to prevent attribute matching failures may affect the optimal form of organization. The story told above shows amply that having blueprints specifying attributes of different parts is not enough to solve the coordination problem. Indeed, in practice, attributes of parts depart from their original blueprint. New adjustments must be made to achieve attribute matching in view of such departures from blueprints which we will call "attribute shocks." When there is a shock affecting an attribute in one task, the attributes in other tasks must be adjusted to achieve matching. Our coordination concept precisely relates to the adjustment of the organization to attribute shocks and more generally to exogenous disturbances and random contingencies. This concept expresses the idea that a need for coordination arises in response to exogenous disturbances to preset plans. Many previous studies of organizations have also explored this idea of coordination, notably Galbraith (1973).

3 Organizational Form

For a given technology and cost structure, the performance of an organization in coordination will depend on how well informed are those who take the coordination decisions on the attribute shock, that is, on the distribution of authority in an organization. Because attribute matching implies coordination in the adjustment of attributes of complementary tasks and since information is usually better on the spot, this implies that it is better (1) to partition organizations so as to group complementary tasks together rather than group similar tasks together and (2) to leave coordination decisions to managers of decentralized units where such attribute

matching takes place rather than to leave coordination decisions to the central management office. In other words, attribute matching requires organizations with decentralized self-contained production units. An omnipotent craftsman in the early ages of car making is an extreme example that the craftsman himself is a self-contained "unit" when the coordination problem was so severe and the communication was so bad.

However, there are other dimensions of coordination inside organizations that are relevant when considering the performance of organizational forms and which give benefits to grouping of similar tasks and to centralization in management. Grouping similar tasks may be better from the point of view of resource reallocation decisions. Indeed, putting similar tasks together allows for more easy substitutability between tasks since similar tasks are by nature more easily substitutable between each other. If there is unexpected shortage for some types of screws, it is easy to make adjustments in the production of other types of screws to substitute for the missing one, rather than to substitute bolts to screws. Moreover, grouping similar tasks together allows to reduce duplication and thus to achieve better economies of scale and cost reductions. Finally, putting together similar tasks allows to achieve better compatibility. Compatibility is also related to attributes of production but affects substitutability of products and parts. An example is compatibility between word processing software—how easy and how well to make their outputs mutually readable or convertible. Another one is to design car engines in a way to make them adaptable (compatible) to different car models, etc.

Once coordination involves trade-offs between the advantages of grouping similar or complementary tasks together, the issue arises of whether coordination decisions are better left at the center or whether they should rather be decentralized to the units; what is the optimal distribution of coordination decision-making? This leads to the central question which we are addressing in this paper: what is the optimal organizational form for coordinating tasks? Indeed, on the one hand each unit has privileged information about coordination inside the unit; on the other hand, however, it has rather bad information about other units and even may not be able to communicate well in any coordination activity with the other unit because of the lack of understanding of problems observed by others. By contrast, the center has less good information about each unit but is better informed about them than managers of each unit about the other unit. This thus gives rise to interesting tradeoffs between different organizational forms, and between centralization and decentralization.

We call an organization a U-form if similar tasks are grouped together into one unit. With similar tasks being grouped together, the U-form is often featured with

task specialization, functional principles, process focus, and so on. In contrast, in an M-form organization complementary tasks are grouped into one unit. This organizational form is often associated with “self-contained” divisions. In our definition, U-form and M-form are general categories of organizational forms and each of them may contain many different subforms. For example, an M-form organization can be designed on the basis of geography, product, or technology, respectively. On top of grouping of tasks, organizations depend on whether coordination is decided at the center or decentralized to the various units. We think this definition is quite general to cover the range of phenomena discussed in different contexts.

4 Coordinating Routine Operation

In this section, we discuss coordination problems associated with routine operation. In routine operation, coordination problems arise from exogenous shocks which make reality different from preset plans. In our framework, we take “tasks” as the basic unit of analysis. We distinguish two kinds of coordination inside an organization. The first concerns attribute matching (or fit) between the tasks and the second relates to resource allocation for carrying out the tasks.

4.1 Basic Elements

We look at the performance of organizations in adjusting to attribute shocks as defined above. For studying routine operation, we also incorporate “capacity shocks,” meaning an unexpected breakdown of a production line, which requires reallocation of resources from substitutable production lines.

Specifically, we assume that there are two tasks 1A and 2A for producing products A and two tasks 1B and 2B for producing B (figure 4.1). Each task has two dimensions: attribute and quantity. Consider attribute first. Tasks $1r$ and $2r$ ($r = A, B$) are complementary and involve attribute matching. Tasks iA and iB ($i = 1, 2$), on the other hand, are similar tasks. As discussed earlier, even if all attributes are perfectly matched in blueprints, various “shocks” are likely to hit the organization. This requires attribute adjustment, which is an important purpose of coordination.

Second, we consider the capacity dimension. Capacities are initially all normalized to 1 when there is no capacity shock (a preset plan). Under a capacity shock, some proportion η of the capacity is lost, and coordination decisions are decisions of capacity transfers t_{ir} ($i = 1, 2$ and $r = A, B$) through resource reallocation for the purpose of production smoothing. We assume that resources used for achieving the task 1A and 1B (2A and 2B) are perfect substitutes, possibly after some costly

adjustment. We assume that it is optimal to reallocate resources so as to equalize production of products A and B (which is implied by a symmetric and concave utility function). Capacity adjustment may also require an adjustment cost per unit of capacity transferred (see figure 4.1).

Organizational forms concern how tasks are grouped into units and how the authority of decision making is assigned to managers. A U-form organization is set up along functional lines and has the structure shown in figure 4.2.

A M-form organization is set up along regional or product lines and has the structure shown in figure 4.3.

Under each organizational form, decision-making can be either centralized (that is, the top manager makes a decision) or decentralized (that is, middle manager 1, 2, A, or B makes decisions).

We make the Hayekian assumption about local information: managers have perfect local information, that is information on attribute shocks and capacity shocks concerning their own tasks. We further assume that communication within

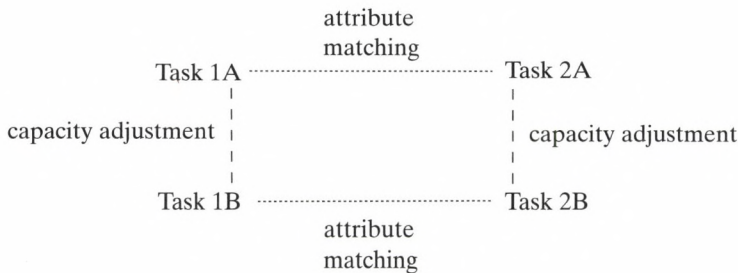


Figure 4.1
Tasks, attributes, and capacities

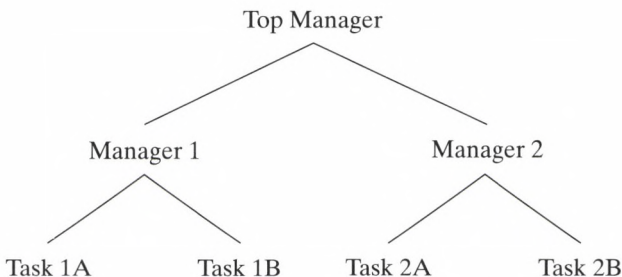


Figure 4.2
A U-form organization

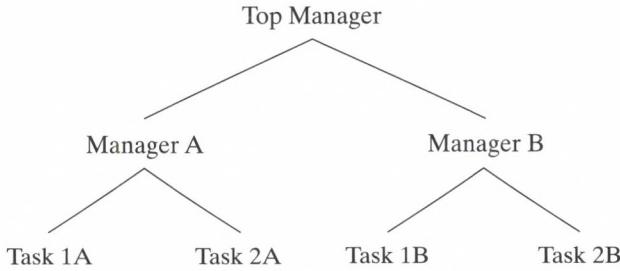


Figure 4.3
An M-form organization

a unit is perfect, but cross-unit (including the one between a middle manager and the top manager) communication is imperfect (noisy). For the sake of simplicity, we assume that there is no horizontal communication between units and that the top manager receives imperfect information about shocks from local managers.

The top manager receives imperfect signals (messages) concerning attribute shocks and capacity shocks of four tasks m_{ir} ($i = 1, 2$, and $r = A, B$) from the middle managers. Suppose $m_{ir} = 1$ (or $m_{ir} = 0$) is a correct (or wrong) message for adjusting attributes or capacities. Then the probabilities of the top manager's getting correct and wrong messages are the following,

$$\text{Prob}(m_{ir} = 1) = \lambda_{ir}; \text{ and } \text{Prob}(m_{ir} = 0) = 1 - \lambda_{ir}.$$

For simplicity, we assume that $\lambda_{ir} = \lambda$ for all i and r . Then the probability of coordination failure is

$$\text{Prob}(m = 0 \text{ and } \varepsilon = 1) = (1 - \lambda)\rho,$$

where ρ is the probability of having a shock. Given that coordination is perfect without any shocks, the probability of no coordination failure is

$$\text{Prob}(\text{there is no shock or attributes are rematched}) = 1 - (1 - \lambda)\rho,$$

which we denote by γ .

4.2 Examples

Example 1: Centrally Planned Economies This example is a brief summary of the comparative study of transition economies by Qian and Xu (1993). Suppose there are two regions in an economy, East (E) and West (W), and two sectors, steel (S) and machine-building (M). The steel industry demands machineries from the

machinery industry and the machine industry needs raw material from the steel industry.

Under a U-form economy (e.g., the Soviet economy), there are two ministries, S and M. All steel-making firms are managed by the S ministry, producing x_s . All the machine-building firms are managed by the M ministry, producing x_m . Under an M-form (e.g., the Chinese economy), there are two regional governments, E and W. Each regional government manages firms of both industries in the region. In region E, the products are x_s^E and x_m^E , and in region W the products are x_s^W and x_m^W .

In the case of centralization, the central government is responsible for all coordination. In the case of decentralization, ministries (U-form) or regional governments (M-form) are responsible for coordination.

The attributes of the tasks in the economy are related to the technical specifications and delivery time of the steel products and machineries, such as delivering type ABCDE steel to machine building factory 12345 at date 5 and delivering type EDCBA machines to steel factory 54321 at date 6. Initially the attributes are designed to match each other in the economic plan. Attribute shocks in the economy include both demand and supply shocks on attributes of steel and/or machines. For example, a different type of steel is available to be supplied because the originally specified type of steel is exported; or an imported new technology in steel industry requires a new type of digital machinery.

Capacity shocks in the economy include changes in demand for steel and/or machine and changes in production capacity of steel and/or machine. For example, tightening up the credit policy or opening to foreign trade may cause such changes. Inventories are used as buffers to deal with capacity shocks. Suppose each machine-producing factory produces semifinished products and final products. Semifinished products of the similar machines are substitutable between different machine-building factories. Each machine-producing factory transfers a proportion of semifinished products to inventory if there is no capacity shock in that region. Similarly, each steel factory transfers a proportion of steel products to inventory if there is no capacity shock in that region.

Example 2: Military Operations—the Normandy Landing In this example, suppose there are two landing fronts: North front (N) and South front (S); and two types of complementary forces: air force (A) and marines (M). There are two ways of organizing the military force for a landing operation.

Under the U-form organization, one general controls all the air forces, that is, the air forces in the north front and in the south front, x_a^N and x_a^S respectively. Another general controls all the marines, which includes the marines in the north front and in the south front, x_m^N and x_m^S respectively. Under an alternative M-form

organization, one general controls all the forces in the north front, which includes the air force and marines, x_a^N and x_m^N respectively. Another general controls all the forces in the south front, which includes the air force and marines, x_a^S and x_m^S respectively.

In the case of centralization, the commander in chief is responsible for all the coordinating tasks between A and M, and between N and S. In the case of decentralization, A and M generals (U-form) or N and S front generals (M-form) are responsible for coordinating tasks.

The attributes of the tasks include the location of bombing and landing, timing of attack, formation of troops, firing specifications, troop specifications, and so on. Mismatching of the attributes between the air force and marines will cause the failure of the operation. Although an operation plan matches all attributes *ex ante*, every general knows that the crucial factor for winning the battle is coordination when the troops make moves, whereas the *ex ante* plans are at their best only guidance or suggestions for the coordination. For example, a weather change may delay the air force operation, so a change in landing time is called for. An unexpected enemy move or deployment may make it necessary to change the landing location of the marines, which in turn might require a change in bombing location and firing specifications by the air force.

Capacities refer to the strength of the troops, volume of firing and bombing, and quantities of airplanes and marines on both the north and south fronts. Capacity shocks occur when there is a reduction of existing capacity due to unexpected casualty or unexpected bad weather, or when there is an increase in demand for more capacity due to unexpected reinforcement of the enemy. To deal with capacity shocks, reserve forces are often used. A proportion of air force and marines on the north and south fronts are reserved if there is no capacity shock on that front.

It is easy to think of more examples from the literature on business organizations, such as the organizational problems faced by Du Pont and General Motors discussed by Chandler (1962) and Williamson (1976), and organizational problems faced by the World Bank (Galambos and Milobsky 1995). In the case of Du Pont in the 1920s, a major organizational problem was how to organize functions such as manufacturing and sales, together with major products such as explosives and fertilizer. In the case of the World Bank, an important organizational problem has been how to organize functions such as lending and economic analysis (e.g., creditworthiness analysis), together with recipient country regions, such as East Asia and Pacific Region and Europe and Central Asia Region.

4.3 Trade-offs between Coordinating Attributes and Resource Allocation

In this section we summarize some results from Qian, Roland, and Xu (1995) on the performance of organizational forms in coordination of routine operations.

The analysis focuses on three different organizational forms: the decentralized U-form where managers 1 and 2 make decisions, the decentralized M-form where managers A and B make decisions, and the centralized organization where the top manager makes decisions. In the centralized organization, the grouping of tasks becomes irrelevant in as much as all coordination decisions are taken by the top manager and all the information has to be sent to the top manager.

Look first at the case where there are only attribute shocks. Observe first that whatever the size of adjustment costs, the decentralized U-form is always dominated by centralization and decentralized M-form. This is because without information exchange across the two units where attribute adjustments have to be carried out in coordination, no action is optimal under the decentralized U-form even when there are attribute shocks. However adjustment is optimally taken within each unit of the M-form given that the information within each unit is perfect. Thus, the decentralized U-form will always be dominated by the decentralized M-form. Moreover, the latter always strictly dominates centralization given that communication from the units to the top is imperfect, that is, $\lambda < 1$.

Take now the case where there are only capacity shocks. A first nonintuitive result is that without adjustment costs (that is, costs of transferring goods across units), the decentralized M-form performs at least as well as the decentralized U-form, and centralization is dominated by both of them. This result is paradoxical and shows an asymmetry to the previous case. Indeed, one would expect that the decentralized M-form where the information structure is worst in coordinating capacity shocks would choose a no action strategy and perform badly, similar to the case of the decentralized U-form under attribute shocks. This turns out to be wrong. Despite the absence of information on capacity shocks in the other unit under the decentralized M-form, there are very simple strategies that allow the decentralized M-form to replicate the tactic of the decentralized U-form where there is perfect coordination in response to capacity shocks. Indeed, under the decentralized U-form when equalization of the capacities is optimal, equalization of capacity can be achieved by the strategy {transfer $\eta/2$ to the other unit when a shock to that unit is observed, and transfer nothing otherwise}, where η is the amount of the capacity being lost due to the shock. Under the decentralized M-form, the above outcome can be replicated by the strategy {always transfer $\eta/2$ of the capacity to the other

unit as long as no shock to the own unit}. To see that, we note that if no units have a shock, transfers will cancel out. If one unit has a shock, the one with no shock will send the appropriate transfers to the other unit whereas if they both have a shock, they send nothing. Optimal resource reallocation is always reached but useless transfers are taking place whenever there is no shock in either unit. Because transfers are costless, this transfer does not cause inefficiency.

A natural question arises immediately: Why can the decentralized M-form replicate the first best of the decentralized U-form under capacity shocks, but the decentralized U-form cannot replicate the first best of the decentralized M-form under attribute shocks? Two basic assumptions about the technology make the attribute matching fundamentally different from capacity substitution. First, the matching function is such that any mismatching will lead to the collapse of production. Second, whenever one attribute of an elementary task changes, the attribute of the other elementary task has to adjust to make a match. Therefore, information about shocks in one task is essential in making a correct adjustment in the other task. But the same is not true for capacity shocks. With the possibility of transfers of resources for substitution, perfect information is not essential in making correct decisions in the face of capacity shocks.

Therefore, information requirements for achieving coordination under decentralization are different for attribute shocks and capacity shocks. The results of our analysis on capacity shocks show that transfer to inventories can be a substitute for communication. But the same does not apply to attribute shocks simply because there is nothing to substitute.

When these transfers are costly, we should not expect the equivalence of the decentralized M-form and U-form. In that case, the decentralized U-form achieves the first best, and the performance of the centralization converges to the decentralized U-form as the quality of communication improves. However, given the assumption of no horizontal communication, the performance of the decentralized M-form can only be improved when transfer costs are reduced. Particularly when the costs go to zero, the decentralized M-form converges to the decentralized U-form. The trade-off between decentralized M-form and centralization depends in a non trivial way on the relation between transfer costs and communication quality between units and the top manager.

Let us now combine attribute shocks and capacity shocks together and assume that these shocks are independent. If adjustment and transfer costs are not too large, then when information gets more precise centralization will dominate strictly both decentralized M-form and decentralized M-form. To summarize, the tradeoffs are the following: The decentralized U-form is good at handling capacity shocks but bad

at handling attribute shocks; the decentralized M-form is good at handling attribute shocks but not at capacity shocks when there are transfer costs; and the centralized organization is not very good at both but improves its performance depending on the quality of the information it receives from below. The theory thus predicts which organizational form should perform better depending on the quality of the communication, the amount of adjustment costs and transfer costs, and the relative seriousness of attribute shocks vs. capacity shocks.

5 Coordinating Changes

In this section we focus on the coordination required when changes are introduced in an organization, such as innovations introduced in business organizations or reforms introduced in transition economies. Successful changes require both a good “program” and successful “coordination” in implementation. Although attribute matching is still a central task of coordination, the major difference between coordinating routine operations and coordinating changes is that in the former case shocks are purely exogenous whereas in the latter case shocks are partly endogenously chosen—the change to be introduced itself represents a set of shocks or is associated with a set of shocks. We argue that organizational forms affect the authority and information structures inside an organization and thus the ability to select a good program through experimentation and to coordinate attribute adjustments.

In their study of the organization of modern manufacturing, Milgrom and Roberts (1990, 1992) underscored the advantage of the flexibility induced by the organizational structure in modern manufacturing and the complementarity of underlying activities. Similarly, we observe strikingly different organizational structures between the Central and Eastern European (CEE) and the Former Soviet Union (FSU) economies on the one hand and the Chinese economy on the other hand before the reform: unlike the unitary hierarchical structure based on specialization principles (the U-form) in the CEE and FSU economies, China’s hierarchical economy has been a multilayer-multiregional one mainly based on the territorial principle (the M-form). This has given China a certain flexibility in experimenting locally with reform, in special economic zones, for example.

5.1 Basic Elements

Consider an organization, that needs to carry out four tasks 1A, 2A, 1B, and 2B in order to make a change. To be focused on more interesting questions related to making changes, we now abstract from capacity shocks by assuming capacities are

normalized to 1. Instead, we introduce “compatibility” as a trade-off from attribute matching. Specifically, we define two sets of attributes between tasks: The first set involves s attributes between tasks $1r$ and $2r$ ($r = A, B$), and the second set involves ks attributes between tasks iA and iB ($i = 1, 2$), where $0 \leq k \leq 1$. Tasks $1r$ and $2r$ ($r = A, B$) are complementary and involve attribute matching as above. Tasks iA and iB ($i = 1, 2$), on the other hand, are similar tasks. Compatibility between tasks iA and iB is required for efficiency. A change is thus well executed only if characteristics of all s attributes in the first set are *matched* and all ks attributes in the second set are made *compatible*.

To take an example, assume that a software company Soft-Micro plans to introduce a major innovation—a Windows2000 operating system together with new application software (e.g., word processor and spreadsheet), such as SM-Word and SM-Excel. To make the whole operating-application system package work, all the related attributes of the Windows2000 and SM-Word (also SM-Excel) should be matched (e.g., memory allocation in SM-Word must be consistent with the requirement of the Windows2000; otherwise, SM-Word will crash in operation). Moreover, to make the whole package being an integrated package, related attributes between SM-Word and SM-Excel should also be compatible (e.g., they should share each other’s output format attributes so that they can share each other’s outputs). Compatibility implies also building in the possibility for conversion from or to other similar software packages with word processors and spreadsheets, as shown in figure 4.4.

In addition to the need for coordination, there may also be uncertainty concerning the nature of a program of change. For instance, the SM Windows2000 system could be either superior or inferior to the existing MS Windows 95 system even when all the attributes are matched or made compatible. We call p the probability of a program being good in the sense that the change increases performance. There is thus a probability $(1 - p)$ that the program turns out to be bad, that is, worse than

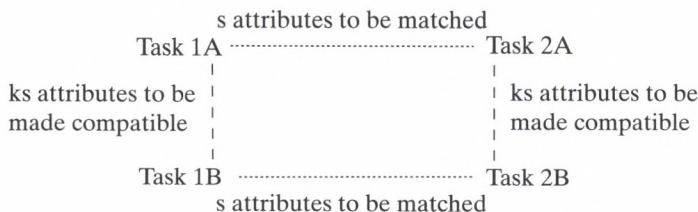


Figure 4.4
Tasks and attributes to be matched and made compatible

current performance. The benefit from making a change depends on whether attributes are matched or not as well as whether the given program is good or bad.

We normalize the status quo output of the organization (without a change) in such a way that the status quo output of a product A (or B) is $\frac{1}{2}$ so that the total output of the organization without change is 1. We assume that if there is no change in the organization, all the attributes of tasks are matched and there are no attribute shocks.

The benefits of changes are defined as follows:

1. When a program is good, if all s pairs of attributes are matched between strict complementary tasks and all ks attributes are made compatible, the outcome of the program in one product (A or B) will be strictly higher than that of the status quo:

$$\frac{A(s)}{2} > \frac{1}{2}.$$

If all s pairs of attributes between strict complementary tasks are matched but at least one of the ks pairs of attributes between compatible tasks are not matched, the outcome of the program in one of products (A or B) is $(1 - k)A(s)/2$.

If all ks pairs of attributes between compatible tasks are matched, but at least one of the s pairs of attributes between strict complementary tasks is not matched, the outcome of the program in one product (A or B) is $\alpha/2$, where

$$0 \leq \frac{\alpha}{2} < \frac{1}{2}.$$

2. When a program is bad, it does not function, that is, the outcome is zero, regardless of the matching of the attributes. However, a bad program can still be useful in providing inventories and/or services if it is fully compatible with the unchanged unit, that is, in this case the outcome of the unchanged unit is $\frac{1}{2}$ (status quo); otherwise, losing part of the inventories and/or services, the outcome of the unchanged unit is reduced to $(1 - k)/2$.

As above, a U-form organization is set up along functional lines (see figure 4.2). Beneath the top manager, there is one manager for each function. Manager i ($i = 1, 2$) is responsible for making ks attributes compatible between similar tasks iA and iB ; and the top manager is responsible for matching the s attributes between distinct tasks $1r$ and $2r$ ($j = A, B$).

An M-form organization is set up along regional lines (see figure 4.3). In this case there is a manager in charge of each region (product) A and B, which are similar to each other. Under the M-form, manager r ($r = A, B$) is responsible for matching the

s attributes between distinct tasks $1r$ and $2r$; and the top manager is responsible for making the ks attributes compatible between similar tasks iA and iB ($i = 1, 2$).

Again we assume that communication is perfect within a unit, but is imperfect (noisy) between units. The top manager receives imperfect signals concerning all four tasks m_{ir} ($i = 1, 2$, and $r = A, B$) from middle managers with

$\text{Prob}(m_{ir} = 1) = \lambda$; and $\text{Prob}(m_{ir} = 0) = 1 - \lambda$.

Here $m_{ir} = 1$ (or $m_{ir} = 0$) is a correct (or wrong) message for adjusting attributes. Thus, the probability of no coordination failure for one pair of attributes is again

$\text{Prob}(\text{there is no shock or attributes are rematched}) = 1 - (1 - \lambda)\rho$,

which we denote by γ . Therefore, the probability of all s pairs of attributes are matched is γ^s .

Under the U-form, manager i ($i = 1, 2$) perfectly coordinates to make attributes between tasks iA and iB compatible, and the top manager imperfectly coordinates to match attributes between $1r$ and $2r$ ($r = A, B$). Similarly, under the M-form manager r ($r = A, B$) perfectly coordinates to match attributes between tasks $1r$ and $2r$, and the top manager imperfectly coordinates to make attributes between iA and iB ($i = 1, 2$) compatible.

In order to implement a change (innovation or reform), managers should be trained and new facilities should be installed (communication, computation, etc.). In other words, there are setup costs. Thus, we assume that there is a setup cost of introducing changes for each manager. The setup cost is paid only once and it is sunk once it is paid. The setup cost for a program with s attributes is $C(s)$, which is increasing and convex. The setup cost can be interpreted as a training cost: before introducing changes all the managers involved need to be trained on the attributes that must be matched. When a change is introduced in two steps, the setup cost can be paid sequentially allowing lower costs to be paid in the first period.

To capture the dynamic nature of making changes, we formulate the model with an infinite number of periods and with a discount factor δ . Some units (one or two in the simplest model) are chosen to do the experiment, and the experiment should be continued within these units until successful coordination is achieved (i.e., in the case of coordination failure, the experiment will be repeated). If a program is good and all s attributes of the program are matched and ks attributes are compatible in period 1, per period gross benefit A will be realized for the rest of time. If a program is bad, the program will be stopped once it is discovered that it is bad.

For each organizational form, there are several strategies available for making changes. These strategies fall into two categories: First, an experiment can be introduced by the top manager and implemented by all the managers together. We call

this a “top-down” strategy. Second, an experiment (or two parallel experiments) can be introduced by the unit level manager(s) first and only the manager of the unit who introduced the experiment is responsible for the implementation. The organization-wide change will be introduced later conditional on the result of the first period experiment. We call this a “bottom-up” approach.

We consider the following four strategies under the U-form:

1. Strategy T2 (top-down with experiment in 2 units): the same experiment is introduced into both units 1 and 2; the two middle managers are responsible for compatibility between similar tasks and the top manager takes care of attributes between the two strict complementary tasks.
2. Strategy T1 (top-down with experiment in 1 unit): an experiment is first introduced into only one of the units and the other unit will be changed later only if the earlier experiment is a success; one middle manager is responsible for compatibility inside his unit and the top manager takes care of attribute matching between the complementary tasks.
3. Strategy B1 (bottom-up with one unit experimenting): one experiment in one unit, one middle manager takes care of compatibility and top manager does not match attributes.
4. Strategy B2 (bottom-up with independent experiments in two units): two independent experiments in two units, two middle managers take care of compatibility and the top manager does not match attributes.

Similarly, we consider the following four strategies under the M-form:

1. Strategy T2 (top-down with experiment in two units): the same experiment is introduced into both units A and B; two middle managers are responsible for matching attributes between complementary tasks and the top manager takes care of attribute compatibility between the two compatible tasks.
2. Strategy T1 (top-down with experiment in one unit): an experiment is first introduced into only one of the units and the other unit will be changed later only if the earlier experiment proves to be a success; one middle manager is responsible for matching attributes between complementary tasks and the top manager takes care of attribute compatibility between the compatible tasks.
3. Strategy B1 (bottom-up with one unit experimenting): the same as T1 but the top manager does not take care of compatibility.
4. Strategy B2 (bottom-up with two units experimenting independently): two independent experiments in two units, two middle managers match attributes and the top manager does not take care of compatibility.

5.2 Two Examples

Example 1: Innovations in Production Suppose the economy has two industries: a machine building industry producing bearings and a truck industry producing trucks. Function 1 is to supply thousands of types of bearings to the truck industry, and function 2 is to manufacture two models of trucks to accommodate the two regions in the economy: regions A and B. Determined by regional industrial features (e.g., an oil-producing region vs. a textile-producing region) and road conditions (e.g., a mountain region vs. a desert region), each region has its idiosyncratic preference on truck models (call them model A and model B trucks respectively). Task 1A is to produce bearings for model A truck. Task 1B is to produce bearings for model B truck. Task 2A is to produce model A trucks for region A. Finally, task 2B is to produce model B trucks for region B.

The attributes of bearings and truck models should be matched with each other in order to produce quality trucks. There is a strong complementarity between attributes of bearings and attributes of trucks (strong complementarity between tasks 1A and 2A, between tasks 1B and 2B). Suppose every truck model has two parts, an engine (engines for the two truck models are called AE and BE respectively) and transmission (transmission systems for the two truck models are called AT and BT respectively), each of them needs to be fitted with a bearing. Bearing AT should match the transmission part AT, bearing BE should fit with the engine part BE, and so on. If one pair of attributes between the bearings and the truck is not matched, for instance, bearing AT does not fit truck part AT, then model A truck cannot move. Moreover, some sets of bearings and the two truck models should be compatible (compatibility between tasks 1A and 1B, and between 2A and 2B). For example, bearing AT should be compatible with bearing BT; transmission system AT should be compatible with transmission system BT but engine AE does not need to be compatible with an engine BE. Coordination here is about looking for solutions to match bearings with parts of truck models, and to make two sets of bearings compatible with each other, and two models of trucks compatible with each other when there are changes. Changes can be due to technological innovation or reforms. For example, a technological innovation in transmission may make a truck more efficient, but it will change the technical requirements on bearings.

Suppose there is an innovation in the transmission system. However, before implementing the innovation, people don't know if the new transmission system is good or bad. Let us first look at the case that the new transmission system is good. If every set A bearing is matched with every part of model A truck, the new model A truck will work well and will be better than the old model A. However, if one of

the bearings does not fit with the corresponding part of the truck, the new truck model will not work. Moreover, the new bearing AT should be compatible to bearing BT, and the new model A truck should be compatible to model B truck in their transmission systems as well. Otherwise, even the two model trucks can run well, the lack of service (or higher costs of inventories) due to the incompatibility of the transmission systems will lower the value of the new truck model.

If the new transmission system is bad, then the truck will not run normally regardless of whether the bearings fit or not fit to corresponding parts of the truck. However, if the innovation is introduced into only one of the models with the other model unchanged, then as long as the new set of bearings is still compatible to the other set of bearings, the inventory provided by the new set of bearings is still useful for the other truck model.

If the economy is organized along the lines of specialized ministries, the machine-building ministry is responsible for the two bearing plants and the truck ministry is responsible for the two truck plants, and the national government is responsible for matching attributes between bearings and trucks (U-form). Under this organization, it is likely that all bearings and trucks are compatible across models, but there may be bad coordination between bearings and trucks due to changes of the model of trucks, say.

If the economy is organized along a regional partition, each regional government is responsible for two plants producing one model of truck and the corresponding bearings, and the national government is responsible for making attributes across model compatible (M-form). Under this organization, it is likely that quality trucks can be produced when there is a change in production due to exogenous reasons, but the drawback is the lack of coordination across models.

Example 2: Enterprise Restructuring and the Creation of Social Safety Net This example reflects stylized facts transition economies are facing. In particular, the recent nationwide implementation in China of a reform package including bankruptcy and dealing with laid-off workers is very close to our example here. In the Chinese case, it is the city governments, not the ministries, who take the major responsibility for implementing the reform package (*People's Daily*, 21 April 1997). Suppose a reform consists of two components: enterprise restructuring (laying off redundant workers and promoting the private sector) and creation of a social safety net (providing social benefits and retraining to laid off workers). The economy has two regions A and B. There are in total four tasks. Task 1A is layoffs in region A. Task 1B is layoffs in region B. Task 2A is paying unemployment benefits through the social safety net in region A. Task 2B is paying unemployment benefits through

the social safety net in region B. First, consider the attributes of the tasks of enterprise restructuring and social safety net that should be matched with each other to make the reform successful. Here attributes of enterprise restructuring are such things as the number of layoffs, characteristics of laid-off workers (age, seniority, family composition, length of residence, sex, types of contracts, current wages, history of employment, skills, and education, etc.), and characteristics of skilled workers demanded in the private sector. Attributes of the social safety net are rules of eligibility (length of employment, special circumstance such as veterans, status of enterprise), of benefits (size and length, types of benefits such as monetary or non-monetary), and characteristics of training provided to laid-off workers, technical support of computers, administration, budget, and so on. If one pair of the attributes of the two tasks is not matched, some categories of laid-off workers may not be compensated, or the training provided does not match the demand such that the laid-off workers are not able to find jobs, which may makes the system unsustainable financially. All of this may lead to social unrest and to a situation worse than the status quo with no reform.

Second, consider the attributes of the tasks of enterprise restructuring and social safety net that should be made compatible across regions to make the reform more efficient. Here attributes of enterprise restructuring are the layoff policies (more or less stringent), sectoral priorities, and so on. If these attributes in the two regions are not compatible, inefficient closedown may occur. If attributes of compensation are not compatible across regions, it may lead to too much or too little labor mobility. For example, if benefits cannot be carried from one region to another region, then too little labor mobility results. A similar problem can arise if the regions use two incompatible computer software for social safety net purposes. The lack of compatibility is inefficient but it is not likely to lead to an outcome as bad as social unrest.

If the reform is organized by specialized committees or ministries, then each specialized ministry is responsible for either enterprise restructuring or social safety net, and the national government is responsible for matching attributes between enterprise restructuring and the social safety net (U-form). It is likely that layoff policies are compatible across regions (same for the compensation rules), but it is also likely that there will be bad coordination between layoff policies and the introduction of the social safety net, leading to a higher probability of social unrest. For example, the rules for eligibility set at the national level may be completely inappropriate in some important regions (example: assume in some regions mainly older workers are concerned by layoffs and that the national rule for eligibility is such that workers close to pension are not entitled to any benefits).

If the reform is organized by regions, each regional government is responsible for matching attributes of enterprise restructuring and social safety net in its own region, and the national government is responsible for making attributes of two reforms in two regions compatible (M-form). Under this organization, it is more likely that layoff policies and institution of social safety nets can be coordinated within each region so that it may be easier to avoid social unrest. The drawback is that the lack of coordination across regions may lead to inefficiencies.

5.3 The Optimal Patterns of Change under Alternative Organizational Forms

We now look at how changes are made optimally under different organizational forms. This is a brief summary of ongoing research by Qian, Roland and Xu (1997). Let us start first with the case that the program is always good, that is, $p = 1$, and the uncertainty in coordination is the only concern. We show that the only undominated strategies under both M-form and U-form are T2 (top-down with two units doing experiments). Indeed, in all other strategies the benefits of reform, which are substantial when $p = 1$, are delayed. Although in other strategies part of the costs are also delayed, the advantage of delaying the costs is offset by the disadvantage of delaying the benefits. So, when the benefits are greater than the costs (otherwise the experiment would not be introduced), there are no benefits of delay. We are thus left with a tradeoff between U-form and M-form both using the same strategy. This tradeoff depends on two parameters: how good is the quality of communication between the units and the top, λ , and how large the scope of attributes to be made compatible compare with the attributes to be matched, k . The U-form will dominate if the communication quality between the units and the top, λ , is high. This is because it will benefit from lower costs while suffering little from losses due to attribute mismatching. The M-form on the other hand can match attributes better in self-contained units and its disadvantage in coordinating compatibility is less severe since losses from absence of compatibility are by definition not as severe. On the other hand, it suffers from duplication in setup costs.

What happens when there is uncertainty about the program quality, that is, $p < 1$? Here the flexibility advantages of the M-form may emerge. Notice first that unlike the case of $p = 1$, bottom-up strategies B1 (with one unit experimenting change) and B2 (with two units doing independent experiments) are not necessarily dominated anymore. Indeed, under B1 fewer setup costs are incurred in case the program proves to be bad. These lower experimentation costs make the organization more willing to innovate especially when uncertainty of the program quality is high. Similarly, B2 has the advantage of parallel experiments increasing the likelihood of introducing a successful program with the same setup costs as in a top-down

approach. One can show that for lower values of p , the U-form will generally be dominated by the strategies of T1, B1, or B2 under the M-form. These strategies under the M-form are more efficient than the T2 strategy under either organizational form when the quality of the program pool is poor, that is, when p is low. Alternatively, these strategies have a lower p -threshold for experimentation. Indeed, this is related to the flexibility of the M-form. Note, however, that these flexibility advantages play only for low values of p . When p goes to 1, the U-form may dominate.

6 Conclusion

The purpose of this chapter is to introduce to the modeling of coordination inside organizations as "attribute matching" and to show that this approach represents a new method of studying organizations which can generate a rich set of results. We have analyzed the comparative performances of two forms of economic organization: the U-form, where similar tasks are grouped together, and the M-form, where complementary tasks are grouped together. Using a team-theoretic framework, we compared the performance of these organizational forms in the coordination of routine activities and in their capacity to introduce innovations and carry out reforms.

Our analysis carries over to the organizational forms in centrally planned economies. The choice of the centralized U-form by Soviet planners (starting with Stalin) is generally explained by their desires to exploit the gains from specialization by copying modern business organizational structure to the national economy when the U-form was the only available successful organizational form in the industrialized countries (in the 1920s). Correspondingly, in addition to Mao's ideological dislike for central bureaucracy, the choice of the M-form organization in China was partly a reaction to the severe coordination problems in their experience of the centralized U-form system in the 1950s, which was copied from the Soviet Union. Interestingly, these initial choices of organizational forms have had important effects on reform strategies. The decentralized M-form in China allowed for experimentation with reforms in an incremental way. However, with strong specialization and centralization, the former centrally planned Central and Eastern European and Soviet economies suffered from the rigidity of their U-form structure that there was not much room for local experiments. Our theory suggests that, given that feature of their system, a top-down radical approach to reform may be optimal (see, however, Dewatripont and Roland 1995 on the possibilities of sectoral gradualism).

Our theory also sheds some lights on the emergence and evolution of U-form and M-form organizations in business history. In the early part of the century, newly emerged large business firms were organized in U-forms. Accompanying the growth of these large firms, market demand became more sophisticated, more heterogeneous, and more volatile. The growth of the firms also depended more on innovations which, along with more sophisticated production, increased the demand for good communication. But communication quality is ultimately limited by a human being's capability to process information and understand its content. According to our model, the growing superiority of the M-form may be explained by the fact that technological progress has dramatically increased the costs of coordination failures in the form of mismatching of attributes, and that this tendency has grown faster than the increase in the quality of vertical information transmission and gains from specialization. As we can see, quality of coordination is crucial for the performance of organizations.

Finally, we want to emphasize that the material reported in this chapter are only basic ideas and preliminary results of our ongoing projects. Complete theories are still in the process of development.

References

- Aoki, Masahiko (1986). Horizontal vs. vertical information structure of the firm. *American Economic Review* (December): 971–983.
- Aoki, Masahiko (1996). An evolving diversity of organizational mode and its implications for transitional economies. *Journal of Japanese and International Economies* 10: 330–353.
- Arrow, Kenneth (1974). *The Limits of Organization*. New York: Norton.
- Athey, Susan, and Armin Schmutzler (1994). Product and Process Flexibility: Implications for Organizational Design. Graduate School of Business, Stanford University.
- Becker, Gary, and Kevin Murphy (1992). The division of labor, coordination costs, and knowledge. *Quarterly Journal of Economics* (November) 107(4): 1137–1160.
- Bolton, Patrick, and Mathias Dewatripont (1994). Firms as networks of communication. *Quarterly Journal of Economics*, 809–839.
- Bolton, Patrick, and Joseph Farrell (1990). Decentralization, duplication and delay. *Journal of Political Economy* 98(4): 803–827.
- Chandler, Alfred Jr. (1966). *Strategy and Structure*. New York: Doubleday.
- Cremer, Jacques (1980). A partial theory of the optimal organization of bureaucracy. *Bell Journal of Economics* 11: 683–693.
- Dewatripont, Mathias, and Gérard Roland (1995). The design of reform packages under uncertainty. *American Economic Review* 83(5): 1207–1223.
- Ericson, Richard E. (1991). The classical Soviet-type economy: nature of the system and implications for reform. *Journal of Economic Perspective* 5(4): 11–27.
- Galambos, Louis, and David Milovsky (1995). Organizing and reorganizing the World Bank, 1946–1972: A Comparative Perspective. *Business History Review* 69: 156–190.

- Galbraith, J. (1973). *Designing Complex Organizations*. Reading, Mass.: Addison-Wesley.
- Gregory, Paul, and Robert Stuart (1989). *Comparative Economic Systems*. Third edition. Boston: Houghton Mifflin.
- Harrison, Mark (1985). Investment mobilization and capacity completion in the Chinese and Soviet economies. *Economics of Planning* 19(2): 56–75.
- Hayes, R. H., S. C. Wheelwright, and K. B. Clark (1988). *Dynamic Manufacturing*. New York: Free Press.
- Henderson, R., and K. Clark (1990). Architectural innovation: the reconfiguration of existing product technologies and the failure of established firms. *Administrative Science Quarterly* 35: 9–30.
- Kornai, János (1959). *Overcentralization in Economic Administration*. London: Oxford University Press.
- Kornai, János (1971). *Anti-Equilibrium*. Amsterdam: North-Holland.
- Kornai, János (1980). *Economics of Shortage*. Amsterdam: North-Holland.
- Kornai, János (1992). *The Socialist System*. New York: Oxford University Press.
- Marschak, Jacob, and Roy Radner (1972). *Economic Theory of Teams*. New Haven, Conn.: Yale University Press.
- Maskin, Eric, Yingyi Qian, and Chenggang Xu (1997). Incentives, Information, and Organizational Form. Mimeo, Harvard University, Stanford University, and London School of Economics.
- Milgrom, Paul, and John Roberts (1990). The economics of modern manufacturing: products, technology and organization. *American Economic Review* 80(3): 511–528.
- Milgrom, Paul, and John Roberts (1992). *Economics, Organization, and Management*. Engelwood Cliffs, N. J.: Prentice Hall.
- Nove, Alec (1980). *The Soviet Economic System*. Second edition. Boston: Allen & Unwin.
- Qian, Yingyi, and Chenggang Xu (1993). Why China's economic reform differs: The M-form hierarchy and entry/expansion of the non-state sector. *Economics of Transition* 1(2): 135–170.
- Qian, Yingyi, Gérard Roland, and Chenggang Xu (1995). Coordination and Information in M-Form and U-Form Organizations. Mimeo.
- Qian, Yingyi, Gérard Roland, and Chenggang Xu (1997). Coordinating Changes in M-Form and U-Form Organizations. Mimeo.
- Radner, Roy (1993). The organization of decentralized information processing. *Econometrica* 62: 1109–1146.
- Stinchcombe, A. (1990). *Information and Organization*. Berkeley: University of California Press.
- Weitzman, Martin (1974). Price vs. quantities. *Review of Economic Studies* 41: 477–491.
- Williamson, Oliver (1975). *Markets and Hierarchies*. New York: Free Press.
- Williamson, Oliver (1985). *The Economic Institution of Capitalism*. New York: Free Press.
- Womack, James P., Daniel T. Jones, and Daniel Roos (1990). *The Machine That Changed the World: The Story of Lean Production*. New York: HarperCollins.

5 Democracy, Egalitarian Redistributions, and the Level of Public Bads

John E. Roemer

1 Introduction

In various writings published recently (Roemer 1993, 1994a, 1996) I have argued that a market economy, which is also a democracy, will tend to have lower levels of certain public bads the more egalitarian the distribution of profits of its firms is. The public bads in question are those that can be viewed as inputs in the production functions of firms. These include, for example, fast assembly-line speeds (as a proxy for stressful working conditions), industrial pollution, and destruction of natural environments for industrial purposes. My informal argument for the stated proposition was as follows. Individuals derive increased incomes when these “inputs” are used in greater quantities, because both profits and wages increase. They also sustain a direct decrease in utility from working faster, inhaling more pollutants, and having fewer redwood forests to frolic in. Individuals who derive huge incomes in the form of profits, I argued, generally want higher levels of these public bads, as the income effect dominates the direct disutility effect. If profits are redistributed in an egalitarian way, there will be no class of individuals whose income increases so rapidly with an increase in these profit-inducing public bads. Now assume that the level of these public bads is set by the polity, through democratic procedures (i.e., labor legislation, environmental regulation, and so on). It then seems possible, maybe even probable, that the democratically arrived at level of regulation, setting allowed levels of these public bads, would demand fewer of them, the more egalitarian the distribution of profits.

The studies referred to in the first sentence constructed general political-economic equilibrium models to study this conjecture. In those models, I endowed all individuals with the same utility function over a single consumption good, produced by firms, and a public bad, which was an input in the firms’ production functions. Usually, I took the utility function to be quasi-linear, of the form

$$u(x, G) = x - \frac{\gamma}{2} G^2,$$

where x is consumption of the good and G is the level of the public bad. A distribution of skill levels and profit shares was postulated, giving rise to a competitive equilibrium for each possible level, G , of the public bad. This enabled me to construct an indirect utility function, $v^i(G)$, for each individual i , giving his utility in the equilibrium with amount G of the public bad. These functions were then used to determine a political equilibrium level of G —call it $\bar{G}$.

A general property of these models is that the ideal level of the public bad for individual i (i.e., the argument-maximum of the function $v^i(G)$) was an increasing function of i 's share of aggregate profits. What happens if profit shares are now redistributed to be closer to the egalitarian distribution? It depends on exactly what political process is used to determine $\bar{G}$. If the share holdings of the median-income voter increase with such a redistribution (which would be the case for any actual capitalist economy), then I showed that her ideal level of the public bad would increase. (This does not follow from my previous statement about monotonicity of ideal levels of the public bad in income, because we have to compute a new general political-economic equilibrium after profit shares are revised.) Thus, if the political process delivers an outcome that is the ideal policy of the median voter, then egalitarian redistributions of profits would *increase* the level of public bads.

But I went on to argue that median-voter theories of politics are naive. Indeed, the wealthy may have much more influence on the outcome of democratic political competition than their weight as voters would suggest. This is because the wealthy have resources enabling them to organize political parties, to influence the views of voters generally through the media they own and through lobbying. I suggested, then, that the decrease in the level of public bads desired by the wealthy, following an egalitarian redistribution of profit shares, might well be much more significant, politically speaking, than the increase in the desired level of public bads of the median voter—and therefore, one might expect that the level of public bads generated by the democratic process would decrease with egalitarian profit redistributions.

Indeed, in Roemer (1994a) and (1996) I proposed an ad hoc model of a political process, in which the level $\bar{G}$ was postulated to be the level that maximized a weighted sum of the indirect utilities of the citizens, with larger than per capita weights given to rich citizens, and smaller than per capita weights given to poor citizens. I demonstrated, in those models, that indeed, for reasonable vectors of such weights, the equilibrium level $\bar{G}$ did decrease with egalitarian redistributions of profits.

This review brings me to the motivation of the present chapter. I was dissatisfied with the ad hoc model of the political process that I just described. In democracies, political competition is carried out through competition of political parties, representing different interest groups in the citizenry. It is not clear that the outcome of that competition can be modeled (in reduced form) as the maximization of some social welfare function, as described in the previous paragraph. In this chapter, then, I study a general political-economic equilibrium model in which the political process is, precisely, modeled as competition between two political parties, one of whom (Labour) represents relatively poor voters, and the other of whom (Tories) repre-

sents relatively wealthy voters. I have, in previous work, proposed such a model of party competition (Roemer 1994*b*, 1994*c*), and here I use that model to study what happens to the democratically arrived at level of the public bad as profits are redistributed in an egalitarian manner.

The conclusion, briefly, is that, with this new specification of political competition, the level of the public bad increases with egalitarian redistributions of profits. This reduces my confidence in the conjecture that I originally put forth, that “market socialism” or “economic democracy”—market economies with rather egalitarian distributions of aggregate profits—would engender a decrease in the levels of profit-inducing public bads.¹

When we work with quasi-linear utility functions of the above form, we have an unambiguous measure of the degree of efficiency of an allocation, namely, how close it is to the surplus-maximizing allocation. In particular, there is a unique level of the public bad associated with Pareto efficiency, and so we can also ask the question: Does an egalitarian redistribution of profits increase the efficiency of the economy? The answer to this question turns out to be yes. Of course, when profits are distributed in a more egalitarian way, the income distribution is more egalitarian as well: in a quasi-linear economy, recall that distribution is completely divorced from efficiency (in the sense that a given level of total surplus is compatible with “any” distribution of income). Thus, it turns out that on the two counts of efficiency and equality, equal profit redistributions are good, but they are bad as concerns the level of the public bad.

A true “economist” is concerned only with efficiency and equality; the level of the public bad, itself, means nothing, and so this “economist” should, if she believes the models below, be undisturbed by the fact that egalitarian profit redistributions increase the level of the public bad. An “environmentalist” cares about the level of the public bad, somewhat independently of its impact on human welfare. Are there ways of consolidating the views of these two types? The environmentalist could argue that the welfare of animals (other sentient beings) is not being considered in the models below, and if it were, higher levels of public bads would actually lead to decreases in efficiency. A more mainline argument she could make is that the preferences of the present generation of individuals are not sufficient to compute the efficiency of an allocation, because the impact of today’s high levels of public bads on future generations is not appropriately counted (there being an incomplete set of futures’ markets). A third argument, less mainline, is that individuals’ preferences are adaptive—preferences do not reflect interests, and in particular, the level of public bads people have come to tolerate (as reflected in their preferences), perhaps through a process of cognitive dissonance, is higher than the level that is good for them. (See, for instance, Akerlof and Dickens 1982.)

I ascribe some importance to the last two arguments just given, and there are many who ascribe importance to the first.

2 The Model

We consider a population of workers with skill levels distributed according to a probability measure S , with density σ , on the interval $\Omega = [0, \infty)$. There is a single firm with production function

$$g(L, x) = \delta L^\beta G^{1-\beta}, \quad (5.1)$$

where L is labor employed in efficiency units, and G is the level of a public bad. (A worker with skill level α yields α efficiency units of labor, when she works for one day.) The ownership structure of the firm is described by a probability measure Z on Ω , with density ζ . That is, for a measurable set $A \subset \Omega$, those whose skill levels α lie in A own fraction $Z(A)$ of the firm. Define

$$f(\alpha) = \frac{\zeta(\alpha)}{\sigma(\alpha)}.$$

Then an agent α receives $f(\alpha)$ per capita units of the firm's profits as his dividend. We assume:

A1. $f(\alpha)$ is an increasing function.

The motivation here is that higher-skilled workers have greater wage income, and will, over time, come to own more shares of the firm than lower-skilled workers.

All individuals have identical utility functions

$$u(x, G) = x - h(G), \quad (5.2)$$

where x is consumption of the good produced by the firm, and $h(\cdot)$ is convex. There is no utility derived from leisure.

I briefly outline the equilibrium concept. For any given level of G , there will be a competitive equilibrium in the economy. Define $v(G; \alpha)$ as the utility an agent with skill α derives at this equilibrium. Given these indirect utilities over the level of public bad, a political process will determine a political-equilibrium level of the public bad. The simplest such political process would be one in which the ideal level of G for the voter with median α is the political equilibrium, but I shall use a more realistic political process in which two parties, representing different interests in the polity, compete. The exercise that interests me is to compare the equilibrium level

of G with the economy just described, to the level that would ensue were profit shares to be equally distributed among the population (i.e., if $f(\alpha) \equiv 1$).

Now to details. Normalize the price of output at unity, and let $w(G)$ be the market wage of the worker with $\alpha = 1$, where the level of public bad is fixed at G . Competition assures that, in equilibrium, the wage of a worker of type α is $\alpha w(G)$. Facing these prices, the profit-maximizing firm, which we will view as maximizing profits per capita, proposes a labor demand function $\varphi(\alpha)$ to maximize

$$\delta G^{1-\beta} \left(\int_{\Omega} \alpha \varphi(\alpha) d\alpha \right)^{\beta} - \int_{\Omega} \alpha w(G) \varphi(\alpha) d\alpha, \quad (5.3)$$

where $\varphi(\alpha)$ is interpreted as the number of units of labor per capita of type α it demands. Call profits per capita, the expression (5.3), $\Pi(\varphi, w, G)$. At these prices the consumption of a worker of type α is

$$\alpha w(G) + f(\alpha) \Pi(\varphi, w, G). \quad (5.4)$$

All workers offer their entire stock of labor to the firm (as long as $w(G) > 0$). We may therefore define:

DEFINITION 1. A *competitive equilibrium* at G is a wage $\bar{w}(G)$, and a labor demand $\bar{\varphi}(\alpha)$ such that $\bar{\varphi}(\alpha) = \sigma(\alpha)$. (That is, in a competitive equilibrium the labor market must clear.)

We may easily compute the unique competitive equilibrium at G . Define $\bar{\varphi} = \int_{\Omega} \alpha \bar{\varphi}(\alpha) d\alpha$. Then we can write profit-maximization (see (5.3)) as

$$\text{Max}_{\varphi} \delta G^{1-\beta} \bar{\varphi}^{\beta} - \bar{w}(G) \bar{\varphi},$$

whose solution is

$$\bar{\varphi} = \left(\frac{\delta \beta G^{1-\beta}}{\bar{w}(G)} \right)^{\frac{1}{1-\beta}}. \quad (5.5)$$

Define $\mu = \int_{\Omega} \alpha \varphi(\alpha) d\alpha$. At equilibrium $\bar{\varphi} = \mu$, and so we may solve (5.5) for the wage:

$$\bar{w}(G) = \frac{\delta \beta G^{1-\beta}}{\mu^{1-\beta}}. \quad (5.6)$$

This is the unique equilibrium wage (at unit skill).

We next define agent α 's utility at the competitive equilibrium at G , from (5.4), as:

$$v(G; \alpha) = \frac{\alpha \delta G^{1-\beta}}{\mu^{1-\beta}} + f(\alpha) \left\{ \delta G^{1-\beta} \mu^\beta - \frac{\delta \beta G^{1-\beta}}{\mu^{1-\beta}} \mu \right\} - h(G),$$

which can be more simply written as:

$$v(G; \alpha) = G^{1-\beta} \psi(\alpha) - h(G), \quad (5.7)$$

where

$$\psi(\alpha) \equiv \delta \left(\frac{\alpha}{\mu^{1-\beta}} + f(\alpha) \mu^\beta (1-\beta) \right).$$

Note that v is concave in G (hence single-peaked); the ideal G for individual α is the solution of the F.O.C. associated with (5.7):

$$(1-\beta)G^{-\beta}\psi(\alpha) = h'(G). \quad (5.8)$$

Calling the G so characterized $\hat{G}(\alpha)$, we implicitly differentiate (5.8) to get:

$$\frac{d\hat{G}}{d\alpha} = \frac{(1-\beta)\hat{G}^{-\beta}\psi'(\alpha)}{h''(G) + (1-\beta)\beta G^{-\beta-1}}. \quad (5.9)$$

Note that $\frac{d\hat{G}}{d\alpha} > 0$, because $h'' > 0$ (convexity of h), and because $\psi'(\alpha) > 0$ (this

uses, importantly, assumption A1). Hence an individual's ideal G increases with his skill. This fact, plus the single-peakedness of the functions $v(G, \alpha)$ in G , proves that there is a unique Condorcet winner among choices for the value of G , given by $\hat{G}(m)$, where m is the median of the distribution σ , that is,

$$\frac{1}{2} = \int_0^m \sigma(\alpha) d\alpha.$$

In most actual capitalist economies, $\zeta(m) = 0$, and so $f(m) = 0$: the worker at the median of the wage distribution typically holds no corporate stock. More weakly, it is surely always the case in actual capitalist economies that $f(m) < 1$. The Condorcet winner, from (5.8), is G^* solving the equation

$$(1-\beta)\psi(m) = h'(G^*)(G^*)^\beta. \quad (5.10)$$

Now consider the comparative static experiment in which shares are equally distributed among all. Then $\psi(m)$ increases (see its definition), because $f(m) < 1$ before redistribution and $f(m) = 1$ after. It follows that G^* , the solution to (5.10), increases. We have thus shown:

Let G be chosen by a political process in which voters exercise their preferences $v(G; \alpha)$, and in which the political equilibrium $\bar{G}$ is the argument-maximum of $v(G; m)$ (i.e., median-voter politics). Then, if profit shares are redistributed in an egalitarian manner, the equilibrium level of the public bad rises.

3 A More Realistic Political Mechanism

I know of three political environments that lead to median voter politics (MVP), that is, a political equilibrium consisting in the ideal point of the voter with median ideal point. The first is the classical Downsian model, where there are two political parties, or candidates, each seeking to maximize the probability of victory. It is well known that if both candidates know the voters' preferences, we get MVP. It is less well known that, even if candidates are uncertain about voter preferences, we get MVP when parties are Downsian (Roemer 1994*b*, theorem 6.1). The third environment is one in which there are two political parties, each with policy preferences, whose ideal policies span the median voter's ideal point. If parties are certain of voter preferences, and if those preferences satisfy a single-crossing property, then again we get MVP (Roemer 1994*a*, theorem 2.1).

I consider all three of these environments unrealistic: I believe that, in capitalist democracies, political parties *do* have policy preferences, *and* that they are uncertain about voter preferences. This case does not lead to MVP: in fact, in political equilibrium, parties put forth different policies. I shall now use this political model to study the comparative statics of the equilibrium level of G when profit shares are redistributed in the egalitarian manner.

Let there be two political parties, which represent different interests in the population. To be simple, I shall assume that one party's preferences over G are exactly those of a voter with skill level α_L , and the other party has exactly the policy preferences of a voter with skill α_H . (One party represents low (L) skilled workers, and the other, high (H) skilled workers.) I shall specify, now, to the case where

$$h(G) = \frac{\gamma}{2} G^2,$$

for I am unable to derive general results in this model.

There are various ways of introducing party uncertainty about voter preferences: I shall do it in two ways. The first, and perhaps the simplest, is to postulate that parties are uncertain about the value of γ (which, for simplicity, is supposed to be identical for all individuals.) I postulate that:

A2. Parties believe γ is uniformly distributed on an interval $[\underline{\gamma}, \bar{\gamma}]$.

I further assume that the two parties' policy preferences are given by

$$\left. \begin{aligned} v^L(G) &= G^{1-\beta} \psi(\alpha_L) - \frac{\gamma^{av}}{2} G^2 \\ v^H(G) &= G^{1-\beta} \psi(\alpha_H) - \frac{\gamma^{av}}{2} G^2 \end{aligned} \right\} \quad (5.11)$$

where

$$\gamma^{av} \equiv \frac{\gamma + \bar{\gamma}}{2}.$$

$v^L(\cdot)$ and $v^H(\cdot)$ are von Neumann-Morgenstern utilities. Each party shall seek to maximize its expected utility, given the policy the other party proposes. To compute the expected utilities, we need to compute $\pi(G_L, G_H)$, the probability, from the parties' viewpoint, that a policy G_L will defeat a policy G_H in an election.

Suppose two policies G_L and G_H are proposed, with $G_H > G_L$. If γ is the true value of the coefficient in voters' preferences, then the set of voters who will vote for G_L over G_H is

$$\left\{ \alpha \left| G_L^{1-\beta} \psi(\alpha) - \frac{\gamma}{2} G_L^2 > G_H^{1-\beta} \psi(\alpha) - \frac{\gamma}{2} G_H^2 \right. \right\},$$

which can be written

$$W(G_L, G_H; \gamma) = \left\{ \alpha \left| \psi(\alpha) < \frac{\gamma}{2Q(G_L, G_H)} \right. \right\}, \quad (5.12)$$

where $Q(G_L, G_H) \equiv \frac{G_H^{1-\beta} - G_L^{1-\beta}}{G_H^2 - G_L^2}$. Now $W(G_L, G_H, \gamma)$ comprises at least one-half the voters if and only if

$$\psi(m) < \frac{\gamma}{2Q(G_L, G_H)}, \quad (5.13)$$

because $\psi(\alpha)$ is increasing in α . Thus, G_L defeats G_H iff

$$\gamma < 2\psi(m)Q(G_L, G_H). \quad (5.14)$$

Now the parties' prior is that γ is uniformly distributed on $[\underline{\gamma}, \bar{\gamma}]$; thus the probability they assign to the event that G_L defeats G_H is

$$\pi(G_L, G_H) = \frac{\bar{\gamma} - 2\psi(m)Q(G_L, G_H)}{\bar{\gamma} - \gamma}. \quad (5.15)$$

[Here, I assume that $2\psi(m)Q(G_L, G_H) \in [\underline{\gamma}, \bar{\gamma}]$, which will indeed, always be the case in equilibrium.]

We can now define the payoff functions of the parties as:

$$\begin{aligned} \Pi^L(G_L, G_H) &= \pi(G_L, G_H)v^L(G_L) + (1 - \pi(G_L, G_H))v^L(G_H), \\ \Pi^H(G_L, G_H) &= \pi(G_L, G_H)v^H(G_L) + (1 - \pi(G_L, G_H))v^H(G_H); \end{aligned} \quad (5.16)$$

that is, a party's payoff is its expected utility.

We define:

DEFINITION 2. A *political equilibrium with partisan parties* is a Nash equilibrium $(\bar{G}_L, \bar{G}_H)$ of the game between players with payoffs $\Pi^L(\cdot, \cdot)$ and $\Pi^H(\cdot, \cdot)$.

Thus, a political-economic environment is specified by a tuple $\langle m, \mu, f(\cdot), \alpha_L, \alpha_H, \delta, \beta, \gamma, \bar{\gamma} \rangle$. Knowing this tuple, we can compute the equilibrium of the political economy described.

Our exercise will be to compare the political economic equilibrium of the environment $\langle m, \mu, f(\cdot), \alpha_L, \alpha_H, \delta, \beta, \gamma, \underline{\gamma}, \bar{\gamma} \rangle$ with the equilibrium of environments with more equal distributions of share ownership. In particular, I shall calculate the equilibrium of environments of the form $\langle m, \mu, \lambda f(\alpha) + (1 - \lambda)1(\alpha), \alpha_L, \alpha_H, \delta, \beta, \gamma, \bar{\gamma}, \gamma \rangle$, where $1(\alpha)$ is the function identically equal to one, and $\lambda \in [0, 1]$. When $\lambda = 1$, we have the original (inegalitarian) environment; when $\lambda = 0$, we have complete equality of share ownership. As λ travels from 1 to 0, the distribution of shares becomes more equal.

As I wrote in the introduction, all Pareto-efficient allocations of the economy are those which maximize the surplus

$$\int_{\Omega} x(\alpha)\sigma(\alpha)d\alpha - \frac{\gamma}{2}G^2,$$

where $x(\alpha)$ is the consumption of type α . This is equal to

$$\delta G^{1-\beta} \mu^{\beta} - \frac{\gamma}{2}G^2.$$

The unique level of the public bad associated with Pareto efficiency is therefore

$$\tilde{G} = \left(\frac{(1-\beta)\mu^{\beta}}{\gamma} \right)^{\frac{1}{1+\beta}}. \quad (5.17)$$

I next discuss the choice of parameters for the simulations. I chose $\sigma(\alpha)$ and $\zeta(\alpha)$ to be exponential distributions:

$$\sigma(\alpha) = \lambda_1 e^{-\lambda_1 \alpha}, \quad \lambda_1 = \frac{1}{40};$$

$$\zeta(\alpha) = \lambda_2 e^{-\lambda_2 \alpha}, \quad \lambda_2 = \frac{1}{120};$$

σ has mean 40 and median 27.7, approximately the mean and median of household income in the United States in 1990 in thousands of dollars. $\zeta(\cdot)$ has a median of 83; thus, one-half of all corporate stock is owned by individuals with income greater than \$83,000. I chose $\beta = \frac{3}{4}$, which gives a wage share in total income of 75 percent. I chose $\alpha_L = 30$ and $\alpha_H = 92$; thus, the "Left" party "represents" a voter whose income is between the median and the mean, and the "Right" party represents a voter at the ninetieth centile of the wage distribution.

We may compute that the value of G in political equilibrium depends only on $\frac{\gamma}{\delta}$. To this end, note that we may write

$$\psi(\alpha) = \delta \bar{\psi}(\alpha), \text{ where}$$

$$\bar{\psi}(\alpha) = \frac{\alpha}{\mu^{1-\beta}} + f(\alpha)\mu^\beta(1-\beta),$$

and then

$$\pi(G_L, G_H) = \frac{\frac{\bar{\gamma}}{\delta} - 2\bar{\psi}(\alpha)Q(G_L, G_H)}{\frac{\Delta\gamma}{\delta}},$$

where

$$\Delta\gamma \equiv \bar{\gamma} - \underline{\gamma},$$

and

$$\frac{\Pi^H}{\delta} = \pi \left(G_L^{1-\beta} \bar{\psi}(\alpha) - \frac{\gamma}{2\delta} G_L^2 \right) + (1-\pi) \left(G_H^{1-\beta} \bar{\psi}(\alpha) - \frac{\gamma}{2\delta} G_H^2 \right),$$

with a similar formula for $\frac{\Pi^L}{\delta}$. Thus, so far as political equilibrium is concerned, we may normalize by choosing $\delta = 1$, and then interpret γ as $\frac{\gamma}{\delta}$. This leaves, finally, the choice of $\bar{\gamma}$, $\underline{\gamma}$ and γ .

Table 5.1

λ	G_L^*	G_H^*	G_H^{opt}	π	$E(G^*)$	$\tilde{G}$
1.0	1.16	1.71	3.07	0.246	1.58	1.75
0.9	1.17	1.73	3.06	0.245	1.59	1.75
0.8	1.18	1.74	3.05	0.244	1.60	1.75
0.7	1.19	1.75	3.04	0.243	1.62	1.75
0.6	1.20	1.76	3.03	0.242	1.63	1.75
0.5	1.21	1.78	3.03	0.241	1.64	1.75
0.4	1.22	1.79	3.02	0.241	1.65	1.75
0.3	1.23	1.80	3.01	0.240	1.66	1.75
0.2	1.24	1.81	3.00	0.239	1.68	1.75
0.1	1.25	1.82	2.99	0.238	1.69	1.75
0.0	1.26	1.84	2.98	0.238	1.70	1.75

I chose many values of $\bar{\gamma}$, $\underline{\gamma}$ and γ , and all simulations had the same character. A typical simulation is represented in table 5.1, where $\bar{\gamma} = 2$, $\underline{\gamma} = 1$, and $\gamma = 1.5$. (G_L^* , G_H^*) are the political equilibrium values of the policies of the two parties; π is the probability that L defeats H ; $E(G^*)$ is the expected value of the public bad (i.e., $\pi G_L^* + (1 - \pi) G_H^*$), and G_H^{opt} is the ideal value of the public bad for the H party (i.e., the α_H voter). $\tilde{G}$ is the Pareto efficient level of the public bad.

The first observation is that as the distribution of shares becomes more egalitarian (λ decreases), $E(G^*)$ increases. Note, in fact, that G_L^* and G_H^* both increase as λ falls. This is a surprise: I expected G_L^* to increase, as the voter α_L is getting more shares as λ decreases, but I also expected G_H^* to fall as the voter α_H is losing shares. In fact, we see that the *ideal* value of G for α_H does decrease as λ decreases: but the *equilibrium* value of G_H^* increases. The explanation is obvious: the H -party is able to propose a higher value of G as the L -party represents a richer voter, because the L -party becomes less radical! This deradicalization of the L -party more than compensates for the fact that α_L 's ideal value of G is falling as λ decreases.

I now turn to the second (promised) way of introducing party uncertainty over voter behavior. I shall suppose that parties are uncertain about the value of λ_1 , that is, about the true distribution of skills (wages). I shall suppose that parties believe that:

A3. λ_1 is uniformly distributed on an interval $[\bar{\lambda}_1 - \varepsilon, \bar{\lambda}_1 + \varepsilon]$.

Let $m(\lambda_1)$ denote the median of the exponential distribution with parameter λ_1 : in fact,

$$m(\lambda_1) = \frac{\text{Log } 2}{\lambda_1}.$$

Because $\psi(\alpha)$ is an increasing function it has an inverse, denoted ψ^{-1} . From (5.13), we may write that G_L defeats G_H exactly when

$$m < \psi^{-1}\left(\frac{\gamma}{2Q(G_L, G_H)}\right). \quad (5.18a)$$

Using the formula for $m(\lambda_1)$, (5.17) is equivalent to

$$\lambda_1 > \frac{\text{Log } 2}{\psi^{-1}\left(\frac{\gamma}{2Q}\right)}. \quad (5.18b)$$

Because $\lambda_1 \sim U[\bar{\lambda}_1 - \varepsilon, \lambda_1 + \varepsilon]$, the probability of the event that G_L defeats G_H is

$$\pi(G_L, G_H) = \frac{\bar{\lambda}_1 + \varepsilon - \frac{\text{Log } 2}{\psi^{-1}\left(\frac{\gamma}{2Q}\right)}}{2\varepsilon}. \quad (5.19)$$

With an expression for $\pi(\cdot, \cdot)$, we can now write down precise expressions for the payoff functions Π^L and Π^H .

It is somewhat trickier to solve for political equilibrium this time, because the function ψ^{-1} is not expressible in closed form. My technique was to simulate a polynomial approximation to the function ψ^{-1} .

This time, I simulated equilibrium just for the two polar cases, of $\lambda = 1$ and $\lambda = 0$. For $\lambda = 0$, we can indeed compute, in closed form:

$$\psi^{-1}[X] = \mu^{1-\beta}[X - \mu^\beta(1-\beta)].$$

I again simulated the problem for various values of $\bar{\lambda}_1$, ε , and γ , always with the same qualitative result, that the expected value of the public bad in political-economic equilibrium increased with the egalitarian distribution of shares. For example table 5.2 reports the results when $\lambda_1 = \frac{1}{40} = 0.025$, $\varepsilon = 0.01$, $\gamma = 1.5$.

4 Conclusion

In this chapter, I check my claim, made in recent writings, that an egalitarian distribution of profit shares would decrease the democratically regulated levels of

Table 5.2

λ	G_L^*	G_H^*	π	$E(G^*)$	$\tilde{G}$
1	1.13	1.72	0.24	1.58	1.75
0	1.28	1.80	0.18	1.71	1.75

certain public bads, namely, ones that can be viewed as inputs in the production functions of firms. In contrast to those recent writings, I here introduce a political model in which parties representing the “poor” and the “rich” compete for votes; the level of the public bad is settled by an election. I find that, in a series of simulations, the expected level of the public bad will *increase* with egalitarian redistributions of profit shares. Although with such redistribution the ideal level of the public bad for the rich decreases and the ideal level for the poor increases, it turns out that the net effect is that the level of the public bad in political equilibrium increases.

It is, however, the case that the higher level of the public bad that accompanies egalitarian redistributions renders the economy more efficient. I have questioned the salience of that efficiency, however, in the introduction.

Note

In many writings, János Kornai has studied, with great insight, the political economy of market socialism, as it was practiced in Hungary and other Eastern European countries. This contribution to his *Festschrift* is written in the spirit of those inquiries.

1. Robert Rowthorn (forthcoming) has criticized my use of quasi-linear utility functions, and demonstrates that, with certain other “reasonable” utility functions, the individual’s ideal level of the public bad will *decrease* with his income. With such a Rowthornian specification of utility, I suppose that my result would be reversed. But I am not sanguine about substituting one of Rowthorne’s utility functions for the one I use, for I believe it to be the case that, in actual capitalist economies, it is capitalists (i.e. those whose incomes comes massively from profits) who fight most sharply for deregulation, allowing increased levels of profit-inducing public bads. Despite the popular image of the environmentally insensitive blue-collar worker, I believe that the evidence shows that the highest levels of these public bads are desired by those who derive massive income from profits. This is consistent with my assumptions.

References

- Akerlof, G., and W. Dickens (1982). The economic consequences of cognitive dissonance. *American Economic Review* 72: 307–319.
- Roemer, J. E. (1993). Would economic democracy decrease the amount of public bads? *Scandinavian Journal of Economics* 95: 227–238.
- Roemer, J. E. (1994a). *A Future for Socialism*. Cambridge, Mass.: Harvard University Press.
- Roemer, J. E. (1994b). A theory of policy differentiation in single-issue electoral politics. *Social Choice and Welfare* 11: 355–380.

Roemer, J. E. (1994c). On the relationship between economic development and political democracy. *Economic Design* 1: 15–39.

Roemer, J. E. (1996). Limitized privatization in the presence of public bads. In J. Roemer, ed., *Property Relations, Incentives, and Welfare*. London: Macmillan.

Rowthorn, R. (1996). Would economic democracy decrease the amount of public bads? Comment. *Scandinavian Journal of Economics* 98: 315–319.

6 Parable and Realism in Overlapping Cohorts Models

András Simonovits

1 Introduction

In modern economics, models play an outstanding role. By definition, every model is only a simplified picture of the reality. Whether a model is good or bad, depends on the purpose of the analysis. The theoretical model-builder has plenty of freedom, he can abstract away from many important features of the reality. What matters is only whether the model can shed new light on certain phenomena.

In his famous paper “The Methodology of Positive Economics,” Friedman (1953) asserted that the descriptive realism of economic and other models is much less important than their forecasting power. As a result, extremely simplified models, which are useful only for the rudest forecasts, have become fashionable. Following Samuelson (1962), I shall refer to them as *parables*.

In contrast, in his seminal essay, Koopmans (1957) warned of the danger of arbitrary assumptions. He argued that economic reality is so complex that forecasts of phenomena and processes are seldom possible, whereas we are frequently only looking for their explanations. And then the realism of the model becomes very important. It will be obvious that I agree with Koopmans and disagree with Friedman.

In this chapter I will distinguish the following two types of criticisms of the neo-classical economics: external and internal. *External criticism* questions the relevance of the qualitative assumptions (for example, optimization) and of the basic concepts (for example, equilibrium) of the foregoing theory. *Internal criticism* accepts the framework but emphasizes quantitative deficiencies (for example, the value of the elasticity of substitution is too high in a given model). The two approaches are more complementary than competing and both prefer relevance to elegance.

János Kornai has always been very interested in the relation of models and reality in economics. In one of his most important books, the *Anti-Equilibrium* (Kornai 1971), he criticized neoclassical economics for its lack of realism.

I was lucky to work in Kornai’s research group between 1969 and 1992. I have experienced during the joint work that Kornai spent a large part of his efforts to ensure that the model closely reflect the reality. Using Samuelson’s (1962) famous metaphor, “*Realism is preferred to parables.*”

In this chapter I try to give an internal critique of a special field of general equilibrium theory. I will analyze the problems of parables and realism in the context of *overlapping cohorts*. I use the word *cohort* rather than *generation* because I found

the former to be more precise than the latter: I want to stress that to increase realism, generations should be broken down into cohorts. For short, I shall use the abbreviations OLC and OLG, respectively.

As is known, OLG and OLC models are used to explain the burning issues of pension problems as well as the highly theoretical issues of general equilibrium models.

It is assumed that the reader already knows the foregoing field, thus very few definitions and explanations are given. (Perhaps the best introduction is Gale 1973.) No pretension is made to completeness. Moreover, the choice of the topics reflects my research interests.

I shall argue that theoretical economists, building these models, have defined many useful concepts which can also be used in practical analysis (section 2). However, the assumptions underlying these models (section 3) are so stringent that the results (section 4) obtained with their help can only rarely be used for practical purposes. Much further work need to be done to remove the hindrances from applicability.

Finally, let me reinforce my arguments with a quote from Kenneth Arrow and Seppo Honkapohja (1985, p. 26): "Two general suggestions about methodology of economic theory were made [at the symposium]: the first was that, in view of the number of cases in which theoretical results were non-existent or too difficult to find, reliance should be placed to a greater extent on numerical simulation of models for various values of the parameters. Simulations would also give a greater feel for the sensitivity of results to specific numerical assumptions. Closely related was the urging that, when theorists model a specific area, they should give some indication of the range of parameter values they would regard as basically consistent with the intended applicability of the model."

2 Concepts

First I try to summarize the concepts introduced in OLC models.

1. In contrast to traditional macroeconomics, Modigliani and others (for example, Modigliani and Brumberg 1954) stressed that individual consumption and earning depend on *age*, this dependence is described by consumption and earning *profiles*. If the society's demographic composition changes with time, so do macroconsumption and macrosaving.

2. Consumption changes much less with age than earning does: children and pensioners do not earn but still consume (for example, Tobin 1967).

3. This discrepancy between consumption and earning profiles and the limited storability of our products necessitate an important *redistribution of income* in the framework of family, private savings, and social security (see Samuelson 1958 for an exchange economy; Diamond 1965 for a production economy). In the forthcoming we shall assume that there is a single homogeneous product which can only be used in the given time period.

4. *Individual budget constraint* (longitudinal) can be described by the equality of present value of life-time consumptions plus bequest to present value of lifetime earnings. *Social budget constraint* (cross-sectional) is given by zero total savings in a closed model. (In an open economy, non-zero total savings result in accumulated external credit or debt.)

5. The *internal rate of return* of a social security system is an interest rate which equalizes the present value of total contributions to that of total benefits.

6. Generalizing the practice of life insurance, a *social insurance* company can be envisioned which replaces deterministic individual budget constraints with expected values of stochastic individual budget constraints (Yaari 1965).

Creating these concepts, theorists have been busy in stating ‘nice’ theorems. They have often picked up the simplest combination of the simplest assumptions (as we will see) to prove the simplest theorems. First I will list the most important assumptions used by OLG/OLG models, then I will present some typical theorems.

3 Assumptions

I will here list and comment on the most important assumptions used in OLC models.

A1 The exogenous variables are time-invariant.

The key exogenous variables of OLC models are as follows: number of cohorts, growth rate of number of births, survival rates of cohorts, earning profiles of cohorts, growth rate of earnings, interest rate, and so on. It is normal that theoretical models are time-invariant, but in our case time-invariance is very far from the truth. In the last two centuries birth rates have been dramatically falling in the industrialized world (including Eastern Europe and Southeast Asia), survival rates have been improving in the developed world (excluding Eastern Europe). Economic trends are always changing abruptly. Remember that certain demographic parameters depend on the pension system; thus they should be treated endogenous variables (for example, Becker and Barro 1988).

A2 People live for n periods, where n is usually equal to 2, seldom to 3 or infinity.

Obviously, 2 is the minimal number where there is some overlap between cohorts. However, two-cohort models are very restrictive, equalizing the length of one period to twenty-five to thirty years. I suspect that the neglect of multicohort dynamics was partly connected with the misinterpretation of the following interesting result of Balasko et al. (1980): any multicohort model can be reduced to a two-cohort model with an appropriate enlargement of the commodity space, although cycles may collapse to steady states or shorter cycles in the process. Unfortunately, many researchers have drawn the wrong conclusion that it is sufficient to consider two-cohort, one-good models.

The alternative assumption of unlimited lifespan also generates serious doubts (see A4 below).

A3 Each cohort is homogeneous.

Modern macroeconomists generally take great effort to base macroeconomic relations on individual behavior, but they often stop at a *representative agent*. (Kirman 1992 gives a penetrating critique of this approach). The introduction of n representative agents of n overlapping cohorts is already a great improvement over the model of one ageless agent. Concentrating on age-specific variables, however, macrotheorists are inclined to forget other equally important real-life dimensions: within each cohort there are women and men, there are the rich and the poor, there are thrifty and thriftless individuals, and so on. Especially in the framework of public finance the lack of distinction of income groups make many issues (e.g., solidarity, free riding, and so on) almost meaningless. It is noteworthy that one of the main criticisms of Kornai (1971) against mainstream economics was just the neglect of heterogeneity. Note that heterogeneity also plays an important role in Hildenbrand's (1983) derivation of well-behaving macro relations.

A4 Survival rates are exogenous.

Survival rates are the basic parameters of demographic and life insurance models. However, most OLC models simply neglect survival rates, contributing to the underestimation of the importance of social insurance. Although there are few papers working with arbitrary or realistic survival rates (for example, Yaari 1965; Leung 1994), there is a new tendency to apply the exponential distribution in this context (Blanchard 1985; Blanchard and Fisher 1989). That assumption has gained much popularity, because then the time-invariant death rate is simply added to the discount rate in continuous-time models or the survival factor is simply multiplied by the discount factor in discrete-time models (see A7). This assumption is very elegant,

but it is only appropriate for atoms, not for humans. Its name, *perpetual youth* uncovers its absurdity: having lived long enough, a million-year-old person has the same survival rate as the reader!

A5 Labor force participation rates are exogenous.

Participation rates are obviously very important in the pension problem: the higher the participation rates are, the lower the social burden of pensions is. There is a strongly held and debated view that the introduction of the pay-as-you-go (PAYG) system diminishes the participation rate of the old cohorts with respect to the capital reserve (CR) system (Feldstein 1974). Note also that increased participation rates may have decreased average real wages in the United States in the 1980s and decreased participation rates may have recently increased average real wages in Hungary (see A6).

There are several theoretical models where participation rates depend on the consumption-leisure choice (for example, Grandmont 1985), but the effects of unemployment and underground economy have been neglected so far.

A6 Wage-dependence on age (wage profile) is exogenous.

It is well known that in a given year the earnings are growing with the age. (In contrast to the earning profile, the earning path combines the changes due to age and progress.) Although this dependence is important, it usually degenerates into *constant* or *exponentially decreasing* profile. The first simplification fits the prevalent two-cohort models when only the young earns. The second simplification fits the perpetual youth models (see A2 and A4).

On the other hand, the existing few realistic models work with realistic wage profiles; moreover, children and pensioners have zero earnings. This may appear as a minor question, but it is important in explaining the source of child care and old-age pensions, and it causes complications in establishing the existence of balanced steady states.

A7 Consumers maximize lifetime utility given by a discounted time-additive utility function.

(1) *Time-additivity* of the utility function is frequently assumed, at least for the multicohort case (A2). (2) *Discounting* in time-additive utility functions is a basic building block of neoclassical economics, but its magnitude is elusive. (3) Constant elasticity of intertemporal substitution (CES) or equivalently under (1), constant relative risk aversion (CRRA) is a key assumption and its magnitude may be very important.

Modern macroeconomics is very proud of having replaced ad hoc behavioral rules by optimally determined decision. On the other hand, optimizers frequently neglect another important feature of microeconomics: competitive markets presuppose many different types of agents (in contrast to A3), yielding complex relations (Kirman 1992).

For a moment I leave the realm of internal critique and note that Kornai (1971) questioned the universal validity of optimization in economics. Diamond (1985) gave a penetrating criticism of life-cycle optimization.

A8 Supporting children is costly.

The bulk of the literature follows Samuelson (1975, p. 337) who acknowledged the existence of the problem but added, "childhood dependency is intrinsically less costly relative to old-age dependency."

There are two ways to include children's consumption into the model (Tobin 1967): (1) to start the analysis at age zero instead of age twenty years; (2) to retain twenty years as a starting point but replace individuals with families. Among others, Augusztinovic (1992) championed the first approach, Lee (1980) the second.

A9 Agents have rational expectations (on future interest and growth rates).

As nonoptimization models have become less and less acceptable to the mainstream economics (A7), so have 'nonrational' expectations lost their application.

Although the case of *rational expectations* is certainly a useful benchmark, it has theoretical and empirical pitfalls. In our framework, (1) it leads to the absurdity that the present interest rate was determined decades earlier (called *indeterminacy*, Laitner 1981, 1984); (2) it puts all burden of adjustment on the youngest cohort's consumption; (3) it excludes learning (see Kornai and Diamond's critique in A7).

A10 Zero bequest condition holds.

This is a very elegant assumption but patently untrue (it presupposes either zero death risk or perfect life insurance, selfishness, and so on). Note that it has a decisive role in making PAYG to be responsible for the fall in savings (Blanchard and Fisher 1989, chapter 3). There is an ongoing debate whether bequest is important (Kotlikoff and Summers 1981) or not (Modigliani 1988).

It is of interest that A4, A9, and A10 are connected: in reality neither lifespan nor earnings are known in advance, thus accidental bequests may arise (see, for example, Abel 1985)

Deaton (1992) also criticized the life-cycle theory of consumption, for example for neglecting cash-constraints.

4 Theorems

What are the main themes of the literature of overlapping generations or cohorts? Without going into unnecessary detail, we discuss four issues.

p1 Comparing PAYG and CR pension systems in open models.

There are two pure pension systems: the pay-as-you-go and the capital reserve systems. We speak of open models when the interest rate and the growth rate of the economy are exogenously given. Evidently an open model describes an open economy. Open models are mainly used to compare the ranking of PAYG and CR systems. The main result, *Aaron's principle*, goes back to Aaron (1966): PAYG is superior to CR (that is, the former assures higher lifetime utility than the latter) if and only if the growth rate of the economy is higher than the real interest rate. This inequality is called *Aaron's condition*. In my opinion, Aaron considered two different types of PAYG without noticing it.

1. According to the generally accepted interpretation, and the verbal introduction to Aaron (1966) supports this view, Aaron implicitly assumed that the size and the structure of the transfer (that is, child support and pension) systems are given and the consumers (workers and pensioners) can borrow at the given interest rate from the banks. We shall distinguish such a PAYG system by index 1. Aaron found that the introduction of PAYG1 increases the present value of lifetime earnings with respect to the CR system if and only if Aaron's condition holds. It is an unattractive feature of that model that the size and structure of PAYG1 are exogenously given, and in the case of Aaron's condition, the larger the size of PAYG1, the larger is this improvement. However, if transfers from the older to the younger (family allowances, students loans) are taken into account, then children and young workers may consume more than they earn (A6 and A8), through receiving transfers from the extended PAYG1.

2. In a strange way, in his *model* Aaron worked with wages and consumption rather than contributions and pensions. (He even equalized pensions with gross rather than net wages but this did not affect the essence of his proof. This error was also committed by two other distinguished authors, as Lee (1980) remarks, and plays an important role in pension reforms, too.) Following this line of thought, I developed a second approach and determined the optimal PAYG2 system without private savings and borrowing but possibly again with family allowances (Simonovits 1995b).

In both approaches, Aaron's principle loses its validity.

Auerbach and Kotlikoff (1987) built a 55 cohort OLC model with endogenous leisure but neglected intracohort heterogeneity (A3) and death risk (A4). In their well-calibrated model, they were able to analyze important policy issues. They emphasized the paramount importance that contributors realize the dependence of benefits on contributions.

The opponents of PAYG claim that PAYG is socially unjust because rich people live much longer than poor people, thus there is a regression in PAYG. A similar "injustice" arises because women contribute for a shorter period and enjoy benefits for a longer period than men (see A3).

In an interesting paper, Leung (1994) has shown that if there are death risks, if complete life insurance is unavailable or very expensive (A4 does not hold) and if people cannot leave bequests (A10 holds), then at the last stage of life they will experience a cash constraint. This underlies the importance and limited availability of inexpensive life insurance.

More generally, even if the internal rate of return of PAYG is lower than that of CR, its risk is much lower than its counterpart.

In the remaining part I will consider exclusively CR systems, and I will gloss over that the consumption of children and pensioners are financed by transfers. I will consider a closed economy where the interest rate is determined by the condition of zero total savings.

p2 The existence and uniqueness of balanced steady states in closed exchange models.

As is known, in closed models there are two types of steady states: golden rule and balanced. The latter's *existence* was proved by Gale (1973, Part II and 1974) for nonnegative earning profiles (zero wages for children and pensioners), "general" (parameter-free, nonadditive) utility functions, but under the unrealistic assumption of strong intertemporal elasticity of substitution (A7c). He conjectured unicity and proved it for Cobb-Douglas utility. For OLG models, the balanced steady state degenerates into the autarky where there is no trade at all.

Moving from general utility functions toward quasi-CES functions, Kim (1983) provided a generalized existence theorem where medium weak elasticities are excluded. As Simonovits (1995a) added, in this important range either no balanced steady state exists or more than one such point exists for CES functions. Augusztinovics (1992) extended Kim's theorem to the Leontief case. It is of interest that P1 and P2 are connected. In closed systems Aaron's principle appears now as the nonoptimality of *balanced steady states*.

Aiyagari (1988) proved the existence and optimality of balanced steady states under the extreme assumption that the number of periods and the lifespan go to infinity.

P3 Local stability of golden rule or balanced steady states in closed exchange models.

Making a shortcut, Gale (1973, Part I) claimed that the lower of the golden rule and the balanced steady states is *locally stable* for a two-cohort, rational expectations model (A2 and A9). This claim is not true in general, but it is true for CRRA utility function. He made the equivalent of the following conjecture for the multi-cohort model: under rational expectations, the steady state with the lower interest rate is stable.

Perhaps Laitner (1981) and (1984) were the first papers to analyze genuine dynamics in a general multicohort setting. In fact, their author studied and solved the problem of the interconnection of indeterminacy and instability (already emphasized by Samuelson 1958) in a most general setting. If the number of unstable modes (characteristic roots) is equal to the number of undetermined expectation values, then all viable paths are stable and indeterminacy is ruled out. In case the former is smaller or larger than the latter, the problem of either indeterminacy or instability remains, respectively. It is to be underlined that Laitner's solution requires very high precision in calculation, which cannot be assumed for ordinary agents. Indeed, in the 1960s monographs on numerical analysis emphasized that at the Mises iteration, unstable characteristic roots cannot be overcome (Ralston 1965, example 10.2).

Using Gale's shortcut in a multicohort model, with children and pensioners (A6) and CRRA utility function (A7), Molnár and Simonovits (1998) showed that all steady states with nonnegative interest rates are unstable and conjectured instability of steady states with negative interest rates for realistic parameter values. Thus we refuted or undermined Gale's conjecture for rational expectations. Utilizing another idea of Gale (1974), Molnár and Simonovits (1998) considered naive expectations (dropping A9). Rather surprisingly, local stability was proved for a number of special cases, for instance, near autarky, and instability was only found for a negligible domain around singularity, thus validating Gale's message for naive expectations in *typical* cases.

Superiority of naive expectations over rational expectations is not new in the literature. For example, under certain assumptions (e.g., 2-cycles are recognized) corollary 1.2 of Grandmont and Laroque (1990) proved that stability under learning with memory *implies* instability under rational expectations. (Because naive expectations

do not detect 2-cycles, the foregoing result does not apply to Molnár and Simonovits's model.)

P4 Existence of cycles in closed exchange models.

Gale (1973, Part I) discovered optimal *cycles* for a very specific utility function in a two-cohort, rational expectations model. Relying on the newly emerged chaos literature, Benhabib and Day (1982) and Grandmont (1985) analyzed the dynamics of OLG models and obtained many more cycles and *chaos*.

Grandmont (1985, p. 1039) himself considered it "an interesting study . . . to extend the analysis to a more general framework where, for instance, traders would live an arbitrary (finite or infinite) number of periods" (cf. A2). As a reaction, Sims (1986) conjectured that in realistic overlapping cohorts models there is no significant instability. However, with the hindsight of P3, we know that Sims's conjecture is incorrect.

Aiyagari (1989) excluded the existence of 2-cycles for his asymptotic economy. Reichlin (1992) introduced populations with exponentially distributed lives, allowed zero earnings as well and obtained 2-cycles. Neither of these models was well calibrated. As Aiyagari (1989, p. 167, n. 4) noted with self-irony: "any resemblance of these examples to reality is purely coincidental." Simonovits (1997) reconsidered the question in a realistic model and confirmed nonexistence for elastic substitution and proved existence for non-substitution (Leontief).

5 Conclusion

Having discussed concepts, assumptions, and theorems in OLC models, I draw several conclusions.

1. OLC models are handy tools to analyze many theoretical macroeconomic problems: comparison of pension systems, existence, optimality and stability of the steady state of these models, and so on.

2. OLC models use many stringent assumptions that often prevent their practical use.

3. A possible solution may be to build complex and realistic models whose building blocks are simple enough to yield analytic solutions but the full analysis of these models requires computers.

Samuelson's (1962) warning still applies: "If all these cause headaches for those nostalgic for the old time parables of neoclassical writings, we must remind our-

selves that scholars are not born to live an easy existence. We must respect, and appraise, the facts of life.”

Note

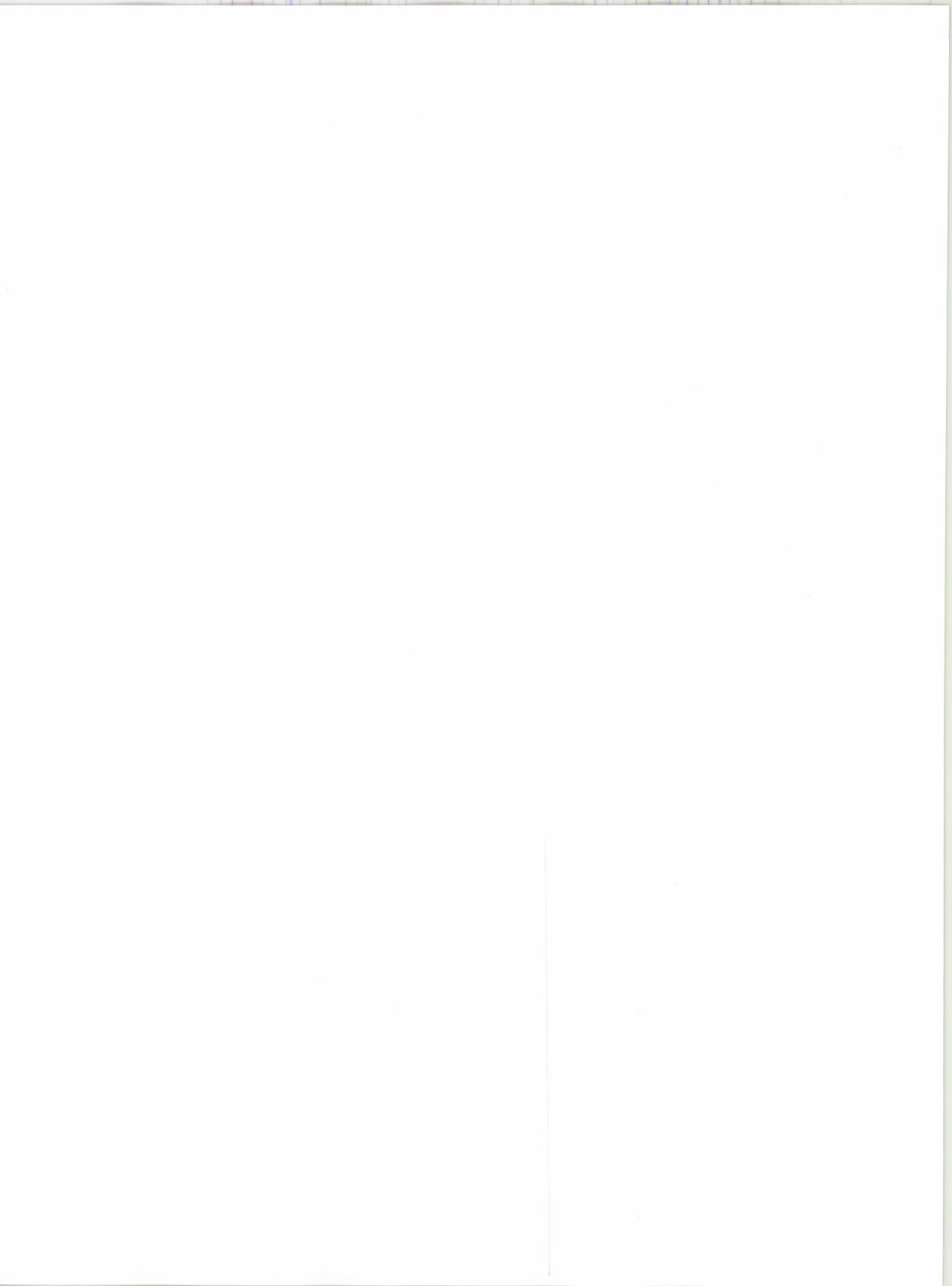
I express my gratitude to Mária Augusztinovics for introducing me into the subject. I had the pleasure to write a joint paper with Gy. Molnár that was a prelude to this work. I am indebted to A. Bródy, J. Brunner, A. Ráftai, E. Siandra, and anonymous referees for their comments on papers underlying this study. J. Gács deserves a special acknowledgment for his thorough critique of the Hungarian version of this paper. None of them is responsible for the content of the paper. Financial support by ACE-Phare Grant Research project No. 095-2139-R and a Fulbright Scholarship at Boston University is acknowledged.

References

- Aaron, H. J. (1966). The social insurance paradox. *Canadian Journal of Economics and Political Science* 32: 371–374.
- Abel, A. (1985). Precautionary saving and accidental bequests. *American Economic Review* 75: 777–791.
- Aiyagari, S. R. (1988). Nonmonetary steady states in stationary overlapping generations models with long lived agents and discounting: multiplicity, optimality, and consumption smoothing. *Journal of Economic Theory* 45: 102–127.
- Aiyagari, S. R. (1989). Can there be short-period deterministic cycles when people are long-lived? *Quarterly Journal of Economics* 104: 163–185.
- Arrow, K. J., and S. Honkapohja, eds. (1985). *Frontiers of Economics*. Oxford: Blackwell.
- Auerbach, A. J., and L. J. Kotlikoff (1987). *Dynamic Fiscal Policy*. Cambridge: Cambridge University Press.
- Augusztinovics, M. (1992). Towards a Theory of Stationary Populations. Manuscript, Institute of Economics, Budapest.
- Balasko, Y., D. Cass and K. Shell (1980). Existence of competitive equilibrium in a general overlapping-generations model. *Journal of Economic Theory* 23: 307–322.
- Becker, G. S., and R. J. Barro (1988). A reformulation of the economic theory of fertility. *Quarterly Journal of Economics* 103: 1–26.
- Benhabib, J., and R. H. Day (1982). A characterization of erratic dynamics in the overlapping generations model. *Journal of Economic Dynamics and Control* 4: 37–55.
- Blanchard, O. J. (1985). Debt, deficits, and finite horizons. *Journal of Political Economy* 93: 223–247.
- Blanchard, O. J., and S. Fischer (1989). *Lectures on Macroeconomics*. Cambridge, Mass.: MIT Press.
- Champsour, P., et al., eds. (1990). *Essays in Honor of Edmund Malinvaud*. Cambridge, Mass.: MIT Press.
- Diamond, P. A. (1965). National debt in a neoclassical growth model. *American Economic Review* 55: 1126–1150.
- Diamond, P. A. (1985). Comment. In Arrow and Honkapohja (1985), 295–306.
- Deaton, A. (1992). *Understanding Consumption*. Oxford: Clarendon Press.
- Feldstein, M. (1974). Social security, induced retirement and aggregate capital accumulation. *Journal of Political Economy* 82: 905–926.

- Fellner, W., et al. (1967). *Ten Economic Studies in the Tradition of Irving Fisher*. New York: Wiley.
- Friedman, M. (1953). *Essays in Positive Economics*. Chicago: university of Chicago Press.
- Gale, D. (1973). Pure exchange equilibrium of dynamic economic models. *Journal of Economic Theory* 6: 12–36.
- Gale, D. (1974). The trade imbalance story. *Journal of International Economics* 4: 119–137.
- Grandmont, J.-M. (1985). On endogenous business cycles. *Econometrica* 45: 995–1045.
- Grandmont, J.-M., and G. Laroque (1990). Stability, expectations and predetermined variables. In Champ-sour et al. (1990), pp. 71–92.
- Hildenbrand, W. (1983). On the law of demand. *Econometrica* 51: 997–1019.
- Kim, O. (1983) Balanced equilibrium in a consumption loans model. *Journal of Economic Theory* 29: 339–346.
- Kirman, A. (1992). Whom or what does the representative individual represent? *Journal of Economic Perspectives* 6: 117–136.
- Koopmans, T. (1957). *Three Essays on the Present State of Economic Science*. New York: McGraw-Hill.
- Kornai, J. (1971). *Anti-Equilibrium*. Amsterdam: North-Holland.
- Kotlikoff, L., and L. H. Summers (1981). The role of intergenerational transfers in aggregate capital accumulation. *Journal of Political Economy* 89: 706–739.
- Kurihara, K. K., ed. (1954). *Post-Keynesian Economics*. New Brunswick, N. J.: Rutgers University Press.
- Lee, R. (1980). Age structure, intergenerational transfers and economic growth: An Overview. *Revue Économique* 31: 1129–1156.
- Laitner, J. P. (1981). The stability of steady states in perfect foresight models. *Econometrica* 49: 319–333.
- Laitner, J. P. (1984). Transition time paths for overlapping-generations models. *Journal of Economics Dynamics and Control* 7: 111–129.
- Leung, S. F. (1994). Uncertain lifetime, the theory of the consumer, and the life cycle hypothesis. *Econometrica* 62: 1233–1239.
- Modigliani, F., and R. Brumberg (1954). Utility analysis and the consumption function: an interpretation of cross-section data. In K. K. Kurihara, (1954), pp. 388–436.
- Modigliani, F. (1988). The role of intergenerational transfers and life-cycle saving in accumulation of wealth. *Journal of Economic Perspectives* 2: 15–40.
- Molnár, G., and A. Simonovits (1998). Expectations, (in) stability and (in) viability in realistic overlapping cohorts models. *Journal of Economic Dynamics and Control* 23: 303–332.
- Ralston, A. (1965). *A First Course in Numerical Analysis*. New York: McGraw Hill.
- Reichlin, P. (1992). Endogenous cycles with long-lived agents. *Journal of Economic Dynamics and Control* 16: 243–266.
- Samuelson, P. A. (1958). An exact consumption-loan model of interest with or without the social contrivance of money. *Journal of Political Economy* 66: 467–482.
- Samuelson, P. A. (1962). Parable and realism in capital theory: the surrogate production function. *Review of Economic Studies* 39: 193–206.
- Samuelson, P. A. (1975). The optimum growth rate for population. *International Economic Review* 16: 531–537.
- Simonovits, A. (1995a). On the number of balanced steady states in overlapping cohorts model. *Acta Oeconomica* 47: 51–67.
- Simonovits, A. (1995b). Pensions and family allowances: a reconsideration of the social insurance paradox. *Acta Oeconomica* 47: 337–347.

- Simonovits, A. (1997). Are There Endogeneous Cycles in a Realistic Overlapping Cohorts Model? Working Paper, Institute of Economics, Budapest.
- Sims, C. A. (1986). Comments. In Sonnenschein, (1986), pp. 37–39.
- Sonnenschein, H. F., ed. (1986). *Models of Economic Dynamics*. New York: Springer.
- Tobin, J. (1967). Life cycle saving and balanced growth. In Fellner et al. (1967), pp. 231–256.
- Yaari, M. E. (1965). Uncertain lifetime, life insurance and the theory of consumer. *Review of Economic Studies* 32: 137–150.



Jörgen W. Weibull

1 Introduction

János Kornai has shown a long-standing interest in the relationship between a firm's external conditions—such as the degree of administrative regulation, state intervention, and competition—and its internal efficiency. The present study considers this relationship with regard to competition and market integration.

Policy discussions concerning internal efficiency and competition frequently presume that market integration enhances the internal efficiency in the participating firms. In fact, this is one of the basic policy arguments for the European union. The European Commission (1988) writes that “the new competitive pressures brought about by the completion of the internal market can be expected to . . . produce appreciable gains in internal efficiency” (p. 126). Such a view indeed has support from Adam Smith: “good management . . . can never be universally established but in consequence of the free and universal competition which forces every body to have recourse to it for the sake of self-defence” (pp. 163–164).

By contrast, current standard microeconomics textbooks do not deal with this issue: all firms by assumption operate at maximal internal efficiency, irrespective of market conditions. A monopoly firm is internally just as efficiently organized as is a firm in an oligopolistic or perfectly competitive market. However, recent theoretical studies of this issue have shown that if the basic model of the firm is expanded to include an agency problem then market conditions may indeed influence the internal efficiency of a firm. More specifically, the incentive power of equilibrium contracts between owners and management may depend on external market conditions. As a result, managerial efforts to improve the internal efficiency of the firm may change when market conditions change, though not always in the expected direction, see Holmström (1979), Hart (1983), Nalebuff and Stiglitz (1983), Scharfstein (1988), Hermalin (1992), Martin (1993), Horn, Lang, and Lundgren (1995), and Schmidt (1994).

The term *internal efficiency* used in this literature is perhaps unfortunate, and was originally kept undefined. In his seminal paper on this topic, Leibenstein (1966) avoided giving definition but introduced the term *X-inefficiency*. Broadly speaking, “*internal efficiency*” and “*X-efficiency*” were used much in the same way as the layman would use the phrase “good management.” Here these terms will be understood in the narrow sense of “low production costs.” A firm that operates at a lower cost at all output levels than another firm will be called “internally more efficient.” In particular, “internal efficiency” has little, if anything, to do with Pareto efficiency (because it neglects welfare effects on managers).

In contrast to the cited information-based approaches, the present study takes the viewpoint that market conditions may influence managers' incentives to improve internal efficiency even in the absence of agency problems. In order to study this possibility in its purest form we thus assume that firms are managed by their owners. Such managers care about profits, but they may also have preferences concerning the effort they put into their firm. An owner who operates his or her own firm may thus face a trade-off between profits and leisure—taken to be the opposite of effort.¹ Exerting more managerial effort—more intense work, longer hours in the office, or less pleasant decisions (such as firing staff)—the owner-*cum*-manager may improve the firm's internal efficiency and thereby its profits. Moreover, this trade-off may depend on market conditions. For instance, under stiffer competition, equilibrium profits may be lower, and the marginal return to increased effort on profit may or may not be higher. The income effect on managerial effort is unambiguously positive in such circumstances if leisure is a normal good. As John Hicks (1935) puts it, "The best of all monopoly profits is the quiet life." However, the total effect also depends on the substitution effect, and, of course, on what exactly is meant by "stiffer competition." In particular, reduction of barriers to entry involves one form of "stiffened competition" while market integration involves another form. Income and substitution effects of trade barriers on managerial incentives, in markets where all firms are price takers, are studied in Corden (1970) and Martin (1978). The present approach can be viewed as an extension of Martin's model to imperfectly competitive markets with and without barriers to entry. The income effect on managerial effort is studied in the context of an agency model in Hermlin (1992).

In terms of analytical tools, the present paper suggests a minor extension of the basic microeconomics textbook model of the firm, perhaps the slightest extension that includes managerial effort as a non-traded input. More exactly, instead of treating a firm's production possibilities as exogenous (and known) to every manager, we here endogenize the production possibility set of a firm by letting it in part depend on its manager's efforts.² The more such efforts the manager makes, the more production opportunities become available to the firm. Indeed, it may be argued that an essential part of management's task is precisely to identify production possibilities available to the firm. This view was advocated by Hayek, who saw economic agent's acquisition of knowledge as a fundamental aspect of an economy: "knowledge . . . is not given to anyone in its totality" (1945, p. 321); "it is only through the process of competition that the facts will be discovered" (1949, p. 96).

Leibenstein's treatment (1966) was essentially informal and empirical. On the basis of his empirical studies he claimed, "The simple fact is that neither individuals nor firms work as hard, nor do they search for information as effectively, as they

could. The importance of motivation and its association with degree of effort and search arises because the relation between inputs and outputs is not a determinate one" (p. 407). He claimed that a significant part of the beneficial effects of competition come about via increased internal efficiency of firms (X-efficiency): "The data suggest that cost reduction that is essentially a result of improvement in X-efficiency is likely to be an important component of the observed residual in economic growth" (p. 408). "Thus we have instances where competitive pressures from other firms or adversity lead to efforts toward cost reduction, and the absence of such pressures tends to cause costs to rise" (pp. 408–409).

The present analysis is performed in a simple formal model, restricted to the case of a Cournot market for a homogenous product, and focuses on two polar cases: (1) a given set of firms in the market, and (2) free entry to and exit from the market. When the set of firms is fixed all managers simultaneously choose their levels of managerial effort. In the case of free entry we imagine a large population of potential entrepreneurs, each of whom first decides whether or not to set up a firm, and thereby become its owner-*cum*-manager, in the studied product market. All entrepreneurs have access to all production factors and inputs, at fixed and given prices. As an alternative to setting up a firm, each entrepreneur has some (unmodeled) outside option. Once the entry decisions have been taken, these decisions become known to all entrants, and these simultaneously choose their levels of managerial effort—just as in the case of a given set of firms. The analysis presumes identical entrepreneurs/managers and is focused on symmetric market equilibria.

It turns out that the comparative statics analysis of how managers adapt their efforts to changed market conditions is formally parallel to classical models of the price-taking consumer. In the case of a fixed set of firms in the market the analysis is similar to the Marshallian demand analysis, whereas under free entry it takes the form of Hicksian demand analysis. In the first case, changed market conditions induce an income and a substitution effect on managers' efforts, whereas in the second case managers are kept at their reservation utility level (the income effect on their effort is "compensated"). In both cases there is a major difference in comparison with classical consumer demand analysis. Instead of a nonstrategic environment (price-taking consumers) we here have a strategic environment (the price of leisure is influenced by each manager's action).

Conditions are identified under which an increased number of firms in a given market induces managers to exert more effort. Hence, stiffer competition so defined leads to increased internal efficiency (lower unit cost of production). A central concern of this study is whether market integration (or trade liberalization) induces

higher internal efficiency. The thought-experiment is simple: put together two markets—"countries"—that before were completely isolated from each other. Conditions are identified under which such a change in external conditions results in increased internal efficiency, both when the number of firms per market (country) is fixed and unaffected by the change, and when there is free entry and exit of firms before and after the integration.

In both cases, consumers benefit more from trade than in the standard Cournot model: to the pressure on the market price from an increased number of competitors is added the effect from reduced production costs in each firm. According to Leibenstein (1966) the second effect is empirically much stronger than the former. Including the welfare effect on entrepreneurs, the net welfare effect of market integration under free entry is unambiguous: consumers face a lower price of the product and all entrepreneurs remain at their reservation utility level. Market integration is thus a Pareto improvement.

By adopting the present approach, which neglects informational problems, I do not suggest that such problems are unimportant for the questions at hand. On the contrary, such problems seem to be of fundamental importance. However, in the spirit of Occam's razor this chapter seeks to find the simplest setting which allows for the possibility that market conditions influence the internal efficiency of firms. Richer and more complex models, such as those based on asymmetric information between owners and managers, can, we hope, be more easily understood and appreciated against the background of such simpler models:

For a discussion of various concepts of efficiency and competition, and the interplay between these, see Vickers (1995). A recent empirical investigation of relations between competition and corporate performance is given in Nickell (1996).

The chapter is organized as follows. Section 2 presents the model and section 3 the analytical results. Numerical results for a parametric special case are given in section 4, which also contrasts Marshallian demand analysis with Hicksian demand analysis of managers' decision. Section 5 concludes with a brief discussion of some potential extensions. Mathematical proofs are not included; they are provided in a companion working paper, Weibull (1996).

2 The Model

The model is developed in two steps. First, the set of firms participating in the product market in question is taken to be fixed. Then entry and exit decisions are introduced in the model.

2.1 The Demand Side

Consider a Cournot product market for a homogeneous good, with n identical firms. The market price p is determined by an inverse demand function, $p = P_m(Q)$, where Q is aggregate output, $Q = \sum_{j=1}^n q_j$. Here m is an exogenous parameter that will be interpreted as the number of identical countries in a free-trade area for the good in question. With linear demand in each country, $D(p) = 1 - p$, and in the absence of transportation costs, aggregate demand in the free-trade area is $mD(p) = m - mp$. Motivated by this example, the following parametric representation will be maintained throughout the analysis:

$$P_m(Q) \equiv 1 - Q/m, \quad (7.1)$$

where $m > 0$.

2.2 Production Costs and Managerial Effort

Each firm is managed by its owner. The key assumption in this study is that managerial efforts can reduce the firm's production costs. The channel from managerial effort to production cost is thought to go via the firm's production possibility set, or, equivalently, via its family of input requirement sets. Here it is assumed that the manager can expand the input requirement set associated with any given output level by exerting more managerial effort. Formally, if $V(q, e)$ is the input requirement set associated with output level q when the manager exerts effort e , then the total cost to produce q is

$$C(q, w, e) = \min_{z \in V(q, e)} w \cdot z, \quad (7.2)$$

where w is the price vector for inputs. The assumption that the manager can expand the input requirement set can be formally stated as: $e < e'$ implies $V(q, e) \subset V(q, e')$ for each output level q . It then follows directly from (7.2) that the production cost $C(q, w, e)$ is nonincreasing in managerial effort e .³

For the sake of analytical simplicity, the subsequent analysis will be focused on the special case when there is no fixed cost, and, at any given level of managerial effort, the firm's unit (or marginal) cost is constant. This unit cost is, moreover, assumed to be continuously decreasing in managerial effort at a nonincreasing rate. The domain of the effort variable is normalized to the unit interval, and the input price vector w is notationally suppressed because this will be held constant:

(A) $C(q, e) \equiv c(e)q$, where $c : [0, 1] \rightarrow [0, 1]$ is a twice continuously differentiable function with $c' > 0$, $c'' \geq 0$, $c(0) = 1$ and $c(1) = 0$.

When aggregate output is Q , price is $p = 1 - Q/m$, by equation (1). Subtracting production costs from revenues, we obtain the following expression for the profit to firm i :

$$\pi_i = \left[1 - \frac{Q}{m} - c(e_i) \right] q_i. \quad (7.3)$$

2.3 Managers' Preferences

The manager of each firm derives utility $u(\pi, e)$ from her firm's profit π and her managerial effort e . The analysis is restricted to utility functions of the following separable form:

(B) $u(\pi, e) \equiv \varphi(\pi) - \psi(e)$, where

(B1) $\varphi : \mathbb{R} \rightarrow \mathbb{R} \cup \{-\infty\}$ is twice differentiable on $\mathbb{R}_{++}$ with $\varphi' > 0$, with relative risk aversion $r_\varphi > \frac{1}{2}$, and such that $\varphi(\pi) = -\infty$ for $\pi \leq 0$.⁴

(B2) $\psi : [0, 1) \rightarrow \mathbb{R}$ is twice differentiable with $\psi', \psi'' > 0$, $\psi(0) = \psi'(0) = 0$, and $\lim_{e \rightarrow 1} \psi(e) = +\infty$.

Note that a manager's utility is assumed to be an increasing function of her firm's profit. This is the case if the manager receives a monetary reward that is an increasing function of her firm's profit, and if she does not consume the product in the studied market (the price of which is endogenous). Moreover, the marginal utility of profit, φ' , is decreasing at a rate that is sufficiently high to keep the relative risk aversion of the subutility function φ above one half (at all positive profit levels). The disutility of managerial effort increases with effort, from zero to plus infinity, as effort goes from its lower to its upper bound. Also the marginal disutility of effort is increasing with effort.

2.4 Solution Concepts

The interaction between the n managers in the product market is modeled as a simultaneous-move game in which each manager i chooses a combination (e_i, q_i) of effort and output in order to maximize her utility $u(\pi_i, e_i)$. The focus will be on symmetric Nash equilibria, that is, Nash equilibria in which all participating managers choose the same output level q and make the same effort e .

In the case of free entry and exit such product market interaction will be embedded as a "postentry" subgame of a larger game that involves an entry stage in which the set of participating firms is determined endogenously. We then imagine an infinite population of identical entrepreneurs, each of whom may set up a firm—become an owner-cum-manager—in the product market in question. The alternative

available to each entrepreneur is to take some outside option that yields a fixed utility level $\bar{u}$. The equilibrium number of participating firms, n , will be treated as a real, rather than integer, variable. Consequently, the equilibrium utility to all entrepreneurs, those who enter and those who stay out, is the same, $\bar{u}$.

3 Analysis

It turns out to be analytically convenient to make a transformation of variables before one embarks on such an analysis. Instead of using each managers effort e_i as a decision variable, we will use her *effective* effort, defined as $x_i = 1 - c(e_i)$. By condition (A), there is a one-to-one relation between e_i and x_i , such that x_i is strictly increasing from zero to one as e_i increases from zero to one. Hence, it is decision-theoretically (and strategically) immaterial if we use e_i or x_i as part of i 's strategy. For later notational convenience we will write b for the inverse c^{-1} to the unit-cost function c .⁵ Hence $e_i = b(1 - x_i)$, where $b(0) = 1$, $b(1) = 0$, $b' < 0$ and $b'' \geq 0$, by condition (A).⁶

Conditions (A) and (B) will be assumed to hold throughout this section.

3.1 Fixed Set of Firms in the Market

As mentioned earlier, the focus is here on symmetric Nash equilibrium in the interaction between the n managers in the product market. Using the above transformation of variables, a *strategy* to manager i is a pair $s_i = (x_i, q_i)$, where x_i is a number in the open unit interval $(0, 1)$ and q_i is a nonnegative number. The *payoff* to manager i , when a strategy profile $s = (s_1, \dots, s_n)$ is played, is

$$U_i(s) = \varphi \left[\left(x_i - \frac{1}{m} \sum_{j=1}^n q_j \right) q_i \right] - \psi[b(1 - x_i)]. \quad (7.4)$$

A strategy profile s is *symmetric* if all managers choose the same strategy, that is, if there exists a pair (x, q) such that $s_i = (x, q)$ for all i . Such a pair (x, q) is said to *represent* the profile s . A symmetric profile s is *interior* if all managers choose a positive output level, $q > 0$. Note that a symmetric Nash equilibrium is necessarily interior by definition of the strategy set available to a manager. For if $q_i = 0$, then no $x_i \in (0, 1)$ is optimal: $U_i(s)$ is then strictly decreasing in $x_i \in (0, 1)$.

A necessary first-order condition for interior Nash equilibrium is, for each $i = 1, \dots, n$:

$$\frac{\partial U_i(s)}{\partial q_i} = 0 \Leftrightarrow q_i = \frac{1}{2} \left(m x_i - \sum_{j \neq i} q_j \right). \quad (7.5)$$

Hence, in a symmetric Nash equilibrium we necessarily have

$$q = \frac{mx}{n+1}, \quad (7.6)$$

a formula familiar from the standard Cournot model with constant unit cost and linear demand. (Set $m = 1$ and $x = 1 - c$.) This is not surprising: because profit has positive marginal utility to the manager, she should, at any effort level that she chooses, adapt her firm's output optimally to its unit (marginal) cost. In symmetric equilibrium all effort levels, and hence unit costs, are the same, and equation (7.6) results.

Equation (7.6) also shows that the more effort managers exert in a symmetric equilibrium, the more output their firms will produce. The exact relation between effort and output depends on market conditions, here represented by market size, m , and the number of firms, n . Market conditions do matter.

Another necessary first-order condition for interior Nash equilibrium is, for each $i = 1, \dots, n$:

$$\frac{\partial U_i(s)}{\partial x_i} = 0 \Leftrightarrow q_i \varphi' \left[\left(x_i - \frac{1}{m} \sum_{j=1}^n q_j \right) q_i \right] + \psi' [b(1 - x_i)] b'(1 - x_i) = 0. \quad (7.7)$$

Combining this with equation (7.6) we obtain the following equation in one variable, the effective managerial effort x :

$$\frac{mx}{n+1} \varphi' \left[m \left(\frac{x}{n+1} \right)^2 \right] + \psi' [b(1 - x)] b'(1 - x) = 0. \quad (7.8)$$

Increased effort has a direct and an indirect effect on utility. The indirect effect comes about via the induced increase in profit, involving optimal adaptation of output. The first term above represents this indirect effect of a marginal increase in effort, and the second term (negative) represents the direct effect. This equation plays a key role in the subsequent analysis.

It is not difficult to show that equation (7.8) has a unique solution x^* in the open unit interval $(0, 1)$. In view of this result the question arises whether the found pair (x^*, q^*) , with q^* determined from x^* in equation (7.6), indeed represents a Nash equilibrium. A sufficient condition for this to hold is that the resulting utility to a manager, $u(x^*, q^*)$, exceeds the utility she would obtain when producing zero and making no effort, $\varphi(0) - \psi(0) = \varphi(0)$: if $u(x^*, q^*) > \varphi(0)$, then there exists exactly one symmetric Nash equilibrium.

Equation (7.8) permits certain comparative statics observations. In particular, it is not difficult to show that its solution, the equilibrium level x^* of effective man-

agerial effort, rises with the number n of firms. An increase in the number n of firms reduces both the equilibrium profits per firm and the marginal return to effort. However, the income effect dominates the substitution effect:

PROPOSITION 1 Equilibrium managerial effort is strictly increasing in the number n of firms in a market of fixed size m .

Increased competition, in this specific sense, enhances firms' internal efficiency. This is, for example, the case if barriers to entry are reduced so that new firms enter the market. Deregulation and privatization policies may clearly have such effects.

When the number of firms in a market increases, then consumers benefit more from increased competition in the present model than they do in the standard Cournot model. On top of the usual beneficial consequence from a larger number of firms operating at given production costs, we here have an incentive effect inside firms that brings down production costs in each firm. Formally, the equilibrium market price, p^* , is given by a convex combination of the unit cost $c(e^*)$ and 1:

$$p^* = \frac{n}{n+1} c(e^*) + \frac{1}{n+1}. \quad (7.9)$$

The weight to the unit cost is larger the more firms there are in the market. Because the unit cost is less than one, the equilibrium price accordingly decreases when the number of firms increases, at any fixed level of managerial effort. In the limit, as the number of firms is infinite, the equilibrium price, at a fixed level of managerial effort, equals the unit cost. Because the equilibrium effort rises with the number of firms, the unit cost in fact decreases with the number of firms in the market. Hence, the "standard" oligopolistic price effect from an increase in the number n of firms is enhanced in the present model.

One would expect that managers' utility falls with an increase in the number of firms in the market. This follows readily from the above observations: manager's equilibrium utility, $v(n, m)$, is a continuous and strictly decreasing function of n (at any given and fixed value of market size m). In view of the downward pressure from an increased number of firms on the market price we conclude that utility is transferred from managers to consumers as the number of firms in the market rises.⁷

3.2 Free Entry and Exit

In equilibrium in the full game with entry and exit no active entrepreneur, that is, an entrepreneur who decided to enter and become the manager-cum-owner of a firm, obtains a utility below her reservation utility $\bar{u}$. In this sense, no active entrepreneur has an incentive to exit. Moreover, the number of firms in the reached

subgame is such that if one more firm had entered, then the resulting equilibrium utility in that subgame would have fallen below $\bar{u}$. In this sense, no passive entrepreneur has an incentive to enter. Because the number of participating firms is here treated as a real, rather than integer, variable, the equilibrium utility to all entrepreneurs, active and passive alike, is exactly $\bar{u}$: a single equation replaces two inequalities. The subsequent analysis concerns only symmetric equilibria in the product market, and it is assumed in this subsection that the outside option is better than setting up a firm and then running it at zero output (and effort) level: $\bar{u} > \varphi(0)$.

As observed above, managers' equilibrium utility level $v(n, m)$ is continuously and strictly decreasing in n . Hence, for any reservation utility level $\bar{u} \in \mathbb{R}$ there exists at most one (real) number $n \geq 1$ of firms such that all entrepreneurs are indifferent between market entry and the outside option. If the reservation utility $\bar{u}$ is too high—above a monopolist's utility level—then no such n exists. Likewise, if the reservation utility $\bar{u}$ is too low—below the utility level of a manager of a firm in a perfectly competitive market—then no such n exists. In the first case the number of firms in the market is zero, and in the latter case it is plus infinity. However, the latter case is excluded because in the limit case of infinitely many firms the utility to a manager is at most $\varphi(0)$, a utility level that by assumption is below $\bar{u}$. Formally, let $n^*(m)$ denote the equilibrium number of firms under free entry and exit. Then $n^*(m) = 0$ if $v(1, m) < \bar{u}$, and otherwise $n^*(m) \geq 1$ is the unique solution to the equation

$$v(n, m) = \bar{u}. \quad (7.10)$$

Not surprisingly, the number of participating firms under free entry and exit increases with the size m of the market: there exists some minimal market size $m^o > 0$ such that $n^*(m) = 0$ for $m < m^o$, $n^*(m^o) = 1$, and $n^*(m)$ is strictly increasing in m for $m > m^o$. Does this rise in the number of firms, when the market is expanded, induce more or less managerial effort?

This question can be answered by way of the following thought-experiment. Imagine that the product market initially is in a symmetric interior Nash equilibrium with n active firms, each producing output quantity q^* , and each manager exerting effective effort x^* . Suppose some manager i contemplates alternative effort/output pairs for herself. If she chooses effort/output pair (x_i, q_i) , while all other firms remain at their equilibrium output level, her firm's profit becomes

$$\pi_i = \left[x_i - \frac{n-1}{m} q^* - \frac{1}{m} q_i \right] q_i. \quad (7.11)$$

Although profit and effort affect her utility directly, her firm's output matters only indirectly to her, via its effect on her firm's profit. Suppose that manager i , given any effective effort x_i that she contemplates to exert, chooses her firm's output level q_i so that her firm's profit is maximized, conditional on her effective effort x_i and under the hypothesis that the other firms produce their equilibrium output q^* . It is easily verified (using (7.6)) that she will then choose

$$q_i = \frac{m}{2} \max \left\{ 0, x_i - \frac{n-1}{n+1} x^* \right\}. \quad (7.12)$$

As expected, the manager will thus choose a higher output level the more effort she has decided to exert. At low levels of effort, her firm's unit cost $c_i = 1 - x_i$ is so high that the optimum output level is zero. From the viewpoint of the resulting utility to the manager, we may without loss of generality assume that she considers only effective effort levels x_i that exceed

$$\frac{n-1}{n+1} x^*.$$

Given such choices of effort, and with optimal adaptation of output to effort, the profit to firm i is the following convex increasing function of its manager's effective effort:

$$\pi_i = \frac{m}{4} \left(x_i - \frac{n-1}{n+1} x^* \right)^2. \quad (7.13)$$

The graph of this function defines the manager's "possibility frontier" in the (π_i, x_i) -plane. These are the best combinations of profit and effective effort available to the manager when all other managers exert their equilibrium effort. In the same plane we may draw indifference curves for manager i . The optimal effective effort for manager i is x^* , a point of tangency between her possibility frontier and one of these indifference curves. Under free entry this indifference curve is determined by managers' reservation utility, that is, by the equation $u[\pi_i, b(1 - x_i)] = \bar{u}$, see figure 7.1.

Market integration here means an increase in the parameter m , accompanied by an increase in the number $n^*(m)$ of firms. Assuming that the reservation utility of managers is unaffected, the effect of market integration is that the tangency point with the reservation utility indifference curve may move. Hence, no income effect is here at work, all hinges on the substitution effect. If the tangency point moves as market size changes, then managerial effort and profit necessarily move in the same direction: either both increase or both decrease (in order to keep utility constant). It turns out that both increase. It is as if managers trade "leisure" for "money" when

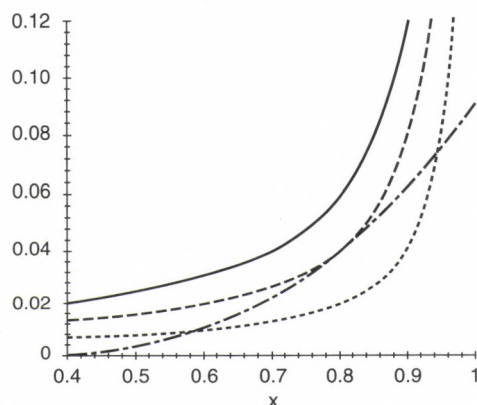


Figure 7.1

The equilibrium "possibility frontier" of manager i , and three indifference curves. (Parametrization as in section 4, with $m = 1$ and $n = 3$.)

the market expands. Consequently, market integration under free entry and exit enhances the internal efficiency of firms.

PROPOSITION 2 Equilibrium managerial effort under free entry and exit is strictly increasing in market size m .

By assumption the utility of all entrepreneurs remains constant under market integration. However, consumers in the product market benefit in two ways. On top of the well-known increased "allocative" efficiency gain due to the increased number of participating firms, resulting in a lower market price, managers work harder and so firms operate under lower costs, adding to the downward pressure on the market price. Granted that consumers' welfare is decreasing in the price of the product in question, we conclude that market integration is a Pareto improvement: a welfare gain for consumers and no welfare loss for managers.

Remark: This conclusion rests, inter alia, on the assumption that managers' reservation utility, $\bar{u}$, is unaffected by the change in market conditions. The qualitative result still holds (by continuity) if $\bar{u}$ increases only slightly. Another presumption is that factor prices (implicit in the definition of the marginal cost) are unaffected.

4 A Parametric Example

The above analysis is particularly simple in the special case of a linear relation between managerial effort and unit cost, combined with Cobb-Douglas preferences.

Let the unit cost $c(e)$ equal one minus effort, $1 - e$, for all effort levels e , and let

$$u(\pi, e) \equiv \log \pi + \lambda \log(1 - e) \quad (7.14)$$

for some $\lambda > 0$.⁸ The residual $z = 1 - e$ may be interpreted as leisure and λ as the intensity in managers' taste for leisure. It is easily verified that conditions (A) and (B) are met. In this special case effective effort and effort coincide: $x = b(1 - e) = 1 - (1 - e) = e$.

4.1 Fixed Number of Firms

The first-order condition (8) has the explicit solution

$$e^* = x^* = \frac{n+1}{n+1+\lambda}. \quad (7.15)$$

A striking feature of this equation is that the market size parameter m is absent. Hence, in this special case equilibrium managerial effort, and hence also the internal efficiency of firms, is independent of market size. As expected, the equilibrium effort level is increasing in the number n of firms and decreasing in managers' taste λ for leisure. Moreover, as the number of firms tend to infinity, the equilibrium managerial effort approaches its upper bound, 1, irrespective of managers' taste $\lambda > 0$ for leisure. Hence, in the limit of perfect competition even the most leisure-loving managers exert maximal effort.

In contrast to effort and price, output and profit per firm do depend on market size,

$$q^* = \frac{m}{n+1+\lambda} \quad \text{and} \quad \pi^* = \frac{m}{(n+1+\lambda)^2}. \quad (7.16)$$

4.2 Free Entry

As expected, the number of firms in the market under free entry is increasing in market size:

$$n^*(m) = \max \left\{ 1, \left[m\lambda^\lambda \exp(-\bar{u}) \right]^{\frac{1}{2+\lambda}} - 1 - \lambda \right\}. \quad (7.17)$$

Note the concavity of the function n^* : market integration results in a reduction of the total number of firms. Expressed in other words: the number of firms per country goes down.

Inserting the expression for $n^*(m)$ in equation (7.17), one obtains the following expressions for managerial effort, $e^{**}(m)$, under free entry and exit:

$$e^{**}(m) = 1 - \left(\frac{\lambda^2 \exp(\bar{u})}{m} \right)^{\frac{1}{2+\lambda}} \quad (7.18)$$

As expected, market integration induces managers to exert more effort.

4.3 Marshall versus Hicks

If we think of the residual $z_i = 1 - e_i$ as the manager's *leisure* and π_i as the *upper bound* on her consumption, then the manager's "budget set" $B(n, m)$ in the leisure/consumption space is obtained from equations (7.13) and (7.15):

$$B(n, m) = \left\{ (z_i, \pi_i) : 0 < z_i \leq \frac{2+\lambda}{n+1+\lambda}, \pi_i \leq \frac{m}{4} \left(\frac{2+\lambda}{n+1+\lambda} - z_i \right)^2 \right\} \\ \cup \left\{ (z_i, \pi_i) : \frac{2+\lambda}{n+1+\lambda} < z_i \leq 1, \pi_i \leq 0 \right\}$$

Figure 7.2 shows the relevant part of this nonconvex set, along with a few indifference curves.

First suppose the number n of firms and market size m both are fixed and given. Any change in market conditions, that is, in these two parameters, results in a shift of the budget set $B(n, m)$, and the effects on managers's demand for leisure—hence, supply of managerial effort—can be studied just as in classical Marshallian demand analysis. The effect on managers' choice of leisure can be decomposed into an

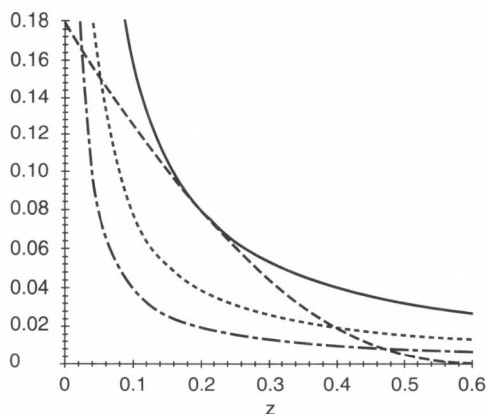


Figure 7.2

The equilibrium "budget set" of manager i , for $n = 1$, $m = 2$, and $n = 3$, and three indifference curves.

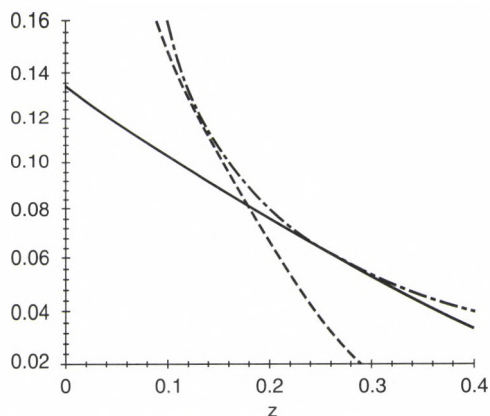


Figure 7.3

The reservation utility indifference curve, and the “possibility frontier” associated with a small market (the flatter curve), and with a large market (the steeper curve), respectively.

“income effect,” that is, a parallel shift of the “budget curve,” and a “substitution” effect, that is a change in the slope of the “budget curve.”

Second, suppose only market size m is fixed and given, while the number of firms is determined by free entry and exit. Then the effects on managers’ choice of leisure from a change in market size can be studied in the spirit of the Hicksian compensated demand analysis. For any change in m is fully compensated by entry and exit of firms—so that all managers remain at their initial utility level. What changes is only the (here non-linear) “relative price” of leisure as against consumption—here given by the “possibility frontier.” Figure 7.3 shows how this frontier changes as the market size m changes. It is as if market integration induces a higher relative price of leisure.

5 Extensions and Variations

The above analysis of the question of whether market conditions may influence the internal efficiency of firms internal efficiency was restricted to a particular setting. Hence, a variety of extensions are called for before robust conclusions can be claimed. Here is a list of a few such extensions. First, the present study has been restricted to managerial incentives to cut production costs. Incentives to promote product quality and implement useful technical innovations are highly relevant potential extensions. In such a setting the present description of “entrepreneurs”

may be enriched. Second, the present analysis rests on the heroic assumption that demand is linear. Are the qualitative results valid under more general demand specifications? Third, an important extension of the present model would be to allow for general equilibrium effects, both with respect to factor prices and to managers' outside options.

Notes

I am grateful for helpful comments from Avinash Dixit, Bengt Holmström, Henrik Horn, Yeongjae Kang, Johan Lagerlöf, Assar Lindbeck, Jim Markusen, Massimo Motta, Roy Radner, Jean Tirole, and Tony Venables.

1. I neglect the realistic possibility that managers associate positive utility with some positive levels of effort. However, a generalization seems straightforward.
2. It is here assumed that all managers are equally able.
3. Cost-reducing managerial effort can also be concerned with adaptation to fluctuating input prices, market conditions, or tax rules. Alternatively, managers may exhibit bounded rationality when they seek a cost-minimizing input vector in a known input requirement set. More effort may then lead to less excessive cost (over the minimum cost).
4. $r_\phi(\pi) = -\pi\phi''(\pi)/\phi'(\pi)$ for all $\pi > 0$.
5. The function c is a bijection from the interval $[0, 1]$ to itself. Differentiation of the identity $b(c(e)) \equiv e$ gives $b'(c(e))c'(e) \equiv 1$. Thus $b' < 0$. Differentiation of the latter identity gives $b''(c(e))[c'(e)]^2 + b'(c(e))c''(e) \equiv 0$. Thus $b'' \geq 0$.
6. It is presumed here that consumers' welfare is decreasing in the market price p .
7. Note that the domain of the subutility function ϕ is here restricted to $\mathbb{R}_{++}$. The domain may be extended to $\mathbb{R}$ by setting $\phi(\pi) = -\infty$ for all $\pi \leq 0$, *mutatis mutandis* (see assumption B1).

References

- Corden, W. M. (1970). The efficiency effects of trade and protection. In I. A. McDougall and R. H. Snape, eds., *Studies in International Economics*. Amsterdam: North-Holland.
- European Commission (1988). The economics of 1992. *European Economy* 35.
- Hart, O. (1983). The market mechanism as an incentive scheme. *Bell Journal of Economics* 14: 366–382.
- Hayek, F. A. (1945). The use of knowledge in society. *American Economic Review* 35: 519–530.
- Hayek, F. A. (1949). The meaning of competition. In *Individualism and Economic Order*. London: Routledge.
- Hermalin, B. E. (1992). The effects of competition on executive behavior. *Rand Journal of Economics* 23: 350–365.
- Hicks, J. (1935). Annual survey of economic theory: the theory of monopoly. *Econometrica* 3: 1–20.
- Holmström, B. (1979). Moral hazard and observability. *Bell Journal of Economics* 10: 74–91.
- Horn, H., S. Lundgren, and H. Lang (1995). Managerial effort incentives, x-inefficiency and international trade. *European Economic Review* 39: 117–138.
- Leibenstein, H. (1966). Allocative efficiency versus x-efficiency. *American Economic Review* 56: 392–415.
- Martin, J. P. (1978). X-inefficiency, managerial effort and protection. *Economica* 45: 273–286.

- Martin, S. (1993). Endogenous firm efficiency in a Cournot principal-agent model. *Journal of Economic Theory* 59: 445–450.
- Nalebuff, B., and J. Stiglitz (1983). Information, competition, and markets. *American Economic Review* 73: 278–283.
- Nickell, S. (1996). Competition and corporate performance. *Journal of Political Economy* 104: 724–746.
- Scharfstein, D. (1988). Product-market competition and managerial slack. *Rand Journal of Economics* 19: 147–155.
- Schmidt, K. (1994). Managerial Incentives and Product Market Competition. Mimeo, Bonn University.
- Smith, A. (1776). *An Inquiry into the Nature and Causes of the Wealth of Nations*. Citation from edition by R. H. Campbell, A. S. Skinner and W. B. Todd, 1976. Oxford: Clarendon Press.
- Vickers, J. (1995). Concepts of competition. *Oxford Economic Papers* 47: 1–23.
- Weibull, J. (1996). Managerial Incentives and Market Integration. Research Institute for Industrial Economics WP 472, Stockholm, Sweden.

8

On Buyers' and Sellers' Markets under Capitalism and Socialism

Martin Weitzman

Low-E and High-E Economies

To tell the basic story with some color, imagine seeing the world through the eyes of a proverbial man from Mars who drops in for a first visit to the planet Earth. The man from Mars would note that Earth is divided into a number of countries, some subset of which declare themselves to be socialist and others capitalist. Our imaginary visitor would quickly discern that although there is no rigid dividing line, whole clusters of correlated characteristics differentiate the two systems. A critical distinction concerns the allocation of resources. In capitalist economies or capitalist sectors of an economy, individuals are basically free to make resource-allocation decisions on their own through the market. In socialist economies or socialist sectors of an economy, the state, through its agents, makes many more resource allocation decisions by central planning. Perhaps the Martian would draw some crude analogy with two different kinds of parking lots he has observed on Earth. In both kinds of lots the total number of automobiles that can be parked is constrained by the total number of available spaces. In the "capitalist" parking lots, drivers can park in any available space. In the "socialist" parking lots, each space has been painted with a different number, and each driver is also assigned a number and can park only in the space marked by the matching number.

The Martian could not help but notice tremendous diversity among capitalist and socialist economies. Capitalist countries include the United States, Japan, West Germany, Sweden, and Brazil. Socialist countries include the Soviet Union, China, Hungary, East Germany, and Yugoslavia. Although there are enormous national differences among the various capitalist economies and among the various socialist economies, certain broad generalizations would strike the visitor. One of them is that the socialist economies all exhibit, to varying-degrees, a kind of generalized seller's market for goods and services. By comparison, the capitalist economies display, also to varying degrees, a sort of pervasive buyer's market for goods and services. What this observation means is perhaps best illustrated by examples taken from consumer behavior, although the point is more general.

Under capitalism, our Martian observer would readily note a kind of perpetual excess supply of goods.¹

If the Martian chose at random some good or service with its corresponding price and asked potential customers whether they wanted to buy more at that price, they would answer no. If the question were posed to the sellers whether they wanted to

sell more at that price, the answer would be almost always be yes. This excess of supply over demand in a capitalist economy puts constant pressure on output markets because at any time firms are actively seeking to expand production and sales. All aspects of economic life under capitalism seem to be shaped by this asymmetric bias. Copious inventories are held by the sellers, who readily make available to potential customers abundant information about their products. The inhabitants of a market economy appear to live in a kind of buyer's paradise. As consumers they are actively courted by the hidden persuaders of advertising; their opinions about what they like to consume are eagerly sought out and listened to; entirely new products and new variations of old products are continually created to gratify their existing wants and to anticipate their future desires; sellers vie with each other in an effort to improve in their buyers' eyes the image of their product and the quality of the services accompanying its sale and maintenance.

Almost everywhere under capitalism, it is the seller who expends effort to find and capture the buyer. On a personal level, the relationship of seller to buyer is generally one of deference. The folklore and literature of capitalism reflect this basic fact. From the informal traveling-salesman stories and the *New Yorker* cartoons about life on Madison Avenue to serious literature like Arthur Miller's *Death of a Salesman*, capitalism is a culture of persistent sellers trying to overcome the sales resistance of coy buyers. Rebates, gifts, easily available credit financing, and ingratiating behavior, not to mention bribes, favors, and kickbacks, are invariably given by sellers to buyers, not the other way around. Colleges and business schools offer innumerable courses on marketing techniques but practically none on buying techniques. A good salesperson is far more highly valued and rewarded in the business world than a good commercial buyer. On a national level, the total money spent on selling, including advertising, far exceeds the resources spent on buying.

The Martian visitor might thus come to the following general conclusion. Although there are many differences between various capitalist economies or sectors, and there may be some exceptions to the rule, capitalism as a whole can legitimately be labeled a "high-E" system, where E stands for generalized selling effort. By contrast, the Martian observer would be forced to conclude that socialist economies or sectors, although showing just as much variation, and also subject to some minor exceptions, are basically "low-E" systems.

The man from Mars would be struck by how the relationship between buyer and seller under socialism represents a kind of anti-world to capitalism.² Under socialism, although to varying degrees in different sectors or countries, it is generally a subservient buyer who must play the role of an aggressive suitor trying to overcome the sales resistance of a coy, indifferent seller. State-enterprise goods are unattrac-

tively packaged and indifferently presented in drab stores. Indeed, the most desirable wares are not even on display, but hidden somewhere in a back storage room where "special favors" rendered by the buyer to the seller may perhaps release them. Sellers have little information about the products they are selling, or if they do have such information they seem to keep it to themselves. (Frequently, it seems, buyers appear to be better informed about the product than are sellers.) Not only can they not count on ready access during all reasonable hours to well-stocked inventories in the sellers' showrooms or stores, but buyers are also forced to stock up on inventories themselves if they do not wish to be caught short. Language itself adapts to socialist economic reality. In Russian, for example, there is a multitude of common expressions used to articulate the degree of difficulty required for obtaining a commodity.

Effort spent on locating and buying goods far outweighs effort spent on selling. It sometimes seems that any fool can sell under socialism. The socialist analogue of the capitalist sales folk-hero is the crafty, highly valued, and well-paid "pusher" of the state enterprise who has the contacts, payoffs, nerve, and clout required to round up the needed input materials when they are in short supply.³ Throughout the economy there is a ubiquitous atmosphere of deficit commodities all over the landscape coexisting with a somewhat paradoxical sense that even a price rise would not much choke off demand. The quality of consumer goods is generally low, service is poor, delivery times are long, spare parts are nonexistent or difficult to obtain, there is not much product innovation, and overall very little effort is expended to please the customer. Whatever its other possible comparative virtues, such as, perhaps, relative employment security or relative income equality, our Martian observer would have to conclude that concerning generalized selling effort *E*, socialism is essentially a low-*E* system.

Having become convinced that the impressionistic evidence of a systemic difference in *E*-ratings is almost overwhelming, our Martian visitor, if he were at all curious, might begin to wonder *why* capitalism tends to be a high-*E* system while socialism is generally a low-*E* system. Standard Walrasian general equilibrium theory, based on the model of perfect competition, would not offer much help. In principle, there should be a basic duality or symmetry between price- and quantity-based general equilibrium systems. This observation suggests that something fundamental is being missed by such an approach. But what, exactly, is left out?

The question itself suggests where to look for an answer. Following Evsey Domar and others, I will argue that under capitalism the widespread existence of imperfect competition—the fact that the overwhelming preponderance of firms face a downward-sloping demand curve—is the driving force behind creating a high-*E*

system. If the logic of this position is accepted, then half of the puzzle has already been solved. The other half reduces to determining how this mechanism is impaired, or even reversed, under socialism. Here I will argue that the widespread existence of taut, firm-specific, output-limiting resource constraints, which is the essence of a centrally planned economy, by itself, even under otherwise ideal circumstances, is sufficient to create a low-E system, without appealing to explanations like bad planning, wrong prices, or insufficient concern about profitability, including budget indiscipline.⁴ In saying this I do not mean to overstate the case by implying that these alternative explanations of socialism's low-E status do not have credence. They clearly do in some cases. The point is not that bad planning, non-profit-maximizing behavior, wrong prices, or other explanations are irrelevant, but rather that even a "reformed" socialist economy whose enterprises maximize profits at appropriate output prices will yield a low-E system so long as each enterprise's inputs are rigidly constrained by taut central plans. The argument will be made largely by performing the thought experiment of contrasting "identical twin" capitalist and socialist systems built on a common foundation of monopolistic competition theory. The primary purpose of the exercise is frankly analytical—to explore what minimal concepts of institutional structure are involved in creating buyers' or sellers' markets in alternative economic systems.

Capitalism as a High-E Economy

I begin by trying to explain the high-E status of capitalism because I think there already exists a relatively powerful, convincing, general theory based on imperfect competition. After formally developing the argument in this section, I attempt to adapt it to explain the low-E status of a socialist economy in the next section, an undertaking which is inherently more controversial.

Consider the following variant of standard monopoly theory. The firm cannot determine output sales Y directly, but rather indirectly by controlling price P and "effort" E . Output Y that can be sold is a function of P and E described by the demand function

$$Y = D(P, E), \quad (8.1)$$

where

$$D_1 = \frac{\partial D}{\partial P} < 0, \quad D_2 = \frac{\partial D}{\partial E} > 0. \quad (8.2)$$

In this formulation, E is treated as a single scalar variable symbolizing "generalized sales effort," although extension to a vector formulation is routine. E could stand for many things, including service, quality, speed of delivery, variety, showroom environment, courtesy, cooperation, information, advertising, packaging, favors, friendliness, and so forth. Of its very nature, E is rather awkward to measure. In a central planning context, for example, it would be difficult for the planners meaningfully to order various levels of E because it is intrinsically hard to specify.

Exerting higher values of E presumably imposes a greater cost on the seller, some of it perhaps psychic rather than monetary. The cost of producing Y units of the single good at a level of generalized sales effort E is denoted

$$C(Y, E), \quad (8.3)$$

where

$$C_1 = \frac{\partial C}{\partial Y} > 0, \quad C_2 = \frac{\partial C}{\partial E} > 0. \quad (8.4)$$

Let $E = 0$ denote the absolute minimal possible sales effort, that is,

$$C(Y, 0) = \min_E C(Y, E). \quad (8.5)$$

Although the model itself cannot speak directly on this interpretation, presumably $E = 0$ represents a bleak world of minimal sales effort resembling something like an austere, unfriendly, warehouse-like environment where customers find and unpack their own purchases. In such a world there is probably widespread mismatching of buyers with their less-than-most-desired product-types and sellers take very little interest in satisfying their customers.

It is difficult to make welfare statements about optimal levels of E without further elaborating the model. On one hand E might represent a genuine increase in welfare from, say, the firm providing an environment leading to better matching and coordination of consumers with the types of products they most desire. More routinely, E could be socially desirable because of improved quality, better service, and so forth. On the other hand, E could stand for things like socially nonproductive, uninformative advertising aimed at manipulating consumers' tastes and largely offset by the equally non-productive advertising of a firm's rivals. Whether there ends up being too much or too little E (relative to a social optimum) in any given situation depends, among other things, on the typical consumer's utility function $U(Y, E)$. In principle the verdict could go either way. In practice, many experienced, fair-minded observers would be prepared to conclude that one of capitalism's genuine

comparative strengths is the attention it pays to its consumers. Indeed, I think it is not unfair to characterize one of the major aims of most socialist reforms as increasing the level of E in the economy.

The problem of the profit-maximizing monopolist firm⁵ can be stated as follows:

$$\min_{P,E} \Pi(P, E) = P D(P, E) - C(D(P, E), E) \quad (8.6)$$

The corresponding first order conditions are

$$\frac{\partial \Pi}{\partial P} = 0 \quad (8.7)$$

$$\frac{\partial \Pi}{\partial E} = 0. \quad (8.8)$$

Equation (8.7) yields the well-known equality of marginal revenue (MR) with marginal production cost ($MC \equiv C_1$):

$$MR = MC. \quad (8.9)$$

An alternative way of writing (8.9) is

$$P = \mu MC \quad (8.10)$$

where the markup coefficient $\mu > 1$ is related to the price elasticity of demand $\varepsilon > 1$ by the formula

$$\mu = \frac{\varepsilon}{\varepsilon - 1}. \quad (8.11)$$

Equation (8.8) yields the condition

$$P - MC = ME, \quad (8.12)$$

where "marginal effort" ME stands for the pure marginal effort-cost (in numeraire equivalent) required to sell one extra unit of the commodity at its posted price:

$$ME = C_2 / D_2 \quad (8.13)$$

The logic of condition (8.12) is straightforward. If an extra sale can be made by the firm, it is worth the difference between price and marginal production cost, $P - MC$. Profit-maximizing behavior dictates that this be equated to the marginal-effort cost of making the extra sale, ME .

The marginal condition (8.12), although not nearly so much emphasized in economic theory as the more standard condition (8.9), is in many respects the more important of the two. Company management typically spends far more time and energy worrying about (8.12) than about (8.9). The wedge between price and marginal cost continually drives a capitalist firm to seek out any and all variables $\{E^j\}$ that might be even remotely correlated with sales, and to push each such variable to the limit where

$$P - MC = ME^j \quad (8.14)$$

for all possible j . To ignore (8.12), or, more generally, (8.14), is in many ways to disregard to essence of what capitalism is all about. Capitalist firms are very busy indeed perpetually casting around for imaginative ways of finding new variables $\{E^j\}$ that might improve the image (real or artificial) of their product in the eyes of their all-important customers. It is not an exaggeration to call the kind of dynamism associated with perpetually high selling effort the driving force behind capitalism.

Note, as an aside, that welfare economics can become quite tricky in the present context. As was previously mentioned, whether so much emphasis on high values of $\{E^j\}$ is a good thing or a bad thing depends very much on particular details of the specification at hand. It may well be that high values of E^j are a good thing, whereas high values of E^i are a bad thing for some particular i and j , depending on the relevant cost and utility functions. Now the hallowed prescription $P = MC$ might be a quite bad rule because it gives the firm no incentive to increase E above zero. (The socially optimal administered price in this second-best problem where the firm chooses E might actually be higher than the monopolist's chosen price, at least in theory, depending, among other things, on the particular form of the prototypical consumer's underlying utility function $U(Y, E)$.)

There is a well-defined sense in which the behavior of the high- E firm described by conditions (8.9), (8.12) will create the atmosphere of a buyer's market. The seller is eager to capture any marginal buyer who randomly comes within buying range, with all the asymmetry of human relations that implies, because each unanticipated extra purchase is worth $P - MC$. Thus, supply exceeds demand even at the profit maximizing price which the seller has freely chosen. A perpetual buyer's market therefore exists in the sense that the seller continually wants to sell more than the buyer wants to buy. And this is the normal state of a capitalist market, not the artificial result of any externally imposed price floor.

The next step in the chain of logic I am trying to develop is integrating the above model into some more comprehensive story of monopolistic competition.⁶ Consider,

then, a Chamberlinian “group” of n symmetric firms. Each seller $i = 1, 2, \dots, n$ faces the demand function

$$Y_i = D(P_i, E_i; \{P_j\}, \{E_j\}) \quad (8.15)$$

where j represent all integers between 1 and n except for i .

The solution concept to be used is a symmetric Nash equilibrium in prices and effort. In words, P^*, E^* is an equilibrium if, whenever all other firms $j \neq i$ have set

$$P_j = P^*, E_j = E^*, \quad (8.16)$$

then it is profit maximizing for firm i also to set

$$P_i = P^*, E_i = E^*. \quad (8.17)$$

The solution (8.17) will satisfy conditions (8.9), (8.12), with all the implications that have previously been discussed for the pure monopoly case.

Throughout this chapter I take the total number of firms in the Chamberlinian group as fixed at n because that is the easiest way of telling the story I want to tell. Although it would not make any essential difference for the purposes of this paper, a more complete analysis might determine n endogenously, say by the zero-profit condition that typically follows from assuming free exit and entry. Although such an extension could be useful for its own reasons, such as giving a more complete picture of equilibrium or highlighting the role of increasing returns, it is essentially irrelevant in the present context.

To focus the analysis very sharply on some basic points, it will be useful, at least initially, to further restrict the demand functions (8.15) in such a way that for all configurations of the n possible firms' prices and the n possible firms' effort levels

$$\sum Y_i = nY^* \quad (8.18)$$

where Y^* is a constant standing for the average demand per firm.⁷

Condition (8.18) represents the idea that the demand curve faced by the “group” is infinitely more inelastic than the demand curve faced by an individual member-firm of the “group.” Demand for Ford Escorts with respect to price or sales effort is much more elastic than demand for automobiles as a whole. If firm i increases its price or decreases its sales effort, it loses customers, but primarily to other sellers in the group. Although the very sharp results to be obtained presently will depend, among other things, on assumption (8.18), less sharp results conveying the same basic message could be derived so long as group demand (Chamberlin's DD curve) is significantly less elastic than individual demand (Chamberlin's dd curve), an

assumption that is typically made. One immediate consequence of (8.18) is that any symmetric Nash equilibrium implies

$$Y_i = Y^* \quad (8.19)$$

for all $i = 1, 2, \dots, n$.

Socialism as a Low-E Economy

I want now to perform a highly artificial thought experiment. The "capitalist twin" group equilibrium has yielded a solution Y^*, P^*, E^* , whose properties were just explained. What would a "socialist twin" group equilibrium look like under otherwise identical conditions?

To answer this question, one first has to answer the even more basic question of what constitutes the analytical essence of a socialist or centrally planned state enterprise. What is the socialist analogue of the capitalist firm whose behavior was described in the previous section? This is not, it must be admitted, an easy question to answer. I would say that the essence of central planning means that each state enterprise is resource constrained because of a state imposed enterprise-specific limitation on inputs. Socialist enterprises almost always have explicit or implicit capacity limitations because, even under the most "reformed" regimen, they cannot truly freely obtain capital, raw materials, or labor on an open market without bureaucratic interference.⁸ The essential difference for the purposes of this paper is that the individual capitalist firm can typically expand output if it wants to by bidding away input resources from alternative uses, whereas the typical socialist firm is effectively limited from expanding on its own without going through bureaucratic channels to obtain the required additional inputs.

I would say, then, that the analytical essence of central planning involves the assignment of taut firm-specific inputs. Obviously it is difficult to draw hard and fast demarcation lines, but if firms can readily acquire inputs on their own, then I would hesitate to call such a system socialist or centrally planned.

Without much further ado, the "socialist twin" firm i is subjected to the additional constraint:

$$Y_i \leq Y^*. \quad (8.20)$$

It should be understood that condition (8.20) comes about not because the central planners do not want the firm to produce any more output, but because by limiting the inputs they have allocated to the firm they are de facto constraining its output.

In order to complete the arbitrarily imposed symmetry between hypothetical capitalist and socialist twins, each firm i will be subjected to the price ceiling

$$P_i \leq P^* \quad (8.21)$$

Although it turns out that (8.21) is a less vital constraint than (8.20) in influencing the basic characteristics of the hypothetical socialist twin system, in a sense to be explained later, at this stage both constraints might be taken merely as symmetric counterparts in the thought experiment to be performed.

In what follows I am modeling some imaginary "idealized form" of state enterprise as representative of a kind of generic socialism. This idealized type of socialist firm is neither so unreformed that profits are not an important success indicator nor so reformed that state-imposed input restrictions do not matter.

Consider, then, the symmetric Nash equilibrium of a hypothetical group of profit maximizing socialist firms facing exactly the same demand and cost functions as their capitalist counterparts, but which differ only in facing the additional constraints (8.20), (8.21). Because the price and quantity ceilings imposed on the socialist twin firms by constraints (8.20) and (8.21) are the very same price and quantity solutions that emerge from the capitalist-twin Chamberlinian equilibrium, it might be conjectured that both groups should have identical equilibria. This turns out to be true for price and quantity, but not, alas, for effort.

The basic result emerging from this hypothetical thought experiment is that in the socialist-twin version of a profit-maximizing symmetric Nash equilibrium, the solution is:

$$Y_i = Y^*, \quad P_i = P^*, \quad E_i = 0, \quad (8.22)$$

for all $i = 1, 2, \dots, n$.

The reason (8.22) constitutes an equilibrium for the socialist-twin group should be readily understandable. With respect to generalized sales effort, why should the firm expend any? Its potential customers cannot obtain merchandise anywhere else, because all rival firms are effectively barred from selling more output, as they cannot produce it. Assured of a captive group of buyers, in that there is nowhere else for disgruntled customers to shop, it is in the firm's interest to offer minimal selling effort. When every firm is so acting, it results in a low-E (in this particular case actually a zero-E) system. The low-E status is an unintended consequence of limiting each firm's ability to take on a rival firm's mistreated customers, even though the output constraint is exactly what the firms themselves would choose to produce in a market economy without any trace of central planning. The socialist firm is indif-

ferent to a marginal customer because what it gains from selling to some new buyer it will lose by not having enough to sell to an old buyer.

In this scenario, the profit-maximizing socialist enterprise would like to produce more output, if only it could get hold of additional inputs. Any such firm is therefore likely to expend effort on treating its suppliers very nicely indeed, because it is always eager to obtain more input. Hence there is a *prima facie* case for expecting the socialist firm to treat its customers and suppliers quite asymmetrically.

The socialist enterprise in this model is chafing under the imposed price ceiling, so it is legitimate for the central planners to be concerned about repressed inflation. With its potential rivals unable to take business away because they cannot respond by expanding output, the profit-maximizing socialist firm would love to be able to raise its price. Actually, in this particular model, the socialist firm would wish to raise its price no matter what ceiling was imposed. A "wrong" imposed price is not the essential cause of repressed inflation or of a low-E system in this model, in as much as they would be present with any imposed price above marginal cost. Rather, it is the tight effective output constraint that does the real damage by limiting the ability of opponent firms to react by taking on dissatisfied customers.

Note that in the socialist-twin economy there does not exist any measurably shortage of goods, disequilibrium, or repressed inflation in the usual macroeconomic sense, even though such an economy can at the same time be fairly characterized as being in the state of a perpetual seller's market. An econometrician could not find in the usual aggregate data any signs of sustained disequilibrium, persistent quantity rationing, or perpetual repressed inflation, for the goods market is always clearing. (Indeed, the prices and quantities are exactly the same as would prevail in a corresponding market economy.) Yet the inability to find statistical evidence of persistent shortages is not inconsistent *per se* with the common, everyday observation that consumers in socialist countries are living in a low-E economy. Thus, there is a sense in which the present model may be interpreted as reconciling the widespread impressionistic evidence that socialism is a "shortage system" with the econometric evidence, which seemingly rejects the hypothesis of sustained widespread shortages.⁹

Table 8.1 contrasts essential properties of the hypothetical socialist and capitalist twin economies.

Extensions and Implications

The model of this chapter was constructed quite specifically to bring sharply into focus the basic message that taut, firm-specific resource constraints can

Table 8.1
Comparison of socialist and capitalist "twins"

Characteristic	Capitalist twin	Socialist twin
selling effort	$ME = P - MC$	$E = 0$
attitude to marginal buyer	eager to capture	indifference
attitude to price	content with price	wishes to raise
attitude to inputs	indifference	wants more

inadvertently cause a low-E system. The essence of that message would remain, with appropriate modification, even under a less crisp formulation. In the general case, central planning will not cause a zero-E solution or other extreme features, but the broad tendencies of the simple model would remain. As a group of related firms changes from market to plan, the *ceteris paribus* result of limiting each firm's output by restricting or hampering its access to above-plan inputs is to diminish any one firm's elasticity of demand with respect to price and effort because its potential opponents' response has been limited. Other things being equal, then, each planned firm will wish to raise price and lower effort because, in effect, it now has a more sheltered market since its rivals are less easily able to take up the slack by expanding their output. When, additionally, firms are prevented from raising prices above some ceiling, that leaves diminished effort as the primary outlet, resulting in a low-E economy.¹⁰

It is not critical to the basic argument that each socialist twin firm of the group be assigned literally the same quota and price as its capitalist twin counterpart. That assumption was made only to illustrate dramatically that even a situation where a socialist system is artificially made to look as similar as possible to a capitalist system still yields low-E behavior. The really crucial assumption is the assignment under central planning of taut, firm-specific, output-limiting resource constraints.

If the theme of this paper rings true, there is a basic incompatibility between central planning—meaning here taut firm-specific physical allocation of inputs—and a high-E economy. Perhaps the most interesting implication concerns the scope of possible limitations on economic reforms under socialism. It is probably difficult, or maybe even impossible, to reform a low-E economy into a high-E economy by instituting piecemeal reforms. The whole economic system is interrelated in such a way that freeing up the output side alone is not sufficient to make firms care about generalized selling effort so long as the planning bureaucracy still maintains control of raw materials and primary factors. To obtain the benefits of economic reform, the input side of labor, capital, and raw materials must also be freed of central control

so that a sufficiently large number of firms can readily expand their output if they so choose without having to go through a time-consuming bureaucratic procedure. The benefits of a high-E economy come ultimately from the threat of competition; and that threat is an empty bluff so long as a firm can feel free to mistreat its customers because potential rivals cannot respond by taking away its business.

Notes

I would like to acknowledge very useful discussions on the subject of this paper with Abram Bergson, Evsey D. Domar, and János Kornai, without attributing to them errors and opinions.

1. This point, along with its ramifications, has been made by Domar, Bishop, and others. See Bishop (n.d.), Domar (1988), and Scitovsky (1986) and the references cited therein.
2. This point has been explicated in some detail in the various works of János Kornai and has been made by Evsey D. Domar.
3. These aspects of the Soviet economy are well described in the works of Joseph Berliner.
4. Kornai stresses heavily the role of what he calls the "soft budget constraint." See Kornai (1986).
5. The problem and its solution have been analyzed in the literature. See Schmalensee (1972) and the references contained therein.
6. The classical reference is Chamberlin (1962). See also Spence (1976).
7. This assumption is frequently made in the context of a situation with one-dimensional attribute-characteristics, as in the standard model of spatial monopolistic competition along a line. See Hotelling for the seminal article and Lancaster for further developments.
8. On this point, see, e.g., Kornai (1986).
9. The careful econometric studies of Richard Portes and his associates seem to show that centrally planned economies are not generally characterized by sustained shortages, persistent disequilibrium quantity rationing, or continual repressed inflation in the standard macroeconomic sense that these terms are usually understood. See especially Portes (1986) for a summary of this line of research, its conclusions, and a bibliography. The conclusions of Portes and his associates have not gone unchallenged, in part because they seem so much at variance with impressionistic evidence. This has given rise to a heated debate; see Kornai (1980, 1982, 1986) for the most prominent example of the other view.
10. This argument—the capitalist status quo is unsustainable under central planning because individualistic firms will wish to raise price and lower effort—remains valid even if a central planning equilibrium fails to exist. In that case, firms presumably oscillate in their behavior, exhibiting lower effort levels on average than their capitalist twin counterparts.

References

- Berliner, Joseph (1976). *The Innovation Decision in Soviet Industry*. Cambridge: MIT Press.
- Bishop, Robert L. (Unpublished). Manuscript on price theory. Book II, Chapter III, 32–42.
- Chamberlin, Edward. (1962). *The Theory of Monopolistic Competition*. 8th edition. Cambridge: Harvard University Press.
- Domar, Evsey D. (1988). The Blind Men and the Elephant: An Essay on Isms. MIT. Department of Economics Working Paper #473.
- Hotelling, Harold (1929). Stability in competition. *Economic Journal* 39: 41–57.

- Kornai, János (1980). *Economics of Shortage*. Amsterdam: North-Holland.
- Kornai, János (1982). *Growth, Shortage, and Efficiency*. Oxford: Blackwell.
- Kornai, János (1986a). The soft budget constraint. *Kyklos* 39(1): 3–30.
- Kornai, János (1986b). The Hungarian reform process: visions, hopes, and reality. *Journal of Economic Literature* XXIV: 1687–1737.
- Lancaster, Kelvin (1979) *Variety, Equity and Efficiency*. New York: Columbia University Press.
- Portes, Richard (1986). The Theory and Measurement of Macroeconomic Disequilibrium in Centrally Planned Economies. CEPR Discussion Paper No. 91.
- Portes, Richard, and David Winter (1980). Disequilibrium estimates for consumption goods markets in centrally planned economies. *Review of Economic Studies* 47(1): 137–159.
- Schmalensee, Richard (1972). *The Economics of Advertising*. Amsterdam: North-Holland.
- Schmalensee, Richard (1979). Market-structure, durability, and quality: a selective survey. *Economic Inquiry* 17: 177–196.
- Scitovsky, Tibor (1985). Pricetakers' plenty: a neglected benefit of capitalism. *Kyklos* 38: 517–536.
- Skurski, Roger. The buyers' market and Soviet consumer goods distribution. *Slavic Review* 31: 817–830.
- Spence, A. M. (1976). Product differentiation and welfare. *American Economic Review* 66: 407–414.

II **THEORY OF TRANSITION**

9

Soft Budget Constraints and Transition

Mathias Dewatripont, Eric Maskin, and Gérard Roland

1 Introduction

“Soft budget constraints”—the refinancing of loss-making enterprises—were a key feature of socialist economies. János Kornai (1979, 1980, 1992), who coined this term, has shown the role of soft budget constraints in explaining the emergence and reproduction of shortages in socialist economies. The concept of soft budget constraints now belongs to the vocabulary of economics. Its importance and relevance is acknowledged beyond socialist economies. There are well-known examples of soft budget constraints in big corporations (for example, the famous Chrysler case) or in the banking sector (for example, the bailout of U.S. savings and loans).

Soft budget constraints have naturally been an important concern in transition economies, in Eastern Europe as well as in Asia. For example, the following ideas have been stressed in the context of the transition process:

- continued soft budget constraints may prevent loss-making enterprises from restructuring because they lack the negative incentives related to the threat of bankruptcy;
- soft budget constraints may be an obstacle to the process of sectoral reallocation as continued subsidies to loss-making state-owned enterprises (SOEs) may prevent efficient private firms from outbidding them for workers (see Sachs and Woo 1993; Castanheira and Roland 1996);
- macroeconomic stability may be jeopardized because continued soft budget constraints in SOEs make it difficult to keep government expenditures under control, as illustrated by the case of Russia in its first years of transition (Litwack 1993).

The necessity of hardening budget constraints of enterprises in economies in transition has clearly been recognized in the by now large literature on transition. How effectively to harden budget constraints has, however, not been the subject of much detailed analysis. Hardening budget constraints is often presented in “reduced form” as a direct choice of action on an exogenous policy variable. In other words, it is as if, when policy-makers decide to harden budget constraints and make such announcements, hard budget constraints automatically follow. The question of the credibility of such announcements and decisions has been given little attention. This is partly because, in an important part of the literature on transition, soft budget constraints are identified with subsidies.¹ Seen in that perspective, hardening budget constraints is, then, nothing other than a decision to cut subsidies. It is then not

astonishing to see the hardening of budget constraints as an exogenous policy variable.

Another line of research, initiated by Dewatripont and Maskin (1995), views soft budget constraints as endogenous to specific institutions. Soft budget constraints are seen as a dynamic incentive problem where a funding source, be it a government or a bank, cannot commit to keep an enterprise to a fixed initial budget. Soft budget constraints represent an inefficiency in that the funding source would like to commit *ex ante* not to bail out firms, but they know they will be tempted to refinance the firm *ex post* because the initial injection of funds is sunk. Here, the interesting question becomes that of the institutional conditions under which one has hard or soft budget constraints. Hardening budget constraints is thus not simply a direct policy variable, but rather the result of institutional design.

This distinction is important for the transition process. Indeed, if soft budget constraints are the result of a dynamic commitment problem, decisions to harden budget constraints cannot be credible unless institutions are in place to this effect. Kornai (1995) himself insists very much on this point in the context of transition: "There are a great many influences on the government tempting it to loosen the financial discipline and soften the budget constraint. . . . Since the period of tougher financial discipline began, exceptional procedures have been followed in several cases, in many of which the bargaining led to agreement. The remnants of the soft budget constraint are clearly visible. There is a danger that the frequency of the exceptions will undermine the credibility of the government's pledges concerning the tough financial discipline" (p. 150).

An area where this question is particularly sensitive is the bailout of banks. Indeed, if macrostabilization programs have drastically cut subsidies, loss-making SOEs have continued to be bailed out via different channels such as interenterprise credits, and most importantly via bank credit. As a consequence, the quality of bank portfolios has sharply deteriorated, and created a "bad loan problem." This problem illustrates very well the soft budget constraint phenomenon. Early analysts of the bad-loan problem have emphasized the need for bank recapitalization as the appropriate solution (Begg and Portes 1992; Mitchell 1994).² At the same time, analysts acknowledged that such recapitalization could only occur once since otherwise expectations of future bailouts would seriously dampen banks' incentives. Despite these clear warnings, there have been repeated bank bailouts, in Hungary for example. Accumulation of bad loans indeed strengthens pressures to bail out banks and expectations of bailouts give fewer incentives to banks to improve their loan portfolio. This is a clear example of the soft budget syndrome.

In this paper, we survey the incentive literature on soft budget constraints and we analyze its relevance for transition. In the next section we survey the Dewatripont-Maskin model and discuss some of its applications to the socialist economy. In the next sections, we look at the effects of various reforms on the hardening of budget constraints: privatization, product market competition, government reform, and banking reform.

2 Soft Budget Constraints as a Time-Consistency Problem

Kornai's work has focused mainly on the consequences of the soft budget constraint, namely the emergence of pervasive shortages under socialism. He primarily attributes the *causes* of the soft budget constraint to political constraints, that is, to the desire of "paternalistic" governments to avoid socially and politically costly layoffs. Dewatripont and Maskin's (1995) analysis of the soft budget constraint stresses instead dynamic commitment problems in the presence of irreversible investment (see also Schaffer 1981 and Qian and Xu 1991).³

They consider the following adverse selection problem. The government faces a population of firms, each needing one unit of funds in initial period 1 in order to start their project. A proportion α of these projects are of the "good, quick" type: after one period, the project is successfully completed, and generates a gross (discounted) financial return $R_g > 1$. Moreover, the manager of the firm (possibly also workers) obtains a positive net (discounted) private benefit E_g . In contrast, there is a proportion $(1 - \alpha)$ of bad and slow projects which generate no financial return after one period. If terminated at that stage, managers obtain a private benefit E_t . Instead, if refinanced, each project generates after two periods a gross (discounted) financial return π_b^* and a net (discounted) private benefit E_b . Initially, α is common knowledge but individual types are private information. A simple result easily follows: if $1 < \pi_b^* < 2$ and $E_b > 0$, refinancing bad projects is sequentially optimal for the government, and bad entrepreneurs who expect to be refinanced apply for initial financing. The government would, however, be better off if it were able to commit not to refinance bad projects, because it would thereby deter managers with bad projects from applying for initial financing, provided $E_t < 0$.

Termination is here, by assumption, a discipline device that allows the uninformed investor (creditor) to turn away bad types and only finance good ones.⁴ The problem is that termination is not sequentially rational if π_b^* is bigger than one: once the first unit has been sunk into a bad project, its net continuation value is positive so that, in the absence of commitment, the soft budget constraint syndrome arises. In this

setup, because irreversibility of investment is such a general economic feature, the challenge for theory is more to explain why hard budget constraints prevail rather than why budget constraints are soft in the first place.

One can use this analysis as a starting point to understand how transition strategies can alleviate the soft budget constraint problem. We now turn to several such strategies.

3 Privatization

One interpretation of the above problem is to consider that π_b^* internalizes a political bias of socialist governments, for example towards excessive employment. Under this interpretation, a profit-maximizing creditor would *not* refinance bad firms. Moreover, if there were any refinancing, it would be accompanied by different restructuring measures from those that would be chosen if the government were directly in control of the firm.

How can a reformist government constrain the ability of future governments to exercise a political bias in economic activity? Boycko et al. (1992) and Shleifer and Vishny (1994) argue that privatization, namely the allocation of control rights and of profit streams to managers, provides the solution. In this case, governments will find it harder and more costly to move firms away from profit-maximizing behavior, because they will have to “bribe” firms instead of simply dictating their behavior. If such bribing is costly for governments, for example because it must be financed through distortionary taxation, one should expect less refinancing of bad firms and fewer deviations from profit-maximizing outcomes. In this story, the ability to bribe managers is reduced if the government is poorer, which then argues for giveaways as opposed to sales as a method of privatization, as is advocated in Boycko et al. (1995).

Note, however, that privatization per se does not always solve soft budget constraint problems inside firms. Indeed, if these problems reflected the political bias of government towards excessive unemployment, privatized firms may exploit the government’s softness to still obtain funds. This issue is particularly relevant in transition economies where governments have gone for “insider privatization,” giving away control to existing management teams, as in Russia. In this case, Debande and Friebe (1996) stress the fact that managers may have a bias in favor of excessive size (i.e., may be “empire-builders”), which is the equivalent of the government bias in favor of employment, and that the reluctance of the government to hand out funds is, if anything, *reduced* by the fact that its financial stake in the company is diminished by privatization: because the government does not get the financial

reward of profit-maximizing strategies after privatization, why should it care at all about the profit consequences of the firm's actions? Faure-Grimaud (1996) stresses, moreover, the fact that by offering more precise signals about the firm's future profitability (through stock market valuation), privatization may enhance the soft budget constraint problem: if the manager is confident that the stock market will value future profits π_b^* appropriately, he will not be afraid to have to report zero profit in period 1, whereas otherwise he might abstain altogether!

4 Product Market Demonopolization

The opportunity cost of refinancing a bad project can also be raised by reducing the cost of terminating it. The cost of termination is in turn related to substitution possibilities across projects, as Segal (1993) stresses. Although Segal makes the point that the soft budget constraint can at times be seen as the result of underprovision of cost-reduction effort by a monopolist in order to extract subsidies from the government, the argument can also be made directly in our framework. Indeed, assume that the government can split each project it finances into two halves at some efficiency cost, for example, because of increasing returns to scale. In other words, two entrepreneurs are selected, and each receives $\frac{1}{2}$. In the case of a good project, the gross return will only be $\frac{1}{2}\theta R_g$ with $\theta < 1$. Similarly, in the case of a bad project that is refinanced by injecting $\frac{1}{2}$ in the second period, the gross return will be $\frac{1}{2}\theta\pi_b^*$. If we assume that there is demand only for the equivalent of one full project, the game stops if both entrepreneurs have good projects. But what if one project is bad? The soft budget constraint problem could remain if $\frac{1}{2}\theta\pi_b^* > \frac{1}{2}$. Assume, however, that if only one project is good, it is optimal for the government to expand its activities instead of refinancing the bad entrepreneur. In other words, if an increase in capacity of the good project through injection of another $\frac{1}{2}$ yields $\frac{1}{2}(2 - \theta)R_g > \frac{1}{2}\theta\pi_b^*$, bad projects will not be refinanced whenever they have been financed together with a good project. Bad projects thus get refinanced only with probability $(1 - \alpha)$. If $\alpha E_t + (1 - \alpha)E_b < 0$, then there is always a unique equilibrium where only good entrepreneurs apply for funds, because then a bad entrepreneur will not find it profitable to submit a project in the first place.

5 Decentralization of Government Decisions

Qian and Roland (1994) take a similar perspective to that in our earlier section on privatization, assuming π_b^* to be too high because of political bias. However, instead of privatization, they investigate *decentralization* as a method for reducing π_b^* . Gov-

ernment remains in control of the financing decisions, and the focus is on altering its incentives by creating competition between local governments through decentralization. Qian and Roland argue that this is one of the main specificities of Chinese reforms so far. Indeed, important improvements in enterprise incentives have taken place in China despite the absence of privatization programs. Most of these improvements have taken place in the township and village enterprises that are not privately owned but started booming after the beginning of the reform process (see, e.g., Weitzman and Xu 1993; Che and Qian 1994; Bolton 1995; and Li 1995).

Qian and Roland insert the setup we saw earlier in a general equilibrium framework with the following objective function W for the government:

$$W = x(K, I) + y + u(z)$$

where K is the level of foreign capital investment into the area and I and z are, respectively, the level of public infrastructure investment and public consumption. Moreover, y is the net return of financing and refinancing firms (including managerial private benefits), while $x(\cdot)$ and $u(\cdot)$ are the net returns of the two forms of investment and of public consumption. Assume positive decreasing marginal returns as well as *complementarity* between K and I .

The degree of decentralization of decisions can be seen to influence the budget constraint firms face through the intensity of capital mobility across regions. Maximizing the objective function W implies refinancing bad firms, and thus a soft budget constraint, if and only if:

$$E_b + \pi_b^* - 1 \geq \frac{\partial x(K, I)}{\partial I} + \frac{\partial x(K, I)}{\partial K} \frac{dK}{dI} = u'(\bar{z})$$

where the left-hand side ($E_b + \pi_b^* - 1$) is the net increase in y when one unit of funds is used to refinance bad firms, while the right-hand side is the net return to infrastructure investment or public consumption ($\bar{z}$ being the equilibrium level of public consumption). Decentralization can then harden the budget constraint of firms because local governments will compete with one another to attract foreign capital to their region by investing more in infrastructure. In other words, decentralization leads to an increase in dK/dI (for simplicity, Qian and Roland assume that dK/dI is zero for the country as a whole, but positive at the regional level). Regional governments will thus divert funds toward infrastructure investment and away from public consumption for the purpose of fiscal competition. Simultaneously, refinancing bad firms will have a higher opportunity cost, in that $u'(z)$ has increased. Call $\bar{z}$

the equilibrium level of public consumption with decentralized government. Provided $u'(\bar{z}) > (E_b + \pi_b^* - 1)$, only good projects are financed under decentralization, inasmuch as bad entrepreneurs expect to be terminated.

It is interesting to compare the results of Qian and Roland (1994) with those of Wang (1991), who sees decentralization as increased autonomy given to enterprises. These receive from the central planner fixed investment and circulating capital, which are combined using a Cobb-Douglas production function. Increased autonomy allows enterprises to decide how to allocate funds between fixed investment and circulating capital, thereby raising the risk that firms might strategically misallocate their funds in order to force the government to increase spending in their favor, which can lead to inflation if government spending is financed by money creation. Partial enterprise autonomy can thus lead to a softening of budget constraints because it gives more room for strategic distortions.

6 Decentralization of Credit and Banking Reform

Decentralization of Credit

The setup we saw earlier is also compatible with a π_b^* that reflects pure profit-maximization motives. Indeed, in the presence of sunk costs, sequential profit maximization can be inferior to *ex ante* profit maximization. In this case, privatization alone will not solve the soft budget constraint problem unless it is accompanied by other institutional changes. Dewatripont and Maskin (1995) show that the decentralization of credit may be a crucial element in hardening budget constraints. As in the previous sections, this is achieved through a reduction in π_b^* .

Assume that the continuation value of bad projects depends on an effort level a to be exerted by the initial creditor. Specifically, assume that the gross (discounted) financial return of a bad project that is refinanced is either 0 or $\bar{R}_b$, and that the probability of $\bar{R}_b$ is a . Finally, assume a to be private information to the initial creditor, who incurs effort costs $\Psi(a)$, assumed to be increasing and convex in a .

In this case, *centralization* of credit means that the initial creditor will also be the one refinancing a bad firm, so that the chosen effort level a^* will fully internalize the benefit of monitoring:

$$\pi_b^* = \text{Max}_a \{a\bar{R}_b - \Psi(a)\} \text{ and } \bar{R}_b = \Psi'(a^*).$$

Under *decentralization* the initial creditor is liquidity constrained, and refinancing has to be performed by a new creditor who has not observed monitoring effort. Given an expected effort $\hat{a}$ and limited resources for the firm and the initial credi-

tor under perfect competition among new creditors, the refinancing contract will grant $1/\hat{a}$ deducted from $\bar{R}_b$ whenever the bad project ends up being "successful" (because, by assumption, no resources are available if the project is unsuccessful). Given $\hat{a}$, the effort level privately chosen by the first creditor will lead to:

$$\pi_b^{**} = \text{Max}_a \left\{ a \left(\bar{R}_b - \frac{1}{\hat{a}} \right) - \Psi(a) \right\}.$$

In equilibrium, this effort level a^{**} is equal to $\hat{a}$ and satisfies: $\bar{R}_b = \Psi'(a^{**}) + 1/a^{**}$. Consequently, a^{**} is lower than a^* , and the associated continuation value of the project π_b^{**} is lower than π_b^* . If $\pi_b^{**} < 1 < \pi_b^*$, decentralization of credit, as defined above, hardens the budget constraint of the firm.

The general insight behind this result is that decentralized finance may lead to externalities that reduce the attractiveness of refinancing. This suggests that bond or equity finance will typically involve a harder budget constraint than bank finance, a point also stressed by von Thadden (1995). Other models explaining why multiplicity of creditors can change refinancing outcomes include Bolton and Scharfstein (1995), Berglöf and von Thadden (1994), Dewatripont and Tirole (1994), and Hart and Moore (1995).

Note that, in the above setup, if π_b^{**} is bigger than 1, decentralization of credit is *worse* than centralization because the refinancing of bad projects is not prevented, but instead occurs with inefficiently low monitoring. If one allows endogenous creditor size in a market economy, however, it is possible to show that in this case a market economy will simply replicate the centralized financing pattern: in equilibrium creditors will have sufficient resources to perform the refinancing themselves (see Dewatripont and Maskin 1995). Although the market system is thus unambiguously better than exogenous centralization, this is not always true under alternative model specifications.⁵

Competition from New Projects

In transition economies, the creation of a decentralized system of credit and financial intermediation has been at the heart of recent policy debates aimed at hardening budget constraints of enterprises. The literature on soft budget constraints in banking has emphasized the importance of the quality of the loan portfolio in determining whether banks are effective in disciplining enterprises. For example, Berglöf and Roland (1997) take a variant of our model but endogenize banks' opportunity cost of refinancing. Assume at time 0 a capital of C_0 is handed over by government to a profit-maximizing bank. This capital is used to finance C_0 projects with the same characteristics as above. At time 1, the bank can use the returns generated in the

first period to finance new projects (assumed to be in infinite supply) and/or to refinance bad projects financed at time 0. After time 1, everything is exactly like it is in the section on soft budget constraints as a time-consistency problem: new projects financed at time 1 can be refinanced at time 2 and will be because $\pi_b^* > 1$.

Will bad projects be submitted at time 0? This depends on the opportunity cost of refinancing these projects at time 1 given the possibility to finance new projects. Because there will be soft budget constraints at time 2, the expected net return to a new project financed at time 1 is $\alpha(R_g - 1) + (1 - \alpha)(\pi_b^* - 2)$, whereas the net return to refinancing a bad project is $\pi_b^* - 1$. One easily sees that hard budget constraints obtain at time 1 (and thus no bad project are submitted at time 0) if:

$$\alpha > \bar{\alpha} = \frac{1}{R_g - (\pi_b^* - 1)}.$$

Instead, if $\alpha < \bar{\alpha}$, soft budget constraints obtain at time 1. One also sees that $\bar{\alpha}$ increases with π_b^* but decreases with R_g . In other words, if the expected quality of projects is high enough, hard budget constraints obtain because, even though refinancing a bad project is in itself profitable, it is less so than financing a new project. Note that by assuming that new projects would be subject to soft budget constraints at time 2, we have made the financing of new projects less attractive than if hard budget constraints were expected.

The first lesson is that soft budget constraints are not an issue if new projects are of sufficiently good quality. This may explain why soft budget constraints are not a more pervasive phenomenon in advanced market economies and why they still are in transition economies, where entrepreneurial skills are only developing. The second lesson is that when there are soft budget constraints at time 1, new projects are crowded out by the refinancing of bad projects. Indeed, under hard budget constraints, funds in amount of $C_0(R_g - 1)$ can go to new projects. Instead, under soft budget constraints, it is only $C_0(\alpha(R_g - 1) - (1 - \alpha))$ that is available for new projects because: (1) fewer returns are generated from the projects financed at time 0, and (2) bad projects must be refinanced.⁶

Rent-Seeking by Banks

Another important issue in transition economies is the relation between banks and government. We have seen above that banks may be soft because of sunk costs. Berglöf and Roland (1995) show that enterprises may still have soft budget constraints even in the case where banks have no intrinsic interest to refinance firms, that is, when $\pi_b^* \leq 1$. Indeed, even in that case, banks may benefit from exploiting

the softness of government. This will be the case if the government, who cares about total welfare, would favor refinancing because $\pi_b^* + E_b > 1$. Banks may then, under certain conditions, prefer rent-seeking in order to obtain subsidies for bailout, rather than being hard toward enterprises. Softness of banks is here related to the weakness of government. As above, softness will depend on the average quality of projects as expressed by α .

Take the same framework as the one just analyzed above. Assume that the government first gives a bank funds to finance n projects at time 0. At time 1, the bank can decide to be hard and liquidate bad projects or instead to ask the government for subsidies to refinance bad projects. Assume that the bailout money provided by government just covers the difference between the total refinancing requirements of the bank and its total funds at time 1, that is $(1 - \alpha)n - \alpha n R_g$, which we shall call G . Assume moreover that the government cannot recover this bailout money, which in effect represents subsidies. The government can, however, monitor the use of funds so that the bank cannot "take the money and run" but has to refinance firms.

Assume the bank has initially attracted a proportion $(1 - \alpha)$ of bad projects. At time 1, it will prefer rent-seeking toward government and softness toward enterprises compared to termination of bad firms if:

$$G - (1 - \alpha)n(1 - \pi_b^*) = n((1 - \alpha)\pi_b^* - \alpha R_g) \geq 0,$$

or:

$$\alpha < \overline{\alpha^H} = \frac{\pi_b^*}{R_g + \pi_b^*}.$$

As above, one obtains soft budget constraints when the proportion of good projects is below a given threshold. In this case, it is because a lower α generates less revenue at time 1 for the bank and thus enables it to obtain more subsidies. In other words, the lower the α , the lower the share of the bank in the costs of refinancing bad projects.

Berglöf and Roland (1995) further show that initial bank recapitalization allows to harden budget constraints provided banks are free to choose the number of projects they want to finance. It is then in the interest of banks to set aside enough reserves as a commitment to be hard. The lower the α , the higher this level of reserves must be, and thus the smaller the number of projects that can be financed. Low initial average project quality thus implies that hardening budget constraints has a high cost in terms of enterprise liquidity. Moreover, hard budget constraints can be obtained if at time 1, a proportion of bad loans are taken away from banks

and put into a "hospital bank," which can remove banks' incentives for rent-seeking and softness. It is, however, costly for the government, which will then bear all the cost of refinancing. One can show that such a solution is attractive for the government only if α is above a given threshold.⁷

7 Conclusion

This short survey has considered the soft budget constraint syndrome as a dynamic commitment problem. It has stressed the various uses this approach can be put to in the context of analyzing the process of transition toward market decentralization. This approach can deliver insights when evaluating instruments such as privatization, demonopolization, government sector reform, and credit market reform. In all cases, taking a dynamic commitment perspective leads one to stress the institutional dimensions of the transition process: hardening budget constraints is by no means a simple mechanical decision, but instead requires credibly altering the incentives of the government and of market participants.

Notes

This paper builds upon Maskin (1996) and Dewatripont and Roland (1996).

1. For example, Aghion, Blanchard and Burgess (1994, p. 1331) write, "subsidies, a plausible characterization of a soft budget constraint." Boycko, Shleifer, and Vishny (1995, p. 314) write, "subsidies from the Treasury to firms, also known as soft budget constraints."
2. Mitchell (1994) shows, with the help of a formal model, that banks tend to be soft with enterprises because initiation of bankruptcy would signal their fragile financial position.
3. Qian (1994) has shown how this formalization of the soft budget constraint problem may lead to the rational use of shortages by the planner as a way to alleviate this problem.
4. This differs from a static problem à la Stiglitz-Weiss (1981) where creditors can at best finance all types, and at worst solely finance bad types.
5. Although hard budget constraints can deter bad entrepreneurs from starting projects, von Thadden (1995) and Dewatripont and Maskin (1995) have pointed out that they can induce short-termism among good entrepreneurs. Specifically, introduce into the above framework the ability for good entrepreneurs to choose between their good, quick project that yields R_g and E_g after one period and a good but slow project that yields 0 after one period (and $E_l < 0$ if terminated) but a gross financial return $R_l > 2$ and a positive private benefit E_l if refinanced. These projects thus have a positive net present value but, at the end of period 1, they cannot be distinguished from bad projects. This can in fact introduce a coordination problem among good entrepreneurs. Indeed, consider the case where decentralization of credit is necessary and sufficient for a hard budget constraint, that is, $\pi_b^* > 1 > \pi_b^{**}$. One can then show that: (1) there always exists an equilibrium where credit is decentralized, the budget constraint is hard and all good entrepreneurs choose short-run projects; (2) for R_1 sufficiently large, there also exists an equilibrium where credit is centralized, the budget constraint is soft and all good entrepreneurs choose long-run projects; (3) the second equilibrium Pareto-dominates the first one. Intuitively, when $\pi_b^* > 1 > \pi_b^{**}$, decentralized financing is the only way to deter bad projects from being started. Expectations by

creditors that all long-term projects are bad are self-fulfilling, because good entrepreneurs expect termination if they do not choose quick projects. Another equilibrium is, however, possible if financing all long-term projects, good and bad, is more profitable than financing solely good, quick projects. In that case, creditors' expectations that all good entrepreneurs choose long-term projects are also self-fulfilling. In this case, the hard budget constraint equilibrium induces "short-termist" behavior, which more than offsets the gain from deterring bad long-term projects from being started.

6. We assume $a > 1/R_g$ so that the returns from the good projects can always serve to refinance the bad projects.

7. For an analysis of the hospital bank solution, see also Aghion, Bolton, and Fries (1996).

References

- Aghion, P., O. Blanchard, and R. Burgess (1994). The behavior of state firms in Eastern Europe pre-privatization. *European Economic Review* 38: 1327–1341.
- Aghion, P., P. Bolton, and S. Fries (1996). Financial Restructuring in Transition Economies. Mimeo.
- Anderson, R., and C. Kegels (1996). *Transition Banking: The Financial Development of Central and Eastern Europe*. Oxford: Oxford University Press.
- Anderson, R., E. Berglöf, and K. Mizsei (1996). *Financial Sector Development in Central and Eastern Europe*. London: CEPR.
- Begg, D., and R. Portes (1993). Enterprise debt and economic transformation: financial restructuring in Central and Eastern Europe. In C. Mayer and X. Vives, eds., *Capital Markets and Financial Intermediation*, pp. 230–254. Cambridge: Cambridge University Press.
- Berglöf, E., and E. L. von Thadden (1994). Short-term versus long-term interests: capital structure with multiple investors. *Quarterly Journal of Economics* 109: 1055–1084.
- Berglöf, E., and G. Roland (1995). Bank restructuring and soft budget constraints in financial transition. *Journal of the Japanese and International Economies* 9: 354–375.
- Berglöf, E., and G. Roland (1997). Soft budget constraints and credit crunches in financial transition. *European Economic Review* 41: 807–817.
- Bolton, P. (1995). Privatization and the separation of ownership and control: lessons from Chinese enterprise reforms. *Economic of Transition* 3(1): 1–12.
- Bolton, P., and D. Scharfstein (1996). Optimal debt structure with multiple creditors. *Journal of Political Economy* 104: 1–25.
- Boycko, M., A. Shleifer, and R. Vishny (1992). Property Rights, Soft Budget Constraints and Privatization. Mimeo, Harvard University.
- Boycko, M., A. Shleifer, and R. Vishny (1995). *Privatizing Russia*. Cambridge, Mass.: MIT Press.
- Castanheira, M., and G. Roland (1994). The Optimal Speed of Transition: A General Equilibrium Analysis. Mimeo, ECARE, Université Libre de Bruxelles.
- Castanheira, M., and G. Roland (1995). Restructuring and Capital Accumulation in Transition Economies: A General Equilibrium Perspective. Mimeo, ECARE, Université Libre de Bruxelles.
- Che, J., and Y. Qian (1994). Boundaries of the Firm and Governance: Understanding China's Township and Village Enterprises. Mimeo, Stanford University.
- Debande, O., and G. Friebe (1995). Privatization, Employment and Managerial Decision-Taking. Mimeo, ECARE, Université Libre de Bruxelles.
- Dewatripont, M., and E. Maskin (1995). Credit and efficiency in centralized and decentralized economies. *Review of Economic Studies* 62: 541–555.
- Dewatripont, M., and G. Roland (1996). Transition as a process of large-scale institutional change. In D. Kreps, ed., *Advances in Economic Theory*. Cambridge: Cambridge University Press, 240–278.

- Dewatripont, M., and J. Tirole (1994). A theory of debt and equity: diversity of securities and manager-shareholder congruence. *Quarterly Journal of Economics* 109(4): 1027–1054.
- Dittus, P. (1994). Corporate Governance in Central Europe: The Role of Banks. Mimeo, Bank of International Settlements.
- Faure-Grimaud, A. (1996). Soft Budget Constraint and Stock Price Information. Mimeo, LSE.
- Hart, O., and J. Moore (1995). Debt and seniority: an analysis of the role of hard claims in constraining management. *American Economic Review* 85: 567–585.
- Kornai, J. (1979). Resource-constrained versus demand-constrained systems. *Econometrica* 47: 801–819.
- Kornai, J. (1980). *Economics of Shortage*. Amsterdam: North-Holland.
- Kornai, J. (1990). *The Road to a Free Economy*. New York: Norton.
- Kornai, J. (1992). *The Socialist System: The Political Economy of Communism*. Oxford: Oxford University Press.
- Li, D. (1995). A Theory of Ambiguous Property Rights in Transition Economies. Mimeo, University of Michigan.
- Litwack, J. (1993). Coordination, incentives and the ratchet effect. *Rand Journal of Economics* 24(2): 271–285.
- Maskin, E. (1996). Theories of the soft budget constraint. *Japan and the World Economy* 8: 125–133.
- Mitchell, J. (1994). Strategic Creditor Passivity in Economies in Transition. Mimeo, Cornell University.
- Mitchell, J. (1995). Cancelling, Transferring or Repaying Bad Debt: Cleaning Banks' Balance Sheets in Economies in Transition. Mimeo, Cornell University.
- Perotti, E. (1993). Collusive Trade Arrears in the Stabilization of Transition Economies. Mimeo, Boston University.
- Qian, Y. (1994). A theory of shortage in socialist economies based on the "soft budget constraint." *American Economic Review* 84: 145–156.
- Qian, Y., and G. Roland (1994). Regional Decentralization and the Soft Budget Constraint: The Case of China. CEPR Discussion Paper 1013.
- Qian, Y., and C. Xu (1991). Innovation and Financial Constraints in Centralized and Decentralized Economies. Mimeo, London School of Economics.
- Schaffer, M. (1981). The credible commitment problem in the center-enterprise relationship. *Journal of Comparative Economics* 13: 359–382.
- Segal, I. (1993). Monopoly and Soft Budget Constraints. Mimeo, Harvard University.
- Shleifer, A., and R. Vishny (1994). Politicians and firms. *Quarterly Journal of Economics* 109: 995–1025.
- Stiglitz, J., and A. Weiss (1981). Credit rationing in markets with imperfect information. *American Economic Review* 71: 393–410.
- Van Wijnbergen, S. (1993). On the Role of Banks in Enterprise Restructuring: The Polish Example. Mimeo, University of Amsterdam.
- Von Thadden, E. L. (1995). Bank finance and long term investment. *Review of Economic Studies* 62: 557–575.
- Wang, Y. (1991). Economic reform, fixed capital investment expansion and inflation: a behavior model based on the Chinese experience. *China Economic Review* 2(1): 3–27.
- Weitzman, M., and C. Xu (1993). Chinese township and village enterprises as vaguely defined cooperatives. *Journal of Comparative Economics* 18: 121–145.

10 Restructuring an Industry during Transition: A Two-Period Model

Richard E. Ericson

1 Introduction

This chapter develops the two-period model of enterprise shutdown during transition that was introduced in Ericson (1994). Its purpose is to provide a formal basis for the study of what Kornai (1994) has called the “transformation of the real structure of the economy” during the transition process. It focuses at the industry level of restructuring, taking a normative industrial-policy perspective.

Most of the discussion in the transition literature of the restructuring of industry deals with creating the institutional environment, and hence the appropriate incentives, for state firms to restructure themselves and for new firms to succeed and grow. Discussions of “liberalization,” “stabilization,” implementing “privatization,” and fostering “competition,” and the analysis of the consequences of such policies, have dominated the literature.¹ There is also a growing empirical literature on the key role played by the development of “new firms,” frequently unrelated to any prior existing organizations in the economy.² Yet the focus in almost all of this literature is on the shifting of labor, employment and unemployment, and the barriers to smooth reallocation of labor between the state and private sectors. It is implicitly assumed that all other factors of production, all other aspects of industrial restructuring, will easily shift to accommodate an appropriate reallocation of labor and modernization of activity within the enterprise.³ All that is required is appropriate incentives to get managers and workers to desire, and hence to implement, the requisite changes. In so assuming, however, the general literature has underestimated the difficulties involved, difficulties arising from the structural legacy of the Soviet-type command economy.

The transition from a command to a market-based economy involves a more substantial change in the structure of production, employment, and factor use than is typically assumed, as much of the prior physical and organizational capital is not viable in a market environment.⁴ Furthermore, the market institutions, in particular financial, that mobilize and reallocate real resources for the ongoing changes characteristic of a market economy⁵ are conspicuously underdeveloped and/or absent. Thus resources currently employed in existing institutions are largely immobile, tied up by those organizations having immediate physical control. This immobility of real resources is only aggravated by the weakness or absence of well defined and protected property rights that facilitate the reallocation/transfer of real assets.⁶ Although appropriate reforms and the passage of sufficient time will ameliorate these problems as the transition proceeds, that transition’s success depends in no

small measure on the restructuring of production and other economic activity, which in turn depends on the mobility of the underlying means of production. Indeed, calls for looser monetary policy to revive enterprise cash flow and thereby stimulate investment are, ultimately, addressing this need.⁷

One way to cut through this problem is to shut down net value-destroying operations,⁸ forcing the release of those resources over which they have operational control (if not well-defined ownership). That both limits the waste of social resources, and makes possible a greater inflow of new, more productive firms. In an increasingly marketized environment this requires allowing failing firms to shut down (bankruptcy), rather than supporting them with subsidies, credits, tax holidays, or tolerating growing payments, tax and/or wage arrears. Of course, there are substantial social costs to such closing of industrial enterprises, costs which are increasing in the scale of such activity. Were there full information on the viability of firms, there would be economically clear and easy, if politically very difficult, trade-offs to be made. However, in the unstable environment of the early to mid-transition, market performance may be only a very poor signal as to the potential viability of the firm. This uncertainty introduces an "option value" to waiting and learning more about the firm's viability, as the decision to close the firm is not costlessly reversible.

As discussed in Ericson (1994), there are two types of mistakes possible: (1) shutting down a potentially viable firm based on weak early performance; and (2) maintaining a net value-destroying firm. The optimal one-time (static) decision on maintaining money losing enterprises trades off the costs of these errors in choosing a rather conservative shutdown policy. Here we continue that investigation in a two-period extension of that model, developing the analysis of the factors that determine the optimal dynamic shutdown policy for apparently failing firms of a given industry. Again it is from the perspective of a weak industrial policy maker, who can withhold state support from firms in the industry, but cannot directly manage them or create new firms in the industry; new firms arise from entrepreneurial initiative responding to perceived opportunities.

The principal result of the model is a complete characterization of the optimal shutdown policy under varying assumptions about information quality and decision horizon. The results support early decisive action in closing sufficiently "bad" firms, although the likelihood threshold (probability that a firm is viable) below which a firm is closed is still rather conservatively chosen. A longer horizon and better information about the "quality" of the firm both raise the threshold for a firm to be allowed to survive. The most aggressive optimal policy occurs under full information with a long horizon. Earlier shutdown is also seen to dominate later, given that it should occur. However, sufficiently improved information in the future may lead

to a more cautious initial policy, inasmuch as that increases the value of waiting, although that effect is usually dominated by the increased marginal value of future shutdown possibilities.

In the next section we present the notation and structure of the model. In the third we analyze the two-stage problem with discrete firms—the direct generalization of the prior model. In the fourth section we explore some policy-relevant comparative statics. Finally, we conclude with a discussion of implications for industrial restructuring in the transition economies.

2 Model Structure

Consider an industry of N firms, indexed $i = 1, \dots, N$. The current net value produced by firm i is given by an i.i.d. random variable η_i with fixed finite support $W = \{w_1, \dots, w_k\}$ and distribution $P\{\eta|\theta_i\}$, where $\theta_i \in \{\underline{\theta}, \bar{\theta}\}$ indicates whether the firm is “bad” or “good,” that is, able to survive in the coming market economy. Let

$$\begin{aligned} p_{\eta}^+ &\equiv P\{\eta|\bar{\theta}\}; & p_{\eta}^- &\equiv P\{\eta|\underline{\theta}\}; \\ y_1 &= E\{\eta|\bar{\theta}\} > 0; & y_0 &= E\{\eta|\underline{\theta}\} < 0; \end{aligned} \quad (10.1)$$

and assume $p^+ \succ p^-$ in the sense of first order stochastic dominance:

$$\sum_{\eta \geq a} p_{\eta}^+ \geq \sum_{\eta \geq a} p_{\eta}^-, \quad \forall a \in R. \quad (10.2)$$

Because θ_i is not observed, let $\pi_i \in (0, 1)$ be the prior estimate that firm i is “good” (of quality $\bar{\theta}$) and $\hat{\pi}_i(\eta)$ be its Bayesian update on observing η :

$$\hat{\pi}(\eta) = \frac{\pi \cdot p_{\eta}^+}{\pi \cdot p_{\eta}^+ + (1 - \pi)p_{\eta}^-}. \quad (10.3)$$

Then the current expected net return to operating a firm with estimate of “quality” π is:

$$y_{\pi} = E\{\eta|\pi\} \equiv \pi \cdot (y_1 - y_0) + y_0. \quad (10.4)$$

If

$$\pi < \bar{\pi} \equiv \frac{-y_0}{y_1 - y_0},$$

then $y_{\pi} < 0$ and the firm is expected to be net value—destroying, that is, in all likelihood that $\theta = \underline{\theta}$, while $\pi \geq \bar{\pi}$ means that the firm is expected to contribute to social

wealth. The expected net social returns to the functioning of all firms in the industry is then given by

$$Y = \sum_{i=1}^N y_{\pi_i}. \quad (10.5)$$

This includes negative expected returns generated by those firms operating with $\pi_i < \bar{\pi}$, and hence could be increased by shutting those firms down.

Closing firms has three impacts: (1) the loss of the (negative) net product of the closed firms (Σy_k); (2) the costs associated with closure, disruption of economic ties, and maintenance of the released factors of production are imposed; and (3) the prospects for new firms entering the industry are enhanced, due to diminished competition and increased availability of released resources to support their operation. The second effect is modeled by a convex, increasing function, $C(m)$, where m is the number (mass/volume) of firms shutdown in the period, and a function giving the continuing (social) cost of maintaining the unemployed resources of the industry, $c(L)$, where L is the number (mass) of potential new firms that could be formed using the available resources. The final effect is captured in a probability measure over the entry of new firms, v^L , which gives the probability that some number (mass), ℓ , of new entrants will operate in the period: $v^L(\ell)$. In the simple discrete case it is given by:

$$\begin{aligned} \forall L \geq 1, \quad v_\ell^L > 0; \\ (\ell-1)v_{\ell-1}^L \geq \ell \cdot v_\ell^L, \quad \ell \in \{2, \dots, L\}; \\ v_0^L = \left(1 - \sum_{\ell=1}^L v_\ell^L\right), \end{aligned} \quad (10.6)$$

where $v_\ell^L \equiv v^L(\ell)$.⁹ We assume that entry takes time, and hence entrants come out of preexisting resources; current shutdowns only add resources to the next period's entry pool. By the selection process that generates new entry, entrants must be *expected* to be "good" (type $\bar{\theta}$), although they may actually turn out to be of type $\underline{\theta}$. Hence we assume a prior quality estimate of $q > \bar{\pi}$ yielding expected firm payoff

$$y_q = q \cdot (y_1 - y_0) + y_0, \quad (10.7)$$

and expected net social payoff from new entrants,

$$Y_e = \sum_{\ell=1}^L v_\ell^L \cdot \ell \cdot y_q > 0$$

when the pool of available resources is L .

The objective of the industrial policy maker in this model is to maximize net social wealth over the course of the transition by a judicious choice of shutdowns—the order and timing of closure of (presumably) “bad” firms. We assume that controlling new entry is beyond the capabilities of the policy maker, and hence arises exogenously from “the market.” The policy maker can, however, “fuel” entry by releasing resources to the market through the shutdown (i.e., failure to subsidize sufficiently) of existing operating enterprises. The general dynamic problem, given total social resources capable of supporting the operation of a set N of firms, facing the policy maker is thus:

$$\max_{\{M_t\}_{t=0}^T} \left\{ \sum_{t=0}^T \beta^t E \left(\sum_{i \in (N \setminus \Lambda_t) \setminus M_t} \eta_{it} + \sum_{j \in \Lambda_t} \eta_j + |\Lambda_t| \cdot r - C(M_t) - c(L_t) \middle| \mathcal{F}_t \right) \right\}, \quad (10.8)$$

where M_t is the set of firms to be shut, L_t —the set of potential (“latent”) firms, and $\Lambda_t \subset L_t$ —the set of new entrants at time t , r is the recovered cost of a latent firm that becomes active, $c(L) = c \cdot L$ is the continuing cost of maintaining the viability of the resource pool L , and $\mathcal{F}_t$ is the information available at time t . The solution to this problem yields a value depending on the initial payoff-relevant parameters (state variables) of the problem: $V^T(\pi, N_0, L_0)$.¹⁰ This problem’s solution also generates the optimal sequences $\{(M_t, \pi_t^*), N_t, L_t\}_{t=0}^T$, where π_t^* is the optimal cutoff expectation for shutting down firms (see below). Clearly $N_{t+1} = (N_t \setminus M_t) \cup \Lambda_t$, $L_{t+1} = (L_t \setminus \Lambda_t) \cup M_t$, and $N_t + L_t \equiv N$.¹¹

3 Optimal Shutdown with Two Periods

In Ericson (1994) I analyzed the discrete, single-period, case [$T = 0$], with immediate new entry (using resources of the firms shutdown in the same period) possible, and firms all of the same resource size (a “standard firm” assumption). The primary analytic result was a complete characterization of the optimal shutdown policy, which we repeat here in the case where immediate reuse of resources released by shutdown is not possible. The result is illustrated in figure 10.1.

PROPOSITION 10.1. There exists a unique solution, (m^*, π^*, σ^*) , to the optimization problem (10.8) with $T = 0$ such that

$$\begin{aligned} \pi_{m^*} &\leq \pi^* \leq \pi_{m^*+1}, \\ -y_{\pi^*} &= \Delta C(m^*), \end{aligned} \quad (10.9)$$

$$\Delta W_{m^*} \geq 0 > \Delta W_{m^*+1},$$

where

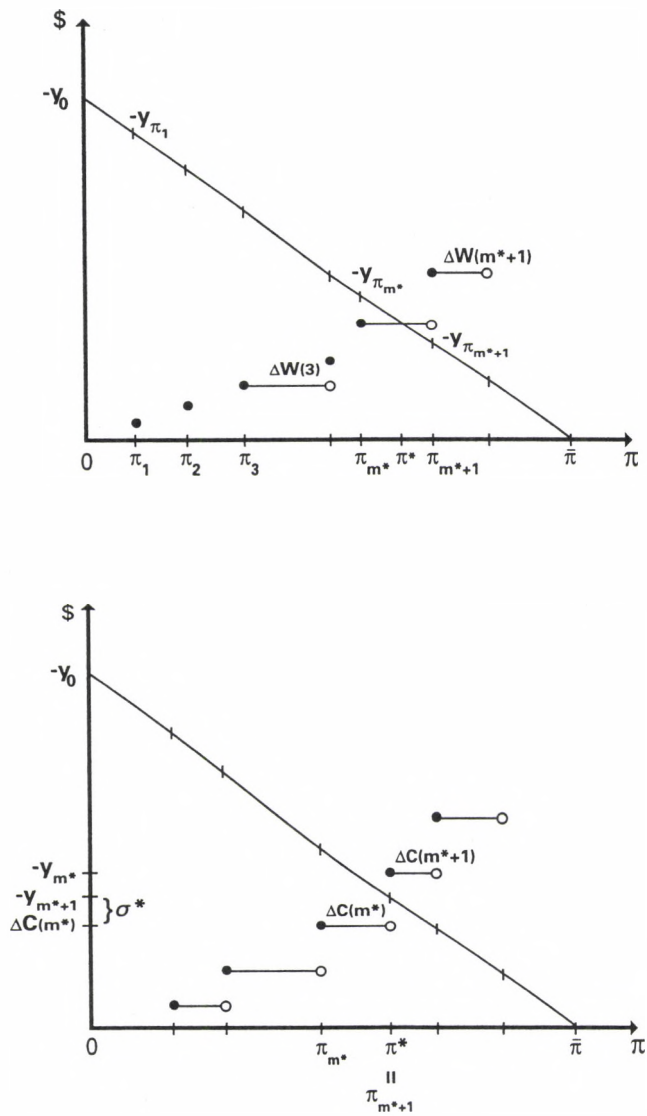


Figure 10.1
Optimal shutdown thresholds in the 1-period case

$$\begin{aligned}
W_m &= E \left\{ \sum_{i=m+1}^{N_0} \eta_i + \sum_{j=1}^{\ell} \eta_j + \ell \cdot r - C(m) - c \cdot L \middle| \mathcal{F} \right\} \\
&= \sum_{i=m+1}^{N_0} y_i + \sum_{j=1}^{L_0} j \cdot [y_q + r] \cdot v_j - C(m) - c \cdot L
\end{aligned}$$

$\Delta C(m) \equiv C(m) - C(m-1)$ and $\Delta W_m \equiv W_m - W_{m-1}$, implies $\sigma^* = 0$.

If

$$\pi^* = \pi_{m^*+1} \quad \text{and} \quad -y_{\pi^*} > \Delta C(m^*),$$

then

$$-y_{\pi^*} = (1 - \sigma^*) \cdot \Delta C(m^*) + \sigma^* \cdot \Delta C(m^* + 1), \quad (10.10)$$

$$(1 - \sigma^*) \cdot \Delta W_{m^*} + \sigma^* \cdot \Delta W_{m^*+1} = 0,$$

and $\sigma^* > 0$.

Furthermore, $\pi^* < \bar{\pi}$ as $\Delta C(\cdot) > 0$.

Proof. See propositions 1 and 2 in Ericson (1994). ■

Remark 1. This is a simpler formulation than in Ericson (1994) as only resources lying fallow at the beginning can be drawn into new enterprises. Thus the impact of shutdown on new entry can only be felt in the future, reducing somewhat the incentive to close failing firms. ■

Remark 2. The first order conditions [(10.9) and (10.10)] follow immediately from placing the firms in increasing order of their likelihood of being good (π 's); then the worst firms, those with the lowest indices, are the first candidates to be shut down. Thus W_m gives the expected net value of operation of the industry conditional on m firms being shut down. ■

Remark 3. The expected value of this problem is: $V^0(\pi, N_0, L_0) = W_{m^*}$. Notice that only m^* firms are ever shut down, even when the conditions (10.10) are relevant; it is never desirable to even randomize over shutting down an additional firm, as that only adds to the expected cost. σ^* is just a measure of slack in the program due to its integer nature—the amount by which $-y_{\pi^*}$ falls short of allowing another firm to be optimally shut. It is useful, as we shall see below, for characterizing the impact of changes in the state variables (π, N, L) on the optimal decision and hence on the value of the problem. ■

Remark 4. With linear shutdown costs ($\Delta C(m) = C, \forall m \geq 1$) only the first order conditions (10.1) are relevant:

$$-y_{\pi^*}^* = C$$

and

$$\pi^* = -\left(\frac{y_0 + C}{y_1 - y_0}\right). \blacksquare$$

Here we extend the analysis with “standard firms” to two periods ($T = 1$), with delayed entry and with an initial set of latent firms that might enter independently of the shutdown decision. That decision, however, affects entry in the next period. The results here provide a foundation for a general T -period analysis.

3.1 Terminal Period Problem

In the dynamic problem, the optimal initial decision of the policy maker must take into account what he will decide to do in the future in the situation created by that initial decision. Thus we must begin with an analysis of the static second period problem where there are $N_1 = N \setminus M_0$ operating firms with “qualities” (likelihood of viability) $\hat{\pi} = (\hat{\pi}_1, \dots, \hat{\pi}_{n_1})$ arranged in increasing order, $L_1 = L_0 \cup M_0$ potential entrants each with expected quality q , and $c(L) = c \cdot L$.

Let us begin with the simplest case: no new entry is possible in the first period ($L_0 = \emptyset$).¹² N_1 is then just the set of initial firms that were not shut down, that is, such that $\pi_i \geq \pi^*$ —the optimal first period cutoff expectation. The set of potential entrants, from which some random ℓ standard firms will enter this (second) period, is built from the resources of the m^* firms optimally shut down in the first period. Thus the expected value of the problem for each possible realization of $\hat{\pi}$ is given by

$$V^0(\hat{\pi}, N - m^*, m^*) = \max_{0 \leq m \leq N - m^*} \left\{ \sum_{i=m^*+m+1}^N \hat{y}_i + \sum_{j=1}^{m^*} j \cdot [y_q + r] \cdot v_j - C(m) - c \cdot m^* \right\} \quad (10.11)$$

where $\hat{\pi}$ contains updated likelihoods of being good for firms that are active, $\hat{y}_i \equiv y_{\hat{\pi}_i}$, $N_1 = N - m^*$, and $L_1 = m^*$.¹³ The solution to this problem, which we label $(\tilde{m}, \tilde{\pi}, \tilde{\sigma})$, is given directly by Proposition 10.1:

$$V^0(\hat{\pi}, N - m^*, m^*) = \sum_{i=\tilde{m}+1}^N \hat{y}_i + E\{\ell | m^*\} \cdot \{y_q + r\} - C(\tilde{m}) - c \cdot m^*, \quad (10.12)$$

where $\Delta C(\tilde{m} + 1) > -\hat{y}_{\tilde{m}} \geq \Delta C(\tilde{m})$ and $\tilde{\sigma} = 0$, $\pi_{\tilde{m}} \leq \tilde{\pi} \leq \pi_{\tilde{m}+1}$ or $\tilde{\sigma} > 0$, $\tilde{\pi} = \pi_{\tilde{m}+1}$. Note that there only a finite number of such problems possible, one for each $(N - m^*)$ -tuple of η -realizations of net value. This characterization allows us to incorporate the discounted $V^0(\hat{\pi}, N - m^*, m^*)$ in the first period problem, inasmuch as it fully

captures the impact of the first period shutdown decision on the industry's future value.

The characterization becomes more complex when we begin with a nontrivial set of latent forms, $L_0 \neq \emptyset$, from which new entrants, not subject to the first period's shutdown threshold, might appear. There is then a similar, but different, problem for each realization of outcomes of surviving firms, for each possible number of new entrants, $0 \leq \hat{\ell} \leq L_0$, and for each realization of outcomes of the new firms. For any such problem, $L_1 = L_0 + m^* - \hat{\ell}$, $N_1 = N_0 - m^* + \hat{\ell}$, and $\hat{\pi}$ contains an $(N_0 - m^*)$ -subvector of updated (from priors π_t —see equation (10.3)) expectations for the survivors and an $\hat{\ell}$ -subvector of updated (from prior q) expected payoffs for the new entrants. Then the problem faced by the industry planner is:

$$\begin{aligned} \max_{m, n \geq 0} W(m, n; \hat{\pi}, N_1, L_1) &= \sum_{i=m^*+m+1}^{N_0} \hat{y}_i + \sum_{j=n+1}^{\hat{\ell}} \hat{y}_j + [y_q + r] \cdot \sum_{h=1}^{L_0+m^*-\hat{\ell}} h \cdot v_h^{L_1} \\ &\quad - C(m+n) - c \cdot (L_0 + m^* - \hat{\ell}) \\ &= \sum_{k=m+1}^{N_0-m^*+\hat{\ell}} \hat{y}_k + [y_q + r] \cdot \sum_{h=1}^{L_0+m^*-\hat{\ell}} h \cdot v_h^{L_1} - C(k) - cL_1. \end{aligned} \quad (10.13)$$

It is straightforward, if somewhat tedious, to derive the following

PROPOSITION 10.2. Let $(\hat{\pi}, \hat{\ell})$ be the realization of expected productivities and new entrants at the beginning of the terminal period. Then there exists a unique solution, $(\tilde{k}, \tilde{\pi}, \tilde{\sigma})$, $\tilde{k} = \tilde{m} + \tilde{n}$, to the problem (10.5) such that

$$\begin{aligned} \pi_{\tilde{k}} &= \max\{\pi_{\tilde{m}}, \pi_{\tilde{n}}\} \leq \tilde{\pi} \leq \min\{\pi_{\tilde{m}+1}, \pi_{\tilde{n}+1}\}, \\ -y_{\tilde{\pi}} &= \Delta C(\tilde{m} + \tilde{n}), \end{aligned} \quad (10.14)$$

$$\tilde{\sigma} = 0,$$

or

$$\begin{aligned} \tilde{\pi} &= \max\{\pi_{\tilde{m}+1}, \pi_{\tilde{n}+1}\}, \\ -y_{\tilde{\pi}} &= (1 - \tilde{\sigma})\Delta C(\tilde{k}) + \tilde{\sigma}\Delta C(\tilde{k} + 1), \end{aligned} \quad (10.15)$$

$$\tilde{\sigma} > 0.$$

Furthermore, $\tilde{\pi} \geq \pi^*$ unless $\hat{\ell} > m^*$ and $\hat{\pi}_j < \tilde{\pi}$ for at least $m^* + 1$ of the $\hat{\ell}$ new entrants, which will occur with probability no greater than

$$\left(\sum_{k=m^*}^{L_1} v_k^{L_1} \right) \times [\text{Prob}\{\hat{\pi} < \pi^* | \pi = q\}]^{m^*+1}. \quad (10.16)$$

Proof. [Outline] The proof is virtually identical to that of Proposition 10.1 (see Ericson 1994 for more details), once both sets of firm expected values ($\hat{y}_i$'s) are placed in increasing order. The marginal cutoff is clearly determined by $\Delta C(k)$, so that firms are shut down from both sets starting with the lowest $\hat{\pi}$'s, until the next shutdown (from either set) will cause the added cost to exceed the savings from that shutdown. Whether that occurs between firm expected values or at the expected value of the next best firm determines the difference between (10.6) and (10.7).

Clearly, if fewer than m^* (necessarily new) firms have expected qualities below π^* then ΔC for shutting down all those firms must be *less than* $-y_{\pi^*}$, so new firms with higher π 's will be optimally shut down. Direct calculation then gives (3.8) as an upper bound on the probability that the quality cutoff will fall in the second period. ■

Remark 5. Note that the probability in (10.16) is extremely small; indeed it is often zero. It naturally depends on the structural probabilities in the problem and the relationship between q and π^* . If the realizations of η are insufficient to drive the $\hat{\pi}$ derived from q to below π^* then that probability is zero. As that distance is quite large,

$$q - \pi^* = \frac{y_q + \Delta C(m^*)}{y_1 - y_0},$$

for $\tilde{\pi} < \pi^*$ at least $m^* + 1$ firms must draw an η such that

$$\left(\frac{p_{\eta}^-}{p_{\eta}^+} \right) > \frac{(1 - \pi^*) \cdot q}{\pi^* \cdot (1 - q)} = \frac{(y_1 + \Delta C(m^*))(y_q - y_0)}{(y_1 - y_q)(-y_0 - \Delta C(m^*))}$$

which can be, depending on the parameters of the problem, quite large. For example, if $y_1 = -y_0 = 5$, $\Delta C(m^*) = 2$ and $y_q = 1$, then

$$\left(\frac{p_{\eta}^-}{p_{\eta}^+} \right) > 3.5$$

for the η 's drawn by $m^* + 1$ firms for the probability that $\tilde{\pi} < \pi^*$ holds to be positive. ■

Remark 6. The notation has suppressed dependence of the state, $(\hat{\pi}, N_1, L_1)$, on the prior shutdown decision, m^* , although that dependence is obvious in the structure of (10.5). Thus it is clear that the current decision will depend nontrivially on m^* : $\tilde{k}(m^*)$, $\tilde{\pi}(m^*)$, $\tilde{\sigma}(m^*)$. ■

These results give, for each $(\hat{\pi}, \hat{\ell})$ realization, a well defined value function for this terminal period problem from which we can derive the future impact of shutdown decisions in the preceding period:

$$V^0(\hat{\pi}, N_1, L_1) = \sum_{i=k+1}^{N_1} \hat{y}_i + [y_q + r] \left(\sum_{j=1}^{L_1} j \cdot v_j^{L_1} \right) - C(\tilde{k}) - cL_1, \quad (10.17)$$

where $N_1 = N_0 - m^* + \hat{\ell}$, $L_1 = L_0 - m^* - \hat{\ell}$ and $(\tilde{k}, \tilde{\pi}, \tilde{\sigma})$ is the optimal decision in the terminal period. Notice that this reduces to (10.13) when $L_0 = 0$, implying that $\hat{\ell} = 0$.

3.2 Initial Period Problem

We can now write the Bellman equation for the first period problem ($T = 1$) when $L_0 = 0$,

$$V^1(\pi, N, 0) = \max_{0 \leq m^* \leq N} \left\{ \sum_{i=m^*+1}^N y_i - C(m^*) + \beta E[V^0(\hat{\pi}, N - m^*, m^*) | \mathcal{F}_0] \right\}, \quad (10.18)$$

and when $L_0 > 0$,

$$V^1(\pi, N_0, L_0) = \max_{0 \leq m^* \leq N_0} \left\{ \sum_{i=m^*+1}^{N_0} y_i - C(m^*) - cL_0 + \sum_{j=1}^{L_0} v_j^{L_0} (j \cdot [y_q + r] + \beta E[V^0(\hat{\pi}, N_0 - m^* + j, L_0 + m^* - j) | \mathcal{F}_0]) \right\} \quad (10.19)$$

Clearly the optimal policies for each of these problems are qualitatively similar, because (10.18) is just a special case of (10.19). This allows generalization to any finite horizon, as well as to the infinite horizon stationary problem.¹⁴

The solution to this problem, an optimal first-period shutdown decision, must be based only on the information available in the first period. Thus iterated expectations must be calculated as $V^0(\cdot)$ is a function of $\mathcal{F}_1 \supset \mathcal{F}_0$. Hence a useful property, a consequence of the Martingale property of Bayesian updating, is given by the following.

LEMMA 10.3. For ℓ fixed, $E[V^0(\hat{\pi}, N_0 - m^* + \ell, L_0 + m^* - \ell) | \mathcal{F}_0] = V^0((\pi^0, q), N_0 - m^* + \ell, L_0 + m^* - \ell)$, where π^0 is the $(N_0 - m^*)$ -vector of initial firm likelihoods for surviving firms, and q is the ℓ -vector of prior quality evaluations of new entrants.

Proof. For any firm (dropping the subscript i) $\hat{\pi}(\eta)$ is given by equation (10.3). Let E_0 be the operator $E[\cdot | \mathcal{F}_0]$. Then, if a firm is “good” $[\bar{\theta}]$,

$$E_0(\hat{\pi} | \bar{\theta}) = \sum_{\eta} p_{\eta}^+ \cdot \frac{\pi \cdot p_{\eta}^+}{(1 - \pi) \cdot p_{\eta}^- + \pi \cdot p_{\eta}^+} = \pi \sum_{\eta} \frac{(p_{\eta}^+)^2}{(1 - \pi) \cdot p_{\eta}^- + \pi \cdot p_{\eta}^+},$$

whereas if it is “bad” $[\underline{\theta}]$ then

$$E_0(\hat{\pi}|\bar{\theta}) = \sum_{\eta} p_{\eta}^{-} \cdot \frac{\pi \cdot p_{\eta}^{+}}{(1-\pi) \cdot p_{\eta}^{-} + \pi \cdot p_{\eta}^{+}} = \pi \sum_{\eta} \frac{p_{\eta}^{-} \cdot p_{\eta}^{+}}{(1-\pi) \cdot p_{\eta}^{-} + \pi \cdot p_{\eta}^{+}}.$$

Therefore,

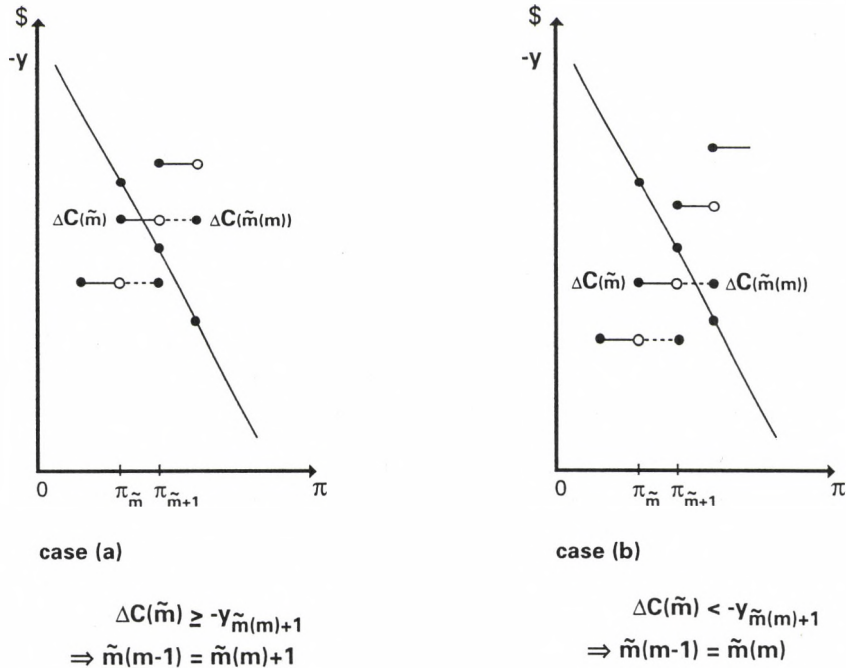
$$\begin{aligned} E(\hat{\pi}) &= \pi \cdot E(\tilde{\pi}|\bar{\theta}) + (1-\pi) \cdot E(\tilde{\pi}|\underline{\theta}) \\ &= \pi \sum_{\eta} \frac{\pi(p_{\eta}^{+})^2 + (1-\pi)p_{\eta}^{+}p_{\eta}^{-}}{(1-\pi) \cdot p_{\eta}^{-} + \pi \cdot p_{\eta}^{+}} \\ &= \pi \sum_{\eta} p_{\eta}^{+} = \pi. \end{aligned}$$

Thus $E_0(\hat{y}_i) = y_i$ for any firm initially in operation, $E_0(\hat{y}_i) = q$ for any entrant, and the expected future value, *conditional on initial information and the number of new entrants*, is unaffected by the output realizations at the end of the first period. ■

The solution to (10.17) and (10.18) clearly depends on how $V^0(\cdot)$ varies in m , the number of firms initially shut down, and in the latter case, on how that varies in the number of new entrants (See remark 6). Let $V_{\ell}^0(m; \cdot)$ represent the optimized value in the terminal period [see (10.16)] as a function of the number of firms shut down in the prior period, m , and the number of new firms, ℓ . That value is generated by the solution $\tilde{m}_{\ell}(m)$ to the terminal problem (10.5) given in Proposition 10.2.

LEMMA 10.4. $\tilde{m}_{\ell}(m) + 1 \geq \tilde{m}_{\ell}(m-1) \geq \tilde{m}_{\ell}(m)$ for all feasible (m, ℓ) , and $\tilde{m}_{\ell}(m) = \tilde{m}_k(m)$ for all feasible (m, ℓ, k) . Hence $\tilde{m}_{\ell}(m)$ is non-increasing in m , and independent of ℓ .

Proof. The conditions determining the choice of $\tilde{m}_{\ell}(m)$ are in equations (10.14) and (10.15). Because there is no dependence on the number of new firms operating in that period, the second assertion is obvious, and we can wlog drop ℓ from our expressions here. Let $\tilde{m} \equiv \tilde{m}(m-1)$ be the number closed, and $k(\tilde{m})$ be the index of that marginal firm. As the lowest expected value firms are shutdown first in any period, closing one more means that there is one less low π_i , so that the same cutoff likelihood, $\tilde{\pi}$, can be reached with one fewer shutdown, that is, at a saving of $\Delta C(\tilde{m})$. Whether that is the optimal response to the extra initial shutdown depends on whether $-y_{k(\tilde{m})+1} \geq \Delta C(\tilde{m})$. If so, $\tilde{m}(m) = \tilde{m}$; the same number of firms are shut down, despite one more having been closed earlier, but $\tilde{\pi}$ increases. If not, then $\tilde{m}(m) = \tilde{m} - 1$, because it is too costly to close down another firm, so $\tilde{\pi}$ remains the same. Notice that a single extra prior shutdown will never lead to more than $\tilde{m}$ closures in the second period, as optimization insures that $\Delta C(\tilde{m} + 1) > -y_{k(\tilde{m})}$. ■

**Figure 10.2**

Variation in prior shutdown of the (period 2) shutdown margin

The different cases in this lemma are illustrated in figure 10.2. This allows us to readily characterize the impact of the prior choice of shutdowns on the subsequent value of the problem. Let $\Delta V_\ell^0(m) = V_\ell^0(m; \cdot) - V_\ell^0(m-1; \cdot)$ be the total impact of increasing prior shutdowns by one on the expected value in the terminal period, and let $\Delta V_\ell^e(m)$ be the impact of increasing the number of latent firms by one through shutting down an m -th firm.

LEMMA 10.5. The variation in the value function due to a unit increase in prior shutdown is given by:

1. $\Delta V_\ell^e(m) = (L_0 - \ell + m) v_{(L_0 + \ell - m)}^{L_1} [y_q + r] - c.$
2. $\Delta V_\ell^0(m) = -y_{\pi(m)} + \Delta V_\ell^e(m).$

Proof. (1) follows from direct calculation:

$$\Delta V_\ell^e(m) = [y_q + r] \sum_{j=1}^{L_0 - \ell + m} j \cdot v_j^{L_0 - \ell + m} - (L_0 - \ell + m)c$$

$$\begin{aligned}
& -[y_q + r] \sum_{j=1}^{L_0 - \ell + m - 1} j \cdot v_j^{L_0 - \ell + m - 1} - (L_0 - \ell + m - 1)c \\
& = -(L_0 - \ell + m)v_{(L_0 - \ell + m)}^{L_1}[y_q + r] - c.
\end{aligned}$$

(2) is immediate from the first order conditions in Proposition 10.10 [(10.14) or (10.15)], where clearly $-y_{\tilde{\pi}(m)} = \max \{-y_{\tilde{m}(m)+1}, \Delta C(\tilde{m}(m))\}$. ■

Remark 7. Although the above results are formulated for the problem (10.19), they all hold as well for (10.18) with $L_1 = m$ ($L_0 = \ell = 0$). ■

Remark 8. These are the key comparative statics results that we need. We could also readily calculate the impact of independent changes in $\hat{\pi}$ or N_1 or L_1 as done in proposition 4 of Ericson (1994). The qualitative results would be the same. ■

Remark 9. Notice that $\Delta V_\ell^0(m)$ is a decreasing function of m , as both its components are such. This ensures that second order conditions are automatically satisfied. ■

The main result of the two-period analysis is the derivation of an optimal policy that can be compared with other, perhaps institutionally constrained, approaches to restructuring of the industry. Although it is by definition better than any other shutdown policy, it has a dynamic structure that is interesting to compare with that of other policies, for example, one in which all shutdown decisions must be made in the initial period (perhaps because politics will not allow a continuation). The key trade-offs revolve around three factors: “bad” firms will continue destroying social wealth if not closed, new entry will not be forthcoming without early shutdowns, but the convexity of full shutdown costs implies that it is socially wasteful to shut-down too many firms too rapidly—those costs are minimized by spreading evenly.

To see this, define ΔW_m to be the variation in $W^1(\pi, N_0, L_0)$ as the m -th firm (starting from the lowest π) is shut down:

$$\Delta W_m = W_m^1(\pi, N_0, L_0) - W_{m-1}^1(\pi, N_0, L_0)$$

where

$$\begin{aligned}
W_m^1(\pi, N_0, L_0) = & \sum_{i=m+1}^{N_0} y_i - C(m) - cL_0 \\
& + \sum_{j=1}^{L_0} v_j^{L_0} (j \cdot [y_q + r] + \beta E[V^0(\hat{\pi}, N_0 - m + j, L_0 + m - j) | \mathcal{F}_0]).
\end{aligned} \tag{10.20}$$

Pulling together and integrating the above results then gives us:

PROPOSITION 10.6. The unique optimal policy for the two-period problem (10.19), $\{(m^*, \pi^*, \sigma^*), (\tilde{k}(m^*), \tilde{\pi}(m^*), \tilde{\sigma}(m^*))\}$, gives total expected value

$$V^1(\pi, N_0, L_0) = W_{m^*}^1(\pi, N_0, L_0)$$

and satisfies the second period marginal conditions

$$\pi_{\tilde{k}} \leq \tilde{\pi} \leq \pi_{\tilde{k}+1}, \quad \tilde{\sigma} = 0,$$

$$-y_{\tilde{\pi}} = \Delta C(\tilde{k}),$$

or

$$\tilde{\pi} = \pi_{\tilde{k}+1}, \quad \tilde{\sigma} > 0,$$

$$-y_{\tilde{\pi}} = (1 - \tilde{\sigma})\Delta C(\tilde{k}) + \tilde{\sigma}\Delta C(\tilde{k} + 1),$$

and the first period marginal conditions

$$\pi_{m^*} \leq \pi^* \leq \pi_{m^*+1}, \quad \sigma^* = 0,$$

$$-(y_{\pi^*} + \beta[y_{\tilde{\pi}(m^*)} - \xi_{m^*}(y_q + r)]) = \Delta C(m^*) + \beta c, \quad (10.21)$$

or

$$\pi^* = \pi_{m^*+1}, \quad \sigma^* > 0,$$

$$-(y_{\pi^*} + \beta[y_{\tilde{\pi}(m^*)} - \xi_{m^*}(y_q + r)]) = (1 - \sigma^*)\Delta C(m^*) - \sigma^*\Delta C(m^* + 1) + \beta c, \quad (10.22)$$

where $\xi_m = \sum_{j=0}^{L_0} \{v_j^{L_0}(L_0 + m^* - j)v_{L_0+m^*-j}^{L_0+m^*-j}\}$,¹⁵ and $\xi_{m^*}(y_q + r)$ is the prior expected value of the marginal entrant when m^* firms are shut. Furthermore, if there are any firms with $\pi_i \in [\pi^*, \bar{\pi})$,

$$\bar{\pi} > E_0[\tilde{\pi}(m^*)] > \pi^*, \quad E_0[\tilde{m}(m^*)] > 0, \quad (10.23)$$

so that shutdown continues to increasingly high quality, albeit still expected value destroying, firms despite the prior elimination of the lowest quality ($\pi_i \leq \pi^*$).

Proof. The results follow from the arguments of Propositions 10.9 and 10.10 and the results of lemmata 10.3, 10.4, and 10.5. The second period characterization is just that of proposition 10.2. To derive the remaining results, notice that lemma 10.3 gives

$$\begin{aligned} W_m^1(\pi, N_0, L_0) &= \sum_{i=m+1}^{N_0} y_i - C(m) - cL_0 \\ &\quad + \sum_{j=0}^{L_0} v_j^{L_0} (j \cdot [y_q + r] + \beta \cdot V^0((\pi, q), N_0 - m + j, L_0 + m - j)) \end{aligned} \quad (10.24)$$

so that by lemma 10.5 the first variation of the total payoff in the initial decision is:

$$\begin{aligned}
 \Delta W_m &= -y_m - \Delta C(m) + \beta \left\{ \sum_{j=0}^{L_0} v_j^{L_0} [\Delta V_j^0(m)] \right\} \\
 &= -y_m - \beta y_{\bar{\pi}(m)} - (\Delta C(m) + \beta c) + \beta \left\{ \sum_{j=0}^{L_0} v_j^{L_0} [\Delta V_j^e(m)] \right\} \\
 &= -y_m - \beta y_{\bar{\pi}(m)} - (\Delta C(m) + \beta c) \\
 &\quad + \beta [y_q + r] \left\{ \sum_{j=0}^{L_0} v_j^{L_0} (L_0 + m - j) v_{L_0+m-j}^{L_0+m-j} \right\} \\
 &= -y_m - \beta y_{\bar{\pi}(m)} - (\Delta C(m) + \beta c) + \beta [y_q + r] \xi_m.
 \end{aligned} \tag{10.25}$$

As ΔW_m is monotonically decreasing in m , $\Delta W_1 > 0$, and $\Delta W_m < 0$ for any m such that $\pi_m \geq \bar{\pi}$, the optimum occurs at m^* such that

$$\Delta W_{m^*} \geq 0 \geq \Delta W_{m^*+1},$$

that is,

$$\begin{aligned}
 -(y_{m^*} + \beta [y_{\bar{\pi}(m^*)} - \xi_{m^*}(y_q + r)]) &\geq \Delta C(m^*) + \beta c, \\
 -(y_{m^*+1} + \beta [y_{\bar{\pi}(m^*+1)} - \xi_{m^*+1}(y_q + r)]) &< \Delta C(m^* + 1) + \beta c.
 \end{aligned} \tag{10.26}$$

If there exists a $\pi^* \in [\pi_{m^*}, \pi_{m^*+1}]$ such that $-(y_{\pi^*} + \beta [y_{\bar{\pi}(m^*)} - \xi_{m^*}(y_q + r)]) = \Delta C(m^*) + \beta c$, then we are in case (10.20). If not, then the relevant FOC's are given by (10.21).

To see (10.22), first note that $\Delta C(1) > 0$ implies that, without a future, no firm with positive expected payoff will ever be shut; indeed, no firm with $-y_i < \Delta C(1)$ will be shut, unless there are sufficient future savings and expected value from entry induced by that closing. That $E_0[\bar{\pi}(m^*)] \geq \pi^*$ follows immediately from Lemma 10.3 and the fact that the expected payoff to all new entrants is $y_q > 0$.

Thus the only firms that might be expected to be shut down are those with $\pi_i \in [\pi^*, \bar{\pi}]$, so the expected cutoff must be in that region. Because $\Delta C(m)$ is increasing in m , and $-y_{\pi^*} = \Delta C(m^*) \gg \Delta C(1)$, if there are any firms with $\pi_i \in [\pi^*, \bar{\pi}]$ at least one of them must be optimally shut down. Hence $\bar{\pi} > E_0[\bar{\pi}(m^*)] > \pi^*$, $E_0[\bar{m}(m^*)] > 0$. ■

The structure of the solution to this industrial policy problem is illustrated in figure 10.3, which shows how the second period solution builds in the first. This proposition provides the basis for a number of comparative statics/dynamics exercises, exploring the dynamic structure of an optimal industrial policy during the tran-

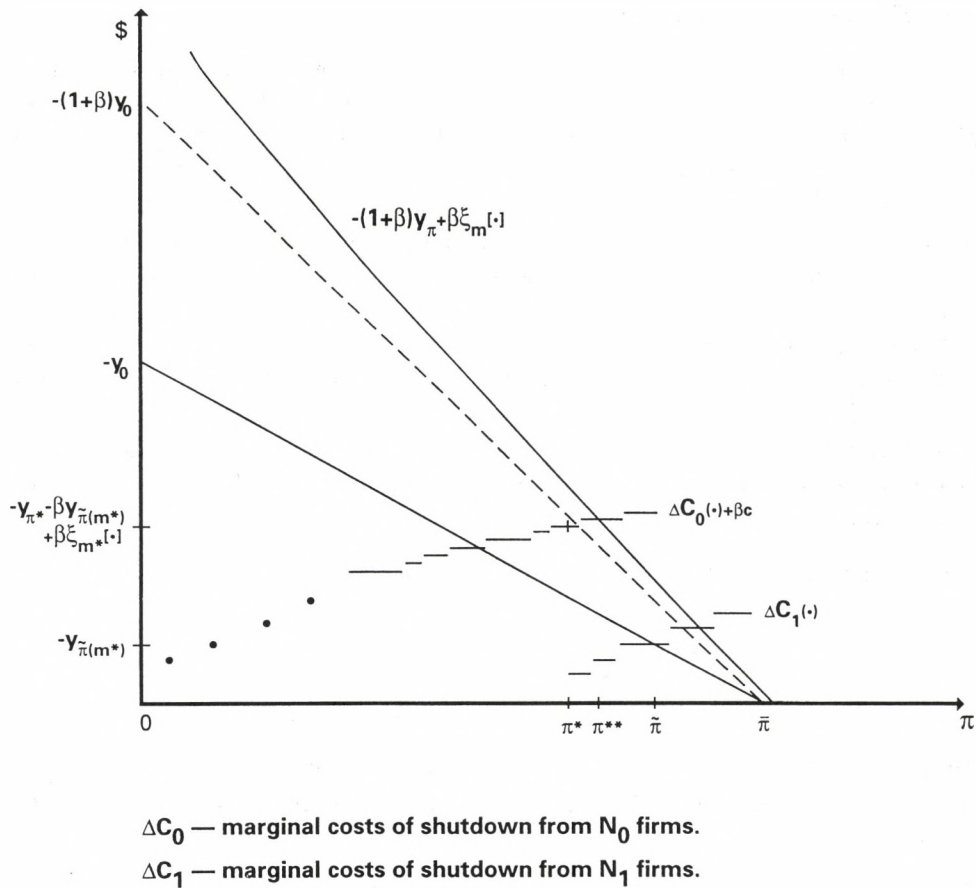


Figure 10.3
Optimal shutdown thresholds in the 2-period case

sition. For example, it shows that the marginal shutdown likelihood, π^* , is clearly greater when a future period is considered, that in its absence (compare with proposition 10.9 and figure 10.1). We now turn to further exploration of such comparisons in the next section.

4 Some Comparative Statics/Dynamics

Here we explore the impact of varying information and the planning horizon on the nature of the optimal solution. The results are largely corollaries of the propositions above, so we avoid full proofs.

4.1 The Value of Information

How does better information affect the optimal shutdown policy and the “option value” of delaying the shutdown decision? Intuitively, it is clear that the optimal policy must become more aggressive with better information as mistakes are less likely to be made and there is less to learn from waiting. The extreme case of this situation is that of perfect information about the quality of all existing firms, although actual outcomes in any period remain uncertain and firms yet to enter are still of uncertain quality $q > \bar{\pi}$. Then the initial N_0 firms can be divided into two sets: N_0^+ “good” firms with expected net return $y_1 > 0$, and N_0^- “bad” firms with expected net return $y_0 < 0$. In the terminal period there will be $N_1 = N_0 - m^* + \hat{\ell}$ operating firms, where m^* is the shutdown number and $\hat{\ell}$ is the number of new entrants from the initial resource pool L_0 , and $L_1 = L_0 + m^* - \hat{\ell}$ potential firms whose resources remain fallow. Clearly the problem faced by the planner is identical to that in equation (10.13) with y_i , $i = 0, 1$, in place of $\hat{y}_k$, for “bad” and “good” firms respectively.

$$\begin{aligned} \max_{k \geq 0} W(k; N_1, L_1, \hat{\ell}) = & (N_0^+ + \hat{\ell}^+)y_1 + (N_0^- - m^* + \hat{\ell}^- - k)y_0 \\ & + [y_q + r] \cdot \sum_{h=1}^{L_1} v_h^{L_1} \cdot h - C(k) - cL_1. \end{aligned} \quad (10.27)$$

The solution is a straightforward corollary of proposition 10.2:

PROPOSITION 10.7. Let $\hat{\ell} = (\hat{\ell}^+, \hat{\ell}^-)$ be the realization of new entrants at the beginning of the terminal period. Then there exists a unique optimal solution, $(\tilde{k}, \tilde{\pi}, \tilde{\sigma})$, to the problem (10.26) such that

$$\tilde{\pi} = 0 \quad (10.28)$$

$$-y_0 = (1 - \tilde{\sigma})\Delta C(\tilde{k}) + \tilde{\sigma}\Delta C(\tilde{k} + 1)$$

$$\tilde{\sigma} \in [0, 1] \quad \text{if} \quad \tilde{k} < (N_0^- - m^* + \hat{\ell}^-).$$

If, however,

$$-y_0 \geq \Delta C(N_0^- - m^* + \hat{\ell}^-)$$

then

$$\tilde{\pi} = 0, \quad -y_0 = \Delta C(\tilde{k}), \quad \tilde{k} = (N_0^- - m^* + \hat{\ell}^-), \quad \text{and} \quad \tilde{\sigma} = 0, \quad (10.29)$$

that is, it is never optimal to close a “good” $[y_1]$ firm.

This solution is illustrated in figure 10.4. It again gives a well-defined value function for the terminal period problem:

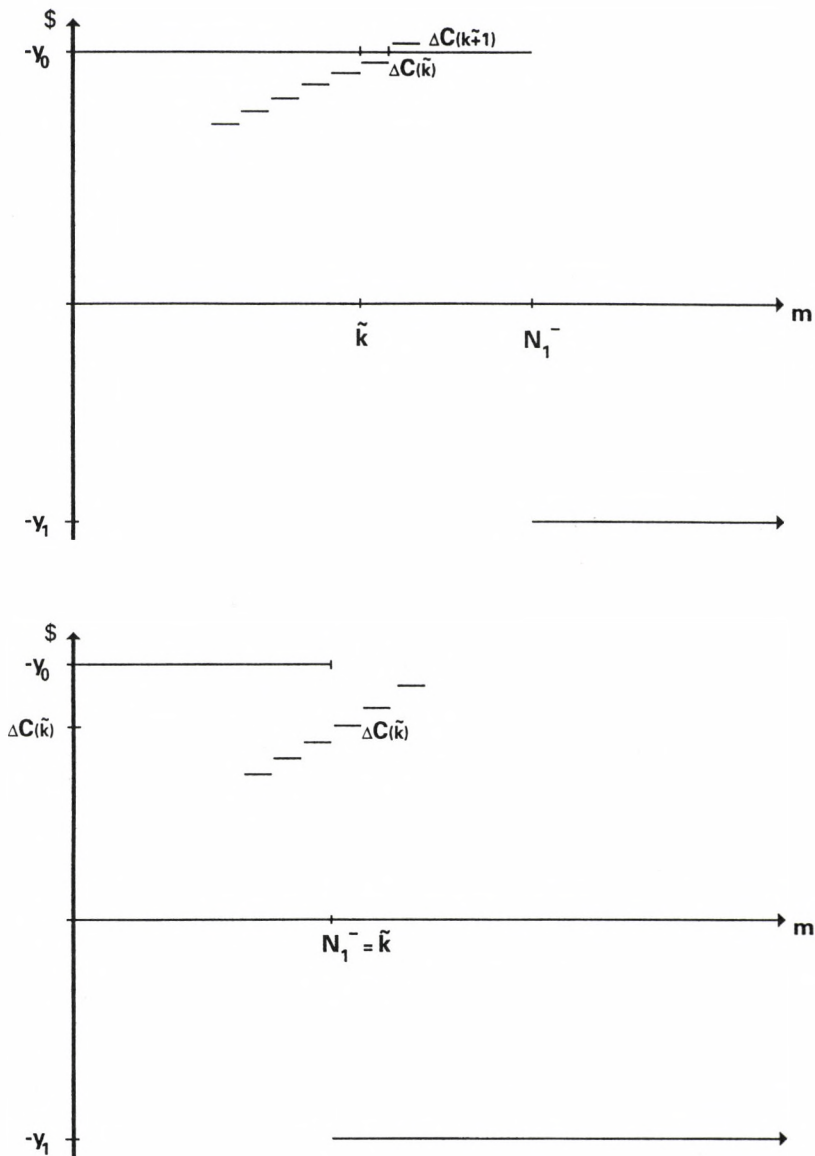


Figure 10.4
Optimal shutdown with full information: terminal period

$$V^{0f}(\hat{\pi}^f, N_1, L_1) = N_1^+ \cdot y_1 + (N_1^- - \tilde{k})y_0 + [y_q + r] \cdot \sum_{j=1}^{L_1} v_h^{L_1} \cdot j - C(\tilde{k}) - cL_1 \quad (10.30)$$

where superscript f indicates “full information” [$\hat{\pi}_i^f \in \{0, 1\}$, $\forall_i$], $N_1^+ = N_0^+ + \hat{\ell}^+$, $N_1^- = N_0^- - m^* + \hat{\ell}^-$, and $L_1 = L_0 + m^* - \hat{\ell}^+ - \hat{\ell}^-$. This value, and its variation in m^* , is a key input into solving the full two-period problem.

In the terminal period all remaining “bad” firms are closed unless the immediate incremental cost of doing so exceeds the maximum expected loss from the continued operation of such a firm. Thus only in the case of extraordinarily large incremental (social) costs of shutdown [$-y_0 < \Delta C(N_1^-)$] will known “bad” firms be left in operation. It is also immediate that enterprise shutdowns will go deeper into the industry, closing some firms left in operation in the first period, unless there were none left and/or $\hat{\ell}^- > m^*$. Thus the optimal policy with full knowledge of the quality of operating firms is substantially more aggressive than without such knowledge; there is no possibility of making a mistake. That is even clearer in the initial period, where any incentive to spread costs is dominated by the sure gains from eliminating value-destroying operations and the absence of any benefit to waiting.

As in the preceding section, the initial period’s Bellman equation is

$$V^{1f}(\pi^f, N_0, L_0) = \max_{0 \leq m^* \leq N_0} \left\{ N_0^+ y_1 + (N_0^- - m^*) y_0 - C(m^*) - cL_0 \right. \\ \left. + \sum_{j=1}^{L_0} v_j^{L_0} (j \cdot [y_q + r] + \beta E)[V^{0f}(\hat{\pi}^f, N_0 - m^* + j, L_0 + m^* - j) | \mathcal{F}_0] \right\}, \quad (10.31)$$

because it never pays to shut down a “good” firm. As above, the solution can be easily characterized:

PROPOSITION 10.8. Necessary and sufficient conditions for an optimal solution $\{(m^f, \pi^f, \sigma^f), (\tilde{k}^f(m^f), \tilde{\sigma}^f(m^f))\}$ to the two-period full-information problem (10.31) are given in proposition 10.7 for the terminal period, and for the initial period by:

$$\Delta C(m^f) + \beta c \leq -(y_0 + \beta[v_0 - \xi_{m^f}(y_q + r)]) < \Delta C(m^f + 1) + \beta c, \quad (10.32)$$

where

$$\xi_m = \sum_{j=0}^{L_0} \{v_j^{L_0} (L_0 + m - j) v_{L_0+m-j}^{L_0+m-j}\},$$

$$\text{and } v_0 = \max\{y_0, -\Delta C(\tilde{k})\},$$

implying that

$$\pi^f = 0,$$

$$-(y_0 + \beta[v_0 - \xi_{m^f}(y_q + r)]) = (1 - \sigma^f)\Delta C(m^f) + \sigma^f \Delta C(m^f + 1) + \beta c,$$

and $\sigma^f \in [0, 1)$, if $m^f < N_0^-$.

If, however,

$$-(y_0 + \beta[v_0 - \xi_{N_0^-}(y_q + r)]) \geq \Delta C(N_0^-) + \beta c,$$

then

$$m^f = N_0^- \quad \text{and} \quad \sigma^f = 0.$$

Again we can see that full information leads to optimal maximal shutdown of “bad” firms as illustrated (heuristically, ignoring integer constraints) in figure 10.5. Indeed, it is clearly far more aggressive than in the single period case as $-v_0 \gg c$,

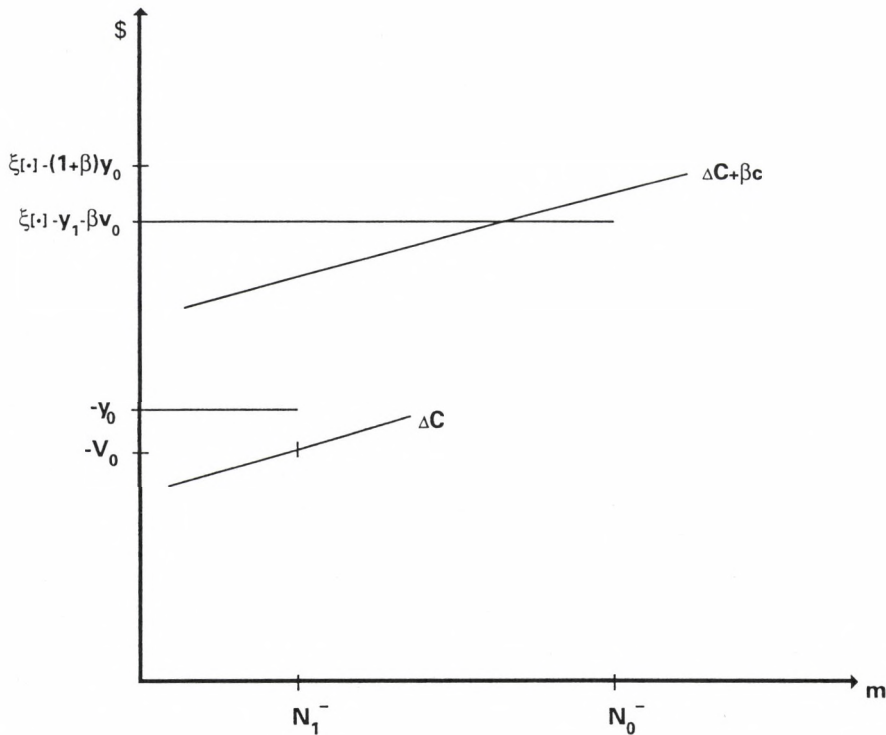


Figure 10.5
Optimal shutdown with full information: 2-period illustration

even without the possibility of new high-quality entry whose marginal value is reflected in $\xi \cdot (y_q + r)$.

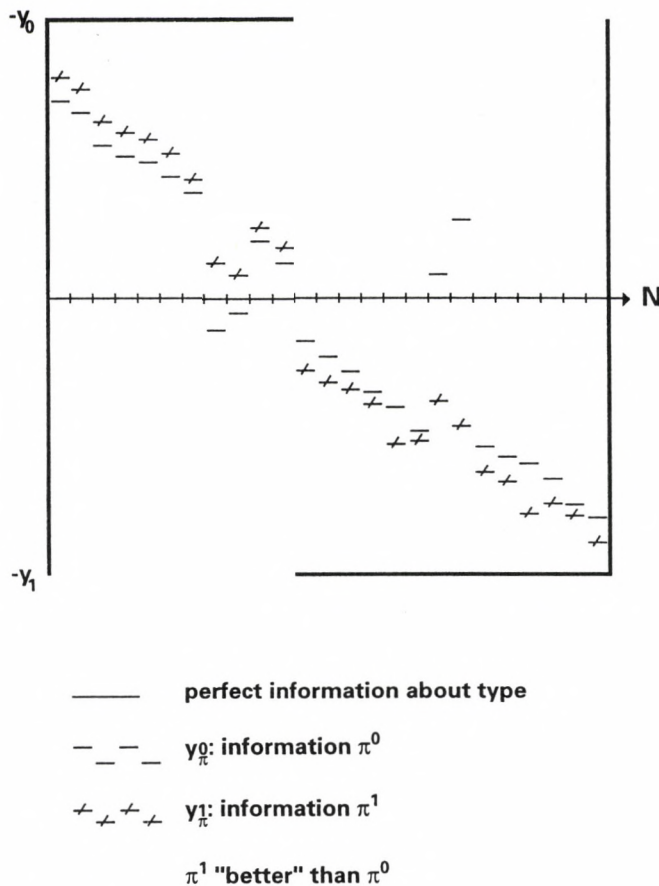
These results give us a possible measure of the value of information to the industry policy maker. Clearly the fact that $m^f > m^*$ and $\tilde{k}^f \geq \tilde{k}$ whenever there is the same number of "bad" firms left in the industry, shows that the information reduces the losses from the operation of "bad" firms and avoids the loss of net value from "good" firms, thereby raising the value of the industry to the planner. By virtue of optimization, the marginal costs of the additional shutdowns, $\Delta C(\cdot) + \beta c$, are no greater than the marginal benefits, $-y_0 - \beta v_0 + \xi \cdot (y_q + r)$, and this is reflected in the increase in the optimal value of the program. This is clear given full information about the quality of firms. From the perspective of a decision maker with only limited information, π , about quality, however, that may not appear to be the case, as some "bad" firms may have high π , thus decreasing $V^1(\cdot)$ (the uninformed apparent value) when shut down, whereas some "good" firms may have low π and so reduce $V^1(\cdot)$ when left in operation. Thus the difference in values, $V^{1f} - V^1$, cannot provide a reliable measure of the value of (in this case, full) information to the industry planner. There is, however, a natural measure of this value of information in this difference *conditioned* on the *better* information¹⁶:

$$\begin{aligned}
 E\{V^{1f}(\pi^f, N_0, L_0) - V^1(\pi, N_0, L_0) | \pi^f\} = \\
 = (1 + \beta)(y_1 - y_0)m^{*+} + \sum_{j=m^*+1}^{m^f} (-y_0(1 - \beta) - \Delta C(j)) \\
 + \beta \left\{ (y_1 - y_0)\tilde{k}^+ + \sum_{k=\tilde{k}+1}^{\tilde{k}^f} (-y_0 - \Delta C(k)) \right\} \\
 > 0,
 \end{aligned} \tag{10.33}$$

where superscript + indicates the number of "good" firms shut under the uninformed policy, whenever N_0 and new entry are the same in both situations. Notice that the increase in costs due to extra shutdown is always dominated by the savings from eliminating a surely "bad" firm, while keeping a good firm is a solid gain from both the cost and benefit side.

We expect this increase value to hold for less extreme improvements in information; the better the information about the underlying quality of the firm, the greater the value of the problem to the planner. To pursue this, one needs a precise definition of "better" information. Here I propose, without exploring its full implications, one such definition.

DEFINITION 10.9. For a given set of firms $N = N^+ \cup N^-$, where N^+ contains firms with expected value y_1 and N^- contains those of expected value y_0 , we say that a likeli-

**Figure 10.6**

An example of "improved" information

hood vector π' is *more informative* than π if the vector inequalities, $\pi'^+ \geq \pi^+$ and $\pi'^- \leq \pi^-$, hold with at least one strict inequality, where the superscripts refer to the corresponding subsets of firms in N .

The definition is illustrated in figure 10.6. Note that this definition is consistent with an improvement in information under Bayesian updating of the likelihoods, π ; as the argument in lemma 10.3 indicates, $\hat{\pi}$ is *increasing to 1* conditional on the firm being "good," while it is *decreasing to 0* when the firm is "bad." It is immediately evident that there will be greater likelihood of shutdown of firms likely to be "bad" and a lesser likelihood of closing firms expected to be "good." This is reflected in an increase in the expected savings from the m^* firms to be shut down and an

increase in the expected value created by those firms remaining in operation. It is always the case, conditional on the *better* information, that the increase in expected savings will compensate for the marginal cost of additional shutdown, thereby further increasing the value of the program $V^1(\pi, N, L)$. This reflects the fact that the expected cost of each of the two types of errors balanced in the optimal policy is reduced by more information. It is just a matter of straightforward calculations, similar to those above, to show:

PROPOSITION 10.10. For given sets, N, L , of active and latent firms, if the likelihood distribution π' is more informative than π , then $V^1(\pi', N, L) > V^1(\pi, N, L)$, conditional on the information in π' .

Thus we can clearly see the value of information in this model. Further, we see that improved information can make the optimal policy no less aggressive (in terms of likelihood cutoff for shutdown), so that the Bayesian learning that takes place with the passage of time will tend to stimulate further shutdowns, even if no further "bad" firms arise. Only the *elimination* of shutdown targets, i.e. those firms with $-y_{\pi} > C(1)$, will lead an optimal restructuring policy to avoid all further shutdowns.

4.2 Planning Horizon and Optimal Timing

In Proposition 10.6 we saw that the optimal cutoff likelihood for shutting down weak firms was increasing in the length of the horizon ($T = 1 > 0$). It is also the case that the marginal shutdown likelihood increases when there are two periods, but no shutdown is allowed in the second period. In that case, the marginal benefit to shutdown increases by a factor of $(1 + \beta)$ as expected savings from closing value-destroying firms now accrue over two periods, whereas the incremental carrying costs of unemployed resources are only $\beta c < \beta C(1) < -\beta y_{\pi^*}$. Hence a greater marginal cost, $\Delta C(\cdot)$, is justified, allowing more firms to be optimally shut down as can be seen by comparing figure 10.1 and figure 10.3.

But a more interesting comparison is, naturally, between the two-period situations with and without the possibility of second period shutdown. The latter might arise if, for example, a change in the political situation were to prevent any further shutdown in the second period. Both cases take into account the fact that current shutdown of failing firms implies savings both now and in the future, but only the former case does it allow the optimal spreading of (convex) costs over both periods. Note that this implies that a single round of closures (if that were all that was politically, say, feasible) should never be put off until the later period; bearing the costs up front allows for two periods of savings, always exceeding the extra maintenance costs of unemployed resources $[C(1) \geq c]$, as well as for potential entry of new productive

firms. Thus we might expect that the policy should be more conservative when there is an option for future shutdown. That this is indeed the case is easy to see, although optimal cumulative shutdown across two periods is naturally greater than the total that is optimal when only one round of shutdown of failing firms is possible.

If there is a future in which no further action can be taken, then the problem becomes

$$\max_{0 \leq m^* \leq N_0} \left\{ \sum_{i=m^*+1}^{N_0} y_i - C(m^*) - cL_0 + \sum_{j=0}^{L_0} v_j^{L_0} \cdot j \cdot [y_q + r] \right. \\ \left. + \beta \cdot \sum_{j=0}^{L_0} v_j^{L_0} E \left[\sum_{i=m^*+1}^{N_0} \hat{y}_i + \sum_{k=1}^j \hat{y}_{qk} + [y_q + r] \left(\sum_{h=1}^{L_1^j} h \cdot v_h^{L_1^j} \right) - L_1^j \cdot c \left| \mathcal{F}_0 \right. \right] \right\} \quad (10.34)$$

where $L_1^j = L_0 + m^* - j$. Again let ΔW_m be the marginal gain from closing the m th firm,

$$\Delta W_m = -(1 + \beta)y_m + \xi_m[y_q + r] - \Delta C(m) - \beta c,$$

where ξ_{m^*} is defined in Proposition 10.6.¹⁷ Then, at the optimum, as in all the prior results (e.g., equation (10.25)),

$$-(1 + \beta)y_{m^{**}} + \beta \xi_{m^{**}}[y_q + r] \geq \Delta C(m^{**}) + \beta c \quad (10.35)$$

$$-(1 + \beta)y_{m^{**}+1} + \beta \xi_{m^{**}+1}[y_q + r] < \Delta C(m^{**} + 1) + \beta c$$

Comparing with equations (10.21) shows that the optimal first period cutoff, π^{**} , must be greater than π^* when there is no second period opportunity to close firms: $-y_{\tilde{\pi}(m^*)} < -y_{m^*}$ as $E[\tilde{\pi}(m^*)] > \pi^*$ (10.22), so it is optimal to bear only a lesser cost of shutdown in the initial period. Further, π^{**} can be no greater than $\tilde{\pi}(m^*)$ without violating the first order conditions. Thus, as again illustrated in figure 10.3, the possibility of cost spreading gives

PROPOSITION 10.11. The optimal first-period shutdown margin, π^* , is lower than π^{**} , and hence fewer firms are shutdown in the first period, when further shutdown is possible in the second period, than is the case when shutdown decisions can only be made in the first period.

COROLLARY 10.12. $m^* + \tilde{k}(m^*) > m^{**}$, that is, more total shutdown optimally occurs when it can take place in both periods.

With two periods in which to make shutdown decisions the optimal policy is hence less aggressive in the first period, while eventually cutting deeper into the pool of firms likely to be “bad.” This result provides one sense in which “shock therapy”

could be considered optimal when there is a narrow political "window of opportunity" as has been argued in, among others, Aslund (1995).

Another comparison of interest deals with the optimal policy trajectory when there will be perfect information in the final, but significant uncertainty in the initial, decision period. The two-period problem is formally still given by (10.19), with V^0 replaced by V^{0f} and N_1^+ and N_1^- depend on the prior shutdown decision, $m^* = m^+ + m^-$, including *mistaken* shutdowns of "good" firms: m^+ .¹⁸ The key to understanding how that dramatic improvement in information affects the first period optimal decision lies in the variational analysis, similar to that in lemmata (10.12) and (10.13), of $V^{0f}(\cdot)$, as the initial period shutdown level, m , changes. It is straightforward to see:

LEMMA 10.13. Let $\tilde{k}_\ell(m)$ be the optimal shutdown in the second (full information) period when ℓ^- "bad" firms enter, and $\Delta V_\ell^{0f}(m)$ be the expected change in value for the entry of $\ell = \ell^+ + \ell^-$ new firms, when m firms are shutdown in the first period. Then

1. $\tilde{k}_\ell(m) = \tilde{k}_\ell(m-1)$ if $\Delta C(N_0^- + \ell^- - m + 1) > -y_0$;
If $\Delta C(N_0^- + \ell^- - m + 1) \leq -y_0$, then $\tilde{k}_\ell(m) = \tilde{k}_\ell(m-1) + 1$;
2. $\Delta V_\ell^{0f}(m) = \Delta V_\ell^0(m)$ if $\Delta C(N_0^- + \ell^- - m + 1) > -y_0$;
If $\Delta C(N_0^- + \ell^- - m + 1) \leq -y_0$, then

$$\begin{aligned} \Delta V_\ell^{0f}(m) &= -\tilde{\pi}(m)y_1 + (1 - \tilde{\pi}(m))\Delta C(\tilde{k}_\ell(m) - (m)) + \Delta V_\ell^e(m) \\ &< -y_{\tilde{\pi}(m)} + \Delta V_\ell^e(m) = \Delta V_\ell^0(m) \end{aligned}$$

[see lemma 10.14].

Proof. This follows almost immediately from proposition 10.7, letting $\tilde{\pi}(m)$ be the cutoff likelihood when m firms are closed in the first period. If the marginal cost of shutting all remaining bad firms is less than $-y_0$, it will be done. If an additional firm is then shut earlier, the cost of doing so later if it is "bad", $\Delta C(\tilde{k}_\ell(m-1))$, is saved, as $\tilde{k}_\ell(m) = \tilde{k}_\ell(m-1) - 1$. If, however, it was "good," then there is an expected loss in value of output, $-y_1$, so the impact on program value (*ex ante*) is determined by the prior probability that the firm is "good" (therefore not requiring shutdown). If, on the other hand, it is too costly to close all "bad" firms in the second period, then that will remain the case with the additional shutdown, $\tilde{k}_\ell(m) = \tilde{k}_\ell(m-1)$, and the impact on (*ex ante*) expected value of the program is the same as with continuing imperfect information: with probability $\tilde{\pi}(m)$, y_1 is lost, and with probability $(1 - \tilde{\pi}(m))$, $-y_0$ is gained. Finally, the impact on the expected value of new entry is unaffected by the improved information in the second period. ■

This result allows us to give some conditions such that full information in the final period leads to greater caution earlier. They depend critically on how many “bad” firms there turn out to be, after shutdowns and new entry have taken place. Of course, in the first period the industry decision maker can never be sure how many “bad” firms there are, and so must probability-weight the two payoff relevant events analyzed in lemma 10.13; his expectation of future value will depend not only on his beliefs about the quality of existing firms but on the distribution over the entry of “bad” firms. To state our result, let $A(m)$ be the event that $\{N_1^-(m, \ell^-) > C^{-1}(-y_0)\}$, $N_1^-(m, \ell^-) = N_0^- + \ell^- - m + 1$, and $B(m)$ be its complement. The Bellman equation, as before, is

$$V^1(\pi, N_0, L_0) = \max_{0 \leq m \leq N_0} \left\{ \sum_{i=m+1}^{N_0} y_i - C(m) - cL_0 + \sum_{j=1}^{L_0} v_j^{L_0} \cdot j \cdot [y_q + r] \right. \\ \left. + \beta \sum_{j=1}^{L_0} v_j^{L_0} (E[V^{0f}(\hat{\pi}^f, N_0 - m + j, L_0 + m - j) | \mathcal{F}_0, j]) \right\},$$

where terminal value is given by equation (10.29), and its expectation is taken with respect to the distribution of “good” $[N_1^+]$ and “bad” $[N_1^-]$ firms.

THEOREM 10.14. An optimal solution $\{(m^0, \pi^0, \sigma^0), (\tilde{k}^0(m^0), \tilde{\sigma}^0(m^0))\}$ to the two period problem with second-period full information satisfies conditions (10.28) on event $A(m^0)$ and (10.29) on $B(m^0)$ for the terminal period. For the initial period, the necessary and sufficient conditions are given by:

$$\pi_{m^0} \leq \pi^0 \leq \pi_{m^0+1}, \quad \sigma^0 = 0,$$

$$-(y_{\pi^0} + \beta[y_{\pi^0} - \xi_{m^0}(y_q + r)]) = \Delta C(m^0) + \beta c,$$

or

$$\pi^0 = \pi_{m^0+1}, \quad \sigma^0 > 0,$$

$$-(y_{\pi^0} + \beta[y_{\pi^0} - \xi_{m^0}(y_q + r)]) = (1 - \sigma^0)\Delta C(m^0) + \sigma^0\Delta C(m^0 + 1) + \beta c,$$

on event $A(m^0)$, i.e. when known “bad” firms are optimally left in operation in the second period. On event $B(m^0)$,

$$\pi_{m^0} \leq \pi^0 \leq \pi_{m^0+1}, \quad \sigma^0 = 0,$$

$$-(y_{\pi^0} + \beta[\pi^0 y_1 - (1 - \pi^0)\Delta C(N_1^-) - \xi_{m^0}(y_q + r)]) = \Delta C(m^0) + \beta c,$$

or

$$\begin{aligned}\pi^0 &= \pi_{m^0+1}, \quad \sigma^0 > 0, \\ -(y_{\pi^0} + \beta[\pi^0 y_1 - (1 - \pi^0)\Delta C(N_1^-) - \xi_{m^0}(y_q + r)]) \\ &= (1 - \sigma^0)\Delta C(m^0) + \sigma^0\Delta C(m^0 + 1) + \beta c,\end{aligned}$$

give the appropriate marginal conditions. Further, $(\pi y_1 - (1 - \pi)\Delta C(N_1^-)) > y_{\pi^0} > v_0 \geq y_0$.

Remark 10. When some “bad” firms must optimally remain in operation in the second period [event $A(m^0)$], the marginal conditions are the same as those when the shutdown decision can only be made in the first period [see (10.35)]. However, if there is some slack [event $B(m^0)$], or even a positive probability of such slack, then the policy with full information in the future period becomes more conservative, $m^0 \leq m^{**}$ and $\pi^0 < \pi^{**}$.

Note that the uncertainty in the first period reduces the *expected* marginal return in the second period, despite full information at that time, because of the marginal likelihood, π^0 , of shutting down a good firm by mistake. Thus this policy is more cautious than the full information first period optimal policy: $m^0 < m^f$. It will usually, however, not be more cautious in the first period than the optimal policy in the uninformed case (theorem 10.6), which, as we have seen, is more conservative than shutdown policy when there is not a second round. Whether it is so, depends on the relationship between $y_{\pi^0}(m^*)$ and $(\pi^0 y_1 - (1 - \pi^0)\Delta C(N_1^-))$ and on the probability of $B(m^0)$. If slack in the second period, implying low marginal costs of second period shutdown, is sufficiently likely, that is, there are few “bad” firms and a low probability of entry of such, then it will be optimal to hold back in the first period, setting $\pi^0 < \pi^*$, thereby spreading some of those costs across the two periods. In event $A(m^0)$, however, $m^0 = m^{**} > m^*$, so better information in the future encourages more aggressive initial restructuring.

5 Conclusions

The analysis of this chapter deals with the simplest formulation of a dynamic industry restructuring problem. It makes a number of quite strong assumptions in order to get an analytic handle on the complexities of enterprise shutdowns during the transition process. Among the assumptions that one would like to relax are those of the homogeneity of firms (the “standard firm” assumption), the direct connection between the market generated signal of “viability” and the social value of the firm, the lack of interdependence among firms, the static environment, and the

inability of firms to alter their own viability through investment. These, as well as extending the analysis to a multiperiod frame work, provide directions for continuing research.

Despite these limitations, the model provides a basis for useful insights into some of the difficulties of the transition. It captures, in tractable ways, critical elements of the process—the randomness in current performance, the uncertainty about underlying viability, the learning from performance outcomes, the convex costs of closing industrial operations, the costs of keeping released resources usable in the future, and finally, the critical connection between shutdowns and new entry.

The results of the model highlight the necessary conservatism of the optimal policy: it balances the marginal costs of errors arising from the lack of information, leaving many firms that are likely to be net value-destroying in operation. This, in part, is a reflection of the option value of “waiting”; the industry policy maker will know better in the future whether the firm is “bad” or not. Despite this, however, the optimal policy is more aggressive in the early period, reflecting two factors. First, the expected savings to any closure in the first period accrue over the full horizon, and hence are worth almost twice savings in the second period. Second, shutdown in the first period allows for greater entry of valuable new firms in the second period. Both factors work to limit the desire of an optimizing industry authority to spread the convex costs of necessary shutdowns over both periods, and indeed counteract, to some extent, the option value of waiting.

The second factor, highlighted in this model, has been, I believe, underappreciated in the transition literature. The release of resources from wasteful industrial operations acts as a countervailing option to that of waiting. By facilitating new startups with an expected viability much greater than that of the closed enterprises, the release of resources from shutdowns mitigates the hysteresis induced by the value of waiting. It must be emphasized that this is an effect that would be absent if the shutdown decision were being taken in a well functioning market economy where resources for new startups are available at their opportunity cost. In transition economies, and in particular the former Soviet Union, that is far from the case; resources, even if unemployed or seriously underemployed, are largely frozen in place by prior socialist overemployment policies, the lack of working factor markets (missing infrastructure and intermediaries), the lack of usable property rights, and transition policies that support existing industrial operations, strengthening their grip on existing resources and factors of production. One of the key messages of this chapter is that shutting down existing industrial operations, by refusing to support them when the expectation that they are market-viable is sufficiently low,

has an additional positive impact through new entry that must be considered in and transition industrial policy.

Thus the optimal shutdown threshold, and number of optimally closed firms, is a complex function of many such factors that balances the countervailing forces introduced in the model. Overall, it supports something of a "big bang" approach in the sense that earlier shutdown generally dominates, even when information is expected to improve dramatically in the future, and the number of firms optimally shut down declines rapidly over time. Convexity of costs and the possibility of mistakenly closing viable firms, however, limit the size of that initial "bang," as well as the number of future shutdowns, unless there is a dramatic improvement in information. Better information about the underlying quality of firms generally leads to a more aggressive first-period shutdown policy, as does a longer horizon, and full information maximizes the amount of shutdown that it is optimal to undertake. These results provide the first steps toward developing a general model of industrial transition, one which must include interdependencies among the firms and the ability of firms to make autonomous decisions about their own future, as well as the (non-)supporting role of central policy-makers.

Notes

I have benefited from the comments and discussion of seminar participants at Duke University, Pennsylvania State University, The William Davidson Institute of the University of Michigan Business School, and CEMI of the Russia Academy of Sciences. I bear responsibility for remaining errors and omissions.

1. See, for example, the survey of economic transitions in Gelb et al. (1996), McMillan (1995), the Blanchard, Froot, and Sachs (1994) collection of papers on the transition in Eastern Europe through 1993, and the more recent comparative study of Aslund, Boone, and Johnson (1996).

2. As Peter Murrell (1996, p. 39) notes, "There is universal agreement that new private firms have been of great significance." For example, how and why they are important is explored in some detail in Johnson and Loveman (1995) for the Polish case and in Richter and Schaffer (1996) and Earle and Rose (1996) for the Russian case.

3. One exception is the unpublished paper of Castenhiera and Roland (1995) which provides a normative model of the optimal rate of restructuring when investment in new capital, reducing current consumption, is required to reemploy labor released from closed firms. Another partial exception is the simple model of "disorganization" in Blanchard and Kremer (1997) and Blanchard (1997), where asymmetric information creates a coordination problem causing input disruptions and an excessive loss of output in the state sector.

4. As I have argued elsewhere (Ericson 1991), economic activity and organizations in the Soviet-type system were built, and functioned, without any consideration of real economic (opportunity) costs or benefits, other than those of the narrowly pursued priorities of the central authorities at the time of their construction. Thus there is no guarantee of any structure's viability, never mind profitability, in a market environment, and much of that structure is, indeed, net value destroying (see Ericson 1996).

5. See Davis, Haltiwanger, and Schuh (1996) for a seminal study of the dramatic, continual change in the U.S. economy.

6. Of course, alternative mechanisms have arisen in the private (shadow economy) sector, providing Gerschenkronian substitutes for legal guarantees and enforcement of property and contractual rights. These substitutes are, of course, highly imperfect, carrying very high attendant "transactions costs" as discussed in Leitzel (1995), chapter 2.

7. This can be seen in the policy prescriptions of Sergei Glaziev, the new chief of the Security Council's Economic Security Administration, interviewed in *Trud*, 26 September 1996, or the group of "Academics" and their American advisors during the recent Russian presidential campaign. See, for example, the latter's manifesto "Novaia ekonomicheskaja politika dlja Rossii," *Nezavisimaia gazeta*, 1 July 1996, pp. 1, 4.

8. By "net value destroying" I mean that the market value of output produced is insufficient to cover the full costs of production, including the opportunity cost of the factors of land, labor, and capital.

9. These conditions are satisfied by, among other distributions, a truncated Poisson distribution,

$p(x) = \frac{\lambda^x}{x!} e^{-\lambda}$, with $\lambda < 2$ and the mass of the tail placed on $x = 0$. Note that the tail conditions in the second line imply that the number of entrants is concave in L —the pressure for any bundle of resources to be reused is diminishing in the volume of unused resources.

10. Note that N_0 is somewhat redundant as it is contained in the dimension of the vector of prior qualities, π . Indeed, in a full dynamic version of the model we set $\pi_i = 0$ for all $i \in L$, so that the vector $\pi \in [0, 1]^N$ contains all relevant information on the state of the industry.

11. For clarity we have assumed that the total resources available to the industry are fixed. It would change little to assume them to be growing or shrinking for exogenous (to the policy of the sector) reasons. Indeed, that will be the case for most industries in a transition economy undergoing structural change.

12. Thus reflects an assumption of initial socialist "overfull employment"; There are initially *no* resources that are *not* fully committed to, and/or controlled by, existing enterprises.

13. With only slight abuse of notation, let N_i , L_i stand for the number of elements in those sets of firms, as well as the sets themselves.

14. The latter would only be interesting as an approximation if the decision and firm response periods were very short relative to the expected length of the "transition."

15. Clearly ξ_m is quite small inasmuch as the probability of the coincidence of a number of unlikely (marginal) events [See equation (10.6)]. These events become even less likely as L_0 and/or m increase. For example, if v_i is truncated Poisson, $L_0 = 3$, and $m = 2$, then $\xi_m \approx .004$; $L_0 = 0$, $m = 2 \Rightarrow \xi_m \approx .068$; $L_0 = 0$, $m = 4 \Rightarrow \xi_m \approx .0008$.

16. Notice that, given identical initial conditions and realizations of η_i , the carrying costs of latent firms and the impact of new entry is identical in both situations. Hence, conditioning on the true quality of operating firms, we need only consider the direct impact of differing shutdown decisions.

17. ξ_{m^*} is the same here, modulus the differing m^* , as it is a function only of the initial conditions and first period decision.

18. This is a critical difference from the two-period full information case where there is never a mistake of shutting down "good" firms, so $m^* = m$ only.

References

- Aghion, Philippe, and Olivier Blanchard (1994). On the speed of transition in Central Europe. In S. Fischer, and J. J. Rotemberg, eds., *NBER Macroeconomics Annual 1994*, pp. 283–320. Cambridge, Mass.: MIT Press.
- Aslund, Anders (1995). *How Russia Became a Market Economy*. Washington, D.C.: Brookings Institute Press, 1995.

- Aslund, Anders, Peter Boone, and Simon Johnson (1996). How to stabilize: lessons from post-communist countries. *Brookings Papers on Economic Activity* 1: 217–313.
- Atkeson, Andrew, and Patrick J. Kehoe (1995). Industry Evolution and Transition: Measuring Investment in Organization Capital. Staff Report 201, FRB Minneapolis.
- Blanchard, Olivier (1997). *The Economics of Post-Communist Transition*. Oxford: Clarendon Press.
- Blanchard, Olivier, and Michael Kremer (1997). Disorganization. *The Quarterly Journal of Economics* 112(4): 1091–1126.
- Castenhiera, Michael, and Gerard Roland (1995). The Optimal Speed of Transition: A General Equilibrium Analysis. Manuscript, IGIER, Milan.
- Davis, Steven J., John C. Haltiwanger, and Scott Schuh (1996). *Job Creation and Destruction*. Cambridge, Mass.: MIT Press.
- Earle, John S., and Richard Rose (1996). Ownership Transformation, Economic Behavior, and Political Attitudes in Russia. Working paper, Stamford CISAC.
- Ericson, Richard E. (1991). The classical Soviet-type economy: nature of the system and implications for reform. *Journal of Economic Perspectives* 5(4): 1–18.
- Ericson, Richard E. (1994). Cost trade-offs in activity shutdowns: a Note on economic restructuring during the transition. In R. Campbell, ed., *The Postcommunist Economic Transformation*, 195–217. Boulder, Colo.: Westview Press.
- Ericson, Richard E. (1996). The Structural Barrier to Transition: A Note on Input-Output Tables of Centrally Planned Economies. Columbia University Economics Department Working Paper #9596-31.
- Gelb, Alan et al. (1996). *From Plan to Market: World Development Report 1996*. New York: Oxford University Press.
- Johnson, Simon, and Gary W. Loveman (1995). *Starting Over in Eastern Europe: Entrepreneurship and Economic Renewal*. Cambridge, Mass.: Harvard Business School Press.
- Kornai, János (1994). Transformational recession: the main causes. *Journal of Comparative Economics* 19(1): 39–63.
- Leitzel, Jim (1995). *Russian Economic Reform*. London: Routledge.
- McMillan, John (1995). Markets in Transition. Paper presented at the 7th World Congress of the Econometric Society, Tokyo.
- Murrell, Peter (1996). How far has the transition progressed. *Journal of Economic Perspectives* 10(2): 25–44.
- Richter, Andrea, and Mark E. Schaffer (1996). The Performance of *De Novo* Private Firms in Russian Manufacturing. CERT Discussion Paper #96/10.
- Svejnar, Jan (1991). Microeconomic issues in the transition to a market economy. *Journal of Economic Perspectives* 5(4): 123–138.
- Svejnar, Jan (1996). Enterprises and workers in the transition: econometric evidence. *AER Papers and Proceedings* 86(2): 123–127.
- Van Frausum, Yves (1995). Industrial restructuring in Romania: diagnosis and strategies. *Europe-Asia Studies* 47(1): 47–65.

11 Shortage and the Soft Budget Constraint: New Interpretations

Paul Hare

1 Introduction

Shortage is a phenomenon that János Kornai considers to be endemic to the centrally planned economy of the Soviet type, and a great deal of his work has been devoted to describing and explaining shortage, the conditions under which it can arise, and the circumstances that sustain it (see, for instance, Kornai 1980). Some of his analysis seeks to be general, applying to all or most economies of Soviet-type (see Kornai 1992), while much of it also refers to or draws on his particular experience and knowledge of the Hungarian case, with its own unique peculiarities (on the latter, see Kornai 1986a, 1995).

For Kornai, one of the cornerstones of the theory of shortage was the notion of the *soft budget constraint* (see Kornai 1986b), to the extent that in many of his writings he came close to insisting that soft budget constraints necessarily accompanied, and indeed should be regarded as a precondition for, shortage. This line of argument raises some interesting and difficult questions concerning both the appropriate definitions of the concepts themselves, the proper theoretical framework within which to analyze them, the types of evidence that can be referred to, and hence the policy implications that can be elicited from the ensuing analysis. Nevertheless, the introduction of the idea of the soft budget constraint into economic thinking owes much to Kornai's pathbreaking analysis.

As far as the theories are concerned, there has been great controversy between the views of Kornai himself and economists more wedded to neoclassical approaches to microeconomic analysis. The former saw shortage as endemic to, and inherent in, the socialist economy as a result of certain key behavioral regularities of that system, which, under given conditions, gave rise to a so called *normal degree of shortage*. Though Kornai might hesitate to use the term, the normal degree of shortage was a feature of the system's *equilibrium* in the strong sense (involving a notion of local dynamic stability) that in the centrally planned economy, forces would be set up to restore normal shortage following any disturbance that gave rise to a departure from this state.¹ On this view of shortage, price reforms—whether they merely removed some of the more glaring distortions in the price system and brought prices more in line with costs, or whether they entailed more comprehensive price liberalization—were not considered likely to do away with shortage or even to change its normal degree.

In contrast, neoclassical economists tended to think of shortage as one of the categories of equilibrium that could occur in a fixed price model in which some desired

transactions could not be realized due to constraints on one or other side of given markets (see Barro and Grossman 1976 for a macroeconomic approach; for a more microeconomic analysis, see Benassy 1982). A shortage equilibrium is then one in which these constraints are widespread, and predominantly binding on the supply side: hence a shortage economy is essentially one that is characterized as having a supply-constrained equilibrium. Based on this line of analysis, there are two possible approaches that can be adopted to remove the shortage and enable the economy to move toward an unconstrained (Walrasian) equilibrium. According to the first, assuming generalized shortage conditions (i.e., generalized excess demand across the economy), the proper remedy would be to bring about a reduction in aggregate demand. The second approach is more applicable to mixed situations in which some markets experience persistent shortage while others simultaneously feature excess supply, and it involves price reform and/or liberalization in order to bring about changes in relative prices.

A great deal of empirical work, both on individual markets and on the economy as a whole, was based on this neoclassical conception of shortage, and some of it sought to discriminate between these two hypotheses of "generalized excess demand" and "wrong relative prices" (see Portes 1989 for a comprehensive survey). Though many of the findings could be interpreted in ways that seemed "reasonable," the results did not generally provide conclusive support for the neoclassical approach.

Since the 1989 collapse of communist governments in Central and Eastern Europe and the subsequent disintegration of the Soviet Union in 1991, central planning and its institutional framework have largely disappeared and most countries of the region have embarked upon extensive reform programs in order to bring about a transition toward a market-type economy (see EBRD 1996 and World Bank 1996, for the most thorough recent reviews of the transition process). One of the first measures taken by most countries was to liberalize both prices and foreign trade by removing most of the former controls and distortions. The result, in most cases, was the elimination of shortage almost overnight. The only markets where shortage remained important were those in which prices continued to be controlled (housing, energy), or where a history of underinvestment implied that no reasonable prices could clear the market (telecommunications, private cars, and associated services) in the short run. Hence, casual observation of the first phase of transition appeared to offer support for the neoclassical view that freeing up relative prices was the key to eliminating shortage.

Nevertheless, observers of the region quickly realised that these liberalizing measures did not automatically result in hard budget constraints for enterprises. Many

firms, especially but not only in the CIS, held onto surplus workers, undertook at most the mildest forms of defensive restructuring (see Grosfeld and Roland 1995; EBRD 1995), and enjoyed continuing subsidies, both direct (i.e., budgetary subventions to particular firms) and indirect (i.e., tolerance of tax and social security arrears, soft credits, spiraling interenterprise debts, and so on) (on the latter, in the Russian case, see Alfandari and Schaffer 1996; on Central Europe, see Newbery 1995). Hence this recent experience suggested that the previously postulated connections between shortage and the soft budget constraint were not completely correct, or only provided part of the story.

Accordingly, I proceed in this chapter to explore the issues raised by these observations in the following way. The second section reviews the basic definitions and concepts in an effort to clarify the relationships between them, and the third section considers the relevant empirical evidence in some detail.

2 Definitions and Concepts

In an economy producing output y (a vector of dimension n , positive components representing net outputs, negative components representing net inputs such as labor supply), enjoying the consumption vector x and with initial endowments $\bar{x}$, excess demands are defined as $z = x - y - \bar{x}$. Under competitive conditions (i.e., price-taking behavior by all agents), demands and supplies, hence excess demands z , depend on the prevailing price vector, p . Moreover, by making suitable assumptions about the underlying behavior of economic agents (firms and households) and the constraints they face (consumption and production possibility sets, budget constraints), the excess demand function² can easily be shown to possess the following properties:

(C) *Continuity* The excess demand function $z(p)$ is defined and continuous for all $p \geq 0$ ($p \neq 0$).

(H) *Homogeneity* The excess demand function is homogeneous of degree zero in p .

(W) *Walras's Law* The excess demand function satisfies Walras's Law, namely $p \cdot z(p) = 0$ for all $p \geq 0$.

A *Walrasian equilibrium* in such an economy is a price vector p^* with the property that $z(p^*) \leq 0$, that is, there is no excess demand in any market. It is a standard result that an economy with properties C, H, and W possesses at least one Walrasian equilibrium (see Arrow and Hahn 1973). Moreover, if we decentralize the allocation to

individual agents (labeling households by h and firms by f), we know that for household h , $x_h(p^*)$ maximizes $u(x_h)$ subject to $p^* \cdot x_h \leq m_h$ and $x_h \in X_h$; and for firm f , $y_f(p^*)$ maximizes $p^* \cdot y_f$ subject to $y_f \in Y_f$ (using standard notation³).

If prices are fixed rather than flexible, then unless by chance they are fixed at p^* (or at some other equilibrium price vector in the case of multiple equilibria) some agents will not be able to complete all their desired transactions at the prevailing prices. Such agents are said to be *constrained*. In such conditions, we can nevertheless define a constrained equilibrium that requires agents not only to take the prevailing prices as given, but also to take as given the constraints that they face (or, more strictly, perceive).

Suppose household h faces constraints of the form $x_h^l \leq x_h \leq x_h^u$ and firm f faces constraints of the form $y_f^l \leq y_f \leq y_f^u$. Then a *constrained equilibrium* for price vector p is a set of constraints for each agent (i.e., x_h^l and x_h^u for household h , and y_f^l and y_f^u for firm f) and transaction vectors for each agent (i.e., x_h^* and y_f^*) with the property that $z^* \leq 0$ and firms and households optimize subject to the constraints they face, including the new transactions constraints.

This rather general conception of a constrained equilibrium is not yet especially interesting inasmuch as it allows far too many equilibria to arise even in the simplest of models, and it fails to make proper use of a great deal of market-relevant information. Hence, it is better to refine the idea by specifying how agents use certain types of market information. Specifically, the definition already proposed allows situations where, in a given market, some agents might face upper limits on their transactions, some might face lower limits, and some might be unconstrained. There are two things wrong with such a formulation: (1) Although agents might face constraints that prevent them reaching a preferred position, they should not be forced to accept individual constraints in any market when better feasible allocations are clearly available; (2) If some agents experience excess supply in a given market, none should experience excess demand, and conversely (this does not exclude some agents perceiving themselves as unconstrained in either case).

For consumers (households), point (1) essentially means that for items of consumption there should be no lower constraint (because it must always be feasible for a household to choose to consume less than some given amount) but there can be upper limits (shortage conditions may limit individual consumption, in ways that would depend on the prevailing rationing or queueing scheme). In relation to labor supply (usually measured as the inverse of "leisure" in the utility function), there can be no upper limit on leisure (people cannot be forced to work), but there can be a lower limit (there might not be enough available work at the prevailing wage rate).

Point (2) means that there are not simultaneously agents experiencing excess demand and agents experiencing excess supply in the same market. It rests on rather strong assumptions about market information, but at least in market situations that are not too volatile it is not unreasonable to suppose that agents would quickly learn about any given market⁴—hence the experience of any agent is compatible with the aggregate picture in that market.

With points (1) and (2) in force, the number of candidates for an equilibrium in the market is greatly reduced, though it is still possible for some markets to experience shortage (excess demand) while others are characterized by excess supply. Moreover, we have not specified an allocation rule. For instance, for markets in excess supply, such as the labor market, there are social/economic rules in force in any given society that help to determine how the given excess supply is to be apportioned among the households. In practice, what tends to happen is that most people are able to work as much as they wish (subject to some constraints on the available set of employment contracts) whereas others cannot work at all and are unemployed. However, in principle one could imagine a rule that ensured a minimum amount of work to all. Similarly, in markets with excess demand, supplies could be allocated to those with the best connections, to those in certain occupations or living in certain locations; they could also be allocated equally by means of an imposed rationing procedure. These details, important as they often are, will not concern us here. Instead, we focus on two polar cases⁵ (others can be defined, as in Malinvaud 1977⁶):

1. A *Keynesian equilibrium* is a constrained equilibrium in which households face excess supply in (all) the goods markets and in the labor market.
2. A *shortage equilibrium* is a constrained equilibrium in which households face excess demand in (all) the goods markets and in the labor market.

The first of these gives rise to the familiar Keynesian multiplier, while the second is associated with the so called supply multiplier (see Barro and Grossman 1974). A great deal of analysis has been carried out using simplified versions of this model, with just three markets: goods, labor, and money. The first two markets experience transactions constraints, the last is unconstrained.

The main point for present purposes, however, is that this general conception of a shortage equilibrium implies an immediate policy recommendation, namely “get the prices right”: specifically, it will be necessary both to cut the real wage (to reduce aggregate demand) and to allow relative goods prices to adjust, too. For the model developed here implies that if only we can shift the prices toward the Walrasian

prices p^* we should be able to move the economy toward an equilibrium without shortage and the accompanying phenomena of rationing, queuing, and the like. Further, under fairly general conditions, this new equilibrium should be a Pareto improvement over the old one: even though the measured real wage is lower than formerly, the removal of transaction constraints should leave households, on balance, better off.

For the centrally planned economies, this approach led to recommendations for price reform as the most effective means of overcoming persistent shortage, and under the more recent conditions of early transition it stimulated advice favouring comprehensive price liberalization. Getting the prices right—in particular, the relative prices—was seen as the key to doing away with shortage.

For Kornai, this essentially neoclassical approach was regarded as fundamentally wrong, especially as a story about centrally planned economies, largely because it omitted some key institutional and behavioral features of these economies from the analysis. These features, as I indicated in the introduction to this chapter, included the *soft budget constraint* and the notion of *normal shortage*. Kornai was especially interested in the behavior of enterprises under socialism,⁷ because the sphere of production was the domain in which the standard neoclassical assumptions—autonomous firms, cost-minimization and/or profit-maximization, an institutional framework in which agents reliably fulfilled their contracts, and so on—were least adequate. The notion of a soft budget constraint in this context was supposed to capture a number of ideas, including:

1. enterprises under socialism could not go bankrupt;
2. they faced little or no competition either through new entry or via foreign trade (i.e., imports);
3. when an enterprise ran into financial problems and incurred deficits it could always obtain additional resources through several channels: budgetary subventions; relief from repayment obligations on existing credits; additional credits; tax relief or deferred taxation; favorable price “adjustments.”

The difficulty with this approach to the soft budget constraint was its apparent implication that enterprises might have access to virtually unlimited resources. However, Kornai himself never claimed this and was always careful to build models of shortage in which these channels for gaining resources gave rise to inefficiencies but never to an unlimited extent. What limited the possibilities for enterprises to gain additional resources was the idea that centrally planned economies tended to operate in such a way that they maintained a normal degree of shortage over time.⁸ In Kornai’s analysis, it is not always easy to understand exactly how a particular

“normal shortage” might have been established in a given economy, but the mechanisms for its reproduction were clear enough (see Hare 1989).

Suppose, for instance, that shortage normally amounted to five per cent of GDP in a certain economy, and that an investment boom not accompanied by a parallel increase in the production of consumer goods was giving rise to increasing shortages of these goods, because the additional wages paid to the workers producing investment goods were not matched by extra output on which to spend the wages. These shortages would lead to *reductions* in labor supply (via the supply multiplier) and increasing pressure on the planners to introduce measures to ameliorate the situation. What could they do? In practice, what they tended to do was ban or delay new investment starts, shift goods from export markets to the domestic market where that was feasible, and possibly shift resources to boost consumer goods production. After a time these measures would simultaneously increase the production of those goods and services where shortages had become most acute and also reduce demand in the economy, especially for investment goods. The result would be a gradual return to the economy’s normal degree of shortage, rather than a return to an equilibrium without shortage.

Conversely, if for some reason, such as an unexpected fall in (planned) exports, shortage was less severe than usual, enterprises demanding additional investment resources would be treated more sympathetically by the planners and corresponding new credits would be approved. Investment would then start to rise and before long the new activity would draw off some of the resources that would otherwise have been consumed by households or used to manufacture consumer items. Hence shortage would worsen again, returning to its normal state.

Two points may be noticed about this mechanism: (1) it is assumed that under all conditions, enterprises would like to demand resources additional to those already allocated to them; and (2) there is little or no use made of price adjustments.

Point (1) is not, of course, a characteristic of a neoclassical firm. However, a feature of a firm with a soft budget constraint is that it always wants more resources, either to help it fulfill the plan, or for additional investment, or for other purposes. Kornai described this quite graphically as the “almost infinite” demand for material inputs, the “almost infinite” demand for investment, or the “almost infinite” demand for credit. As a description, this terminology conveys what is meant very well, though it does not provide a very natural starting point for a more formal analysis (what are we supposed to mean by “almost infinite,” for instance?). At the very least, it implies that firms under socialism always demand more than a corresponding capitalist firm with a hard budget constraint would do.⁹ If a firm always wants more resources, and by implication it feels no particular responsibility to

ensure that they are used to achieve high returns, then there must be someone in the economy to limit what resources firms can actually obtain.¹⁰ Under central planning, this can only be the planners themselves, and they operate in the manner described above to sustain whatever degree of shortage has become the accepted norm.

Point (2) simply reflects that fact that under socialism firms mostly deal with their superiors, as well as with their customers and suppliers, by means of quantitative targets and controls, so that prices hardly matter at all. There is an obligation to supply which generally has nothing to do with the profitability of a given item or even with the state of demand—it only depends on the output “demanded” by the planners. Within the sphere of production, therefore, price changes hardly affect anything other than the cash flows passing between enterprises and hence their resulting financial balances at the end of the plan period. But as we have already implied, the prevalence of soft budget constraints makes these largely inconsequential.¹¹ However, although price changes might have little effect on current demands and supplies, persistently irrational prices can have other effects not allowed for in this analysis. Thus irrationally low prices of freight transport, energy, and land (the last being treated as free) encouraged the socialist economies to adopt patterns of investment that now, after-price liberalization, seem wildly inefficient and call for huge amounts of economic restructuring.

For households prices are typically more important, because even when there are widespread shortages, price rises for rationed goods result in income effects on the demand for unrationed goods. Households do not have soft budget constraints and are therefore responsive to price signals even under socialism. However, because of adjustments by firms to maintain the established, normal degree of shortage (as described earlier), household responses to price changes cannot usually eliminate shortage by themselves.

Although households behave more or less in the neoclassical manner, firms in Kornai's framework cannot be assumed to be responsive to prices in the usual way. Instead, production, y , depends on planned production, y^c , and on the (financial) resources available to firms, r . Note that r is certainly not assumed to be unlimited; it will depend on the degree of softness of the budget constraint. Hence $y = y(y^c, r)$, though this might be a very complex relationship lacking any especially nice properties; and as usual, $x = x(p)$ (where I am neglecting any distinction between consumer and producer prices for simplicity). Then excess demand, $z = x(p) - y(y^c, r) - \bar{x}$, and an *equilibrium with a normal degree of shortage* is a (consumer) price vector p and a plan y^c with the property that $z = z^n$, the latter vector representing the established normal shortage.

Rather than seek to prove formal results about this Kornai-type model of a shortage equilibrium, I simply assert that under reasonably general conditions an equilibrium could be found. It is clear in this framework that shortage has little or nothing to do with simple-minded explanations that run in terms of planners' mistakes, excessively taut plans, or insufficient production of certain goods. It is the institutional conditions of central planning, notably the enterprise-level soft budget constraints, that allow shortage to be sustained once it has become an established condition. This is demonstrated by Kornai in a variety of interesting models—both sectoral and economy-wide—including a growth model along whose equilibrium growth path the established degree of shortage is maintained (see Kornai 1982).

Kornai's analysis of the soft budget constraint is descriptively very appealing (to the extent, as indicated above, that the term is now very widely used, even by many who may be unaware of its origins) and other economists have responded by seek to develop analytical models to explain the phenomenon more formally. One such response has been to construct models of agents' behavior under central planning, to understand better the circumstances under which soft budget-type phenomena might arise, another has been to refine the concept itself to distinguish between different aspects or manifestations of the phenomenon. A particularly interesting instance of the latter is Gomulka (1986, notably chapter 5), which introduces the notion of *budget flexibility*. Gomulka correctly points out that in arriving at their plans for a given period, managers are competing for resources that, in aggregate, are plainly limited. Hence they cannot demand virtually unlimited inputs to fulfil very modest plans, for the managers would then be risking dismissal or other penalties for gross inefficiency. Managers need to arrive at a form of compromise between what they would like and what they perceive as acceptable to their superiors in the planning hierarchy, and *ex ante* they are likely to arrive at a point where they perceive their budgets to be hard at the margin (because if they perceived additional resources to be readily available, they would ask for them). *Ex post*, however, it may turn out to be impossible to fulfill the original plan or economic conditions may change adversely, and some or all firms might then need additional financial resources to keep going. Gomulka then defines budget flexibility as "the speed at which the firm's total income, including subsidies and the effect of increasing the prices of its outputs, changes in response to a change in the value of its inputs." His point is that an essential aspect of softness depends on timing: it is not just *whether* an enterprise can secure additional resources, but *how rapidly*. This distinction turns out to be important when thinking about the policy implications of soft budgets.

Drawing on this discussion among others, Schaffer (1989) develops a simple game theoretic model to analyse the situation where the central authorities cannot make credible commitments about their future behavior toward enterprises; in particular, they cannot credibly commit not to provide additional resources or an easier plan target when an enterprise does not perform as expected. Then bailouts form part of the equilibrium in a repeated game with perfect information. With imperfect information—firms do not know at the start whether the state will be “tough” or “weak,” it can pay the state to build a reputation for toughness since then enterprises will believe that there will be no rescues, and will perform better. The perfect information case, with a paternalistic state, corresponds most closely to Kornai’s soft budget constraint analysis.

As far as policy is concerned, Kornai’s analysis has very different implications from the neoclassical model examined above. Under socialism, price reform might be justified to remove some of the irrational distortions referred to above, but it will do little or nothing to eliminate shortage *per se*. Instead, what is needed is strong measures to harden enterprise budget constraints to ensure that enterprise “demand” becomes more responsive to price signals; this in turn paves the way for a more effective price reform that would indeed affect enterprise behavior and hence might even reduce the economy’s equilibrium (or normal) degree of shortage. Unfortunately, no one discovered a way of achieving this result under the socialist system, though it should not, in principle, have been impossible to enforce harder budget constraints. Even without going as far as bankruptcy, this could have been done by making it harder for enterprises to obtain additional credits, for instance.

In the early transition years, this conception of the socialist economy also has implications for the likely outcome of price liberalization. As McKinnon (1991) observed, liberalizing prices while the economy is still dominated by large, state-owned monopolistic enterprises with soft budget constraints—confident that they will not be permitted to go under—is a recipe for disaster in the form of more or less unbounded inflation. This, indeed, was observed in several countries, especially those in the CIS, where enterprises had operated under central planning for longer and where they felt correspondingly more secure (not in the sense that they really believed their economic position was sound, but in the sense that they believed that could rely on continuing state financial support). Instead, a successful price liberalization must be accompanied or even preceded by determined measures to convince enterprises that the era of soft budget constraints was over, as was done in much of central Europe. This requires significant measures in a number of areas of policy having to do with enterprise taxes and budgetary subsidies to enterprises, access to

credit and credit repayment conditions,¹² and the legal framework governing bankruptcy and liquidation. In addition, as we shall note below, certain aspects of the competitive environment are also crucial.

3 Empirical Evidence

For most countries, the first steps of the transition were the formal abandonment of central planning,¹³ measures to bring about or restore macroeconomic stabilization, and surprisingly quick and comprehensive price and trade liberalization. Where this combination of measures was put in place with determination and commitment, in a way that assured the government a reasonable degree of credibility, conditions of shortage disappeared almost overnight in most markets. Moreover, most households and firms appeared to be sufficiently convinced that their "world" had changed that there was a general expectation that shortage would not recur. Thus in Kornai's sense, the economy's institutional framework and policy stance had shifted in such a way as to reduce the "normal degree of shortage" to zero. Thinking in neoclassical terms, one could say that relative prices in the economy had finally been freed sufficiently to enable them to adjust to move the economy toward a Walrasian equilibrium. However, closer investigation of these lines of argument shows that neither is wholly satisfactory on its own.

In those countries where elements of central planning were retained for a time—such as the branch ministerial structure, the system of state orders, elements of the centralized supply system, central controls over foreign trade, and the like—either shortage was not eliminated so quickly or so completely, or lax or nonexistent macroeconomic control allowed it to be transformed into unbearably high inflation. Following the dissolution of the USSR, for instance, most of the successor states failed to understand the importance of (or were politically unable to impose) a firm monetary and fiscal policy to limit inflationary pressures. The result, for a couple of years, was rates of inflation exceeding 1000 per cent per annum in many countries. In a few countries, such as Russia, other reforms like privatization were going ahead long before stabilization was effectively achieved, contrary to the normally assumed "correct" sequencing (on which see Portes 1991).

Thus price and trade liberalization were clearly necessary conditions for the elimination of shortage, but equally clearly they were not sufficient and required accompanying measures of macroeconomic restraint to prevent enterprises from translating shortage into inflation. However, this link between shortage and inflation needs further analysis, inasmuch as the standard neoclassical story would lead one

to expect no more than a sharp, one-off adjustment in relative prices—such as a rise in most consumer goods prices relative to the money wage, and/or relative to money balances in order to eliminate the so-called monetary overhang—rather than a sustained inflationary episode. Thinking about this brings us back to the question of soft budget constraints.

If budget constraints are hard for all agents (i.e., both firms and households) then agents are responsive to prices in the standard ways assumed in microeconomic theory and a round of price increases should be enough both to choke off any initial shortage *and* to create a situation where there is no pressure for further rounds of price changes (especially increases). The economy simply moves to a new equilibrium characterized by a different set of relative prices (and possibly the same money wage, if institutional conditions make that, in effect, the numeraire). This is a grossly oversimplified story, of course, even in a well functioning market-type economy, because, as Stiglitz (1994) emphasizes, there are many markets in which information other than prices is used by agents to form the equilibrium. Such non-price information is crucial, for example, in situations where there is asymmetric information giving rise to problems of moral hazard and adverse selection, or where price signals alone would give rise to coordination failures. Nevertheless, even in these markets we typically assume that agents behave as if they possessed hard budgets in the sense that they at least had to break even and should preferably earn normal profits. Then conventional approaches to stabilization basically work and do not normally stimulate a sustained inflationary spiral.

In the former centrally planned economies, conventional approaches to stabilization were employed (often on IMF advice) without the right institutional conditions being in place at the microeconomic level. In the presence of soft budget constraints for enterprises, getting inflation down requires tougher policies than would otherwise be needed (that is, in a more price-responsive economy). Likewise, with a moderate stabilization based on typical western experience, transition economies experienced larger drops in output and higher and more sustained inflation than would otherwise have been expected (for details, including analysis of the connections between stabilization, reforms and growth in the transition economies, see Hare 1996; Sachs and Warner 1996; Fischer, Sahay, and Vegh 1996a and 1996b; Aslund, Boone, and Johnson 1996).

In the postcommunist world, therefore, it is evident that neither the initial measures of liberalization, mostly concerned with prices and foreign trade, nor macroeconomic stabilization policies, have been sufficient to eliminate the phenomenon of soft budget constraints for enterprises. Soft budget constraints in a world in which the restraining influence of the planners has been removed are evidently economi-

cally disastrous, largely for reasons that relate more to the institutional conditions discussed by Kornai rather than to the formal supply and demand relationships taken for granted in the neoclassical analysis. These institutional conditions have to do with several aspects of the environment in which firms operated under central planning, and still operate in the early years of transition in many countries. For space reasons I cannot provide here a full analysis, but a few key points must suffice to highlight the features of interest; these refer back to, and modify, the points made about the soft budget constraint when it was defined in the second section of this chapter.

- State-owned enterprises, and even in many cases, recently privatized enterprises, are not permitted to go bankrupt (even in those countries that have formally adopted quite reasonable bankruptcy laws); whether bankruptcy laws are adequate or not, structures of corporate governance are often not yet sound enough to enforce effective adaptation of firms to market conditions.
- Firms often still face very limited competition either through new entry or via foreign trade (that is, imports), despite the liberalization that has occurred.
- When an enterprise runs into financial problems and incurs deficits it can still often obtain additional resources through several channels: official tolerance of the non-repayment of existing credits; additional (new) credits, usually from the state-owned bank it is used to dealing with; tolerance of delayed tax payments and delayed social security contributions. In part, what is at issue here is the credibility of the government and the engrained expectations of enterprises that the government should act, in a very real sense, as their protector. Under central planning, after all, the government and its agencies were responsible for guiding and developing firms, allocating investment resources across the economy, and allocating production targets and input supplies. Few firms entered any given market, and virtually none left. Hence what we think of as the normal processes of restructuring in a market-type economy, an active selection process in which some firms survive and prosper while others decline and disappear and yet other new ones enter the market, hardly functioned at all under central planning.

The resulting mental attitude concerning survival and government responsibility is still widely entrenched in the transition economies, especially in the former Soviet Union. This has made it immensely difficult for governments wanting to be firmer with their business sector to gain the necessary credibility. However, much of the reason for this difficulty is also linked to weaknesses in the implementation mechanisms associated with economic reform programs, and related institutional deficiencies.

Under central planning, the principal implementation mechanism operated through the Communist Party, which established priorities and enforced official policies at all levels where it was necessary. When the party cells were removed from firms and other organizations relevant for the economy, this left an institutional vacuum that has not yet been filled effectively in many countries in the former Soviet Union (though the situation is incomparably better in much of Central and Eastern Europe). The resulting institutional "space" is being occupied in a variety of ways: maintenance of traditional links between firms; the development of more or less openly criminal networks to siphon off funds, block competition and maintain established monopoly positions; corrupt relationships between enterprises and their formal supervisory bodies in connection with such issues as privatization and corporate governance; strengthening of links between enterprises and "their" banks; and so on.

The presence of these mechanisms need not be wholly undesirable, for they are perfectly compatible, as much evidence on enterprise adjustment shows, with many enterprises taking some significant steps toward restructuring along economically sensible lines (see Carlin et al. 1995; Johnson and Loveman 1995). Moreover, in the presence of weak and relatively ineffective governments and widespread lack of respect for (or even understanding of) the Western notion of the "rule of law," there may not have been much alternative. Also, even in developed market economies one can find many instances of such informal and even semilegal institutional arrangements that are commonly regarded as "oiling the wheels of commerce." What is different in some of the transition economies, therefore, is not just the existence of these mechanisms but their pervasiveness and the lack of a wider, law-based and generally accepted institutional structure within which they can operate subject to clear constraints.

In these conditions soft budget constraints (essentially, the generalized lack of financial discipline at enterprise level, evidenced most significantly in easy access to additional credit or additional government funding), and the associated lower-level institutional deficiencies (including very weak corporate governance) that can be found still in many transition economies, actually make macroeconomic stabilization far harder than it would otherwise be. Thus, for instance, if firms are delaying tax and social security payments on a large scale, then governments wishing to avoid large government deficits (and usually lacking an adequate securities market through which they could otherwise finance a deficit) are compelled to cut public spending far more drastically than they might otherwise do. In this way, microeconomic "misbehavior" exacerbates macroeconomic stabilization problems that were already severe enough. This was precisely Russia's situation in 1996, where tax rev-

enues barely reached two-thirds of what was expected, the government was apparently unable to pull in any more revenue, and hence to meet something like the deficit target agreed with the IMF real spending had to be pruned.¹⁴ In contrast, countries like Hungary and the Czech Republic (where tax rates may, from some points of view, be too high) are able to raise the tax revenue they need and have already developed tolerably well functioning securities markets. In those countries, however, enterprises undoubtedly feel far less protected by the authorities—budget constraints are no longer especially soft.

4 Conclusions

What have we learned about the connections between *shortage* and the *soft budget constraint* from this discussion?

First, under central planning, both shortage and the soft budget constraint were widespread phenomena. Second, the two main theoretical approaches focused on different elements in the story. Thus the neoclassical analysis usually ignored soft budget constraints, paying attention instead to the combinations of distorted relative prices and transaction constraints that could give rise to shortage in many markets in a constrained equilibrium. In contrast, Kornai devised an analytical framework that placed the soft budget constraint at center stage, to the extent of explaining shortage and its reproduction in models that made almost no reference to relative prices. It was conceded that prices could influence financial flows between firms but they were not expected to influence significantly firms' resource allocation decisions. Kornai's analysis implied that a pervasive soft budget constraint—an important innovation in economic thinking for which Kornai deserves full credit—was associated with persistent shortage under socialism.

Third, the early experience of transition showed beyond doubt that rapid price and trade liberalization could indeed eliminate the bulk of shortages astonishingly rapidly, apparently lending credence to the neoclassical explanation of shortage that ran in terms of "incorrect relative prices." However, those countries that liberalized, while also keeping in place significant elements of the former economic system, and while failing to develop the sort of institutions needed to operate a market-type economy effectively, quickly encountered other serious problems. Initially, these mostly took the form of rampant inflation. Thus in the countries worst affected, even when the most visible shortages largely disappeared, financial indiscipline both at governmental and enterprise levels allowed inflation to take off and crucial micro-level adjustment was greatly delayed as a result.

Fourth, what this demonstrated was that although most shortages could be eliminated by means of relative price adjustments, the soft budget constraint could remain a problem if many of the specific institutional characteristics that gave rise to soft budgets under socialism remained in place. Thus one might still be able to argue that the soft budget constraint was necessary for shortage, but it has turned out not to be a sufficient condition. In the early post-transition years soft budgets can persist and can continue to inflict substantial economic damage by removing or weakening enterprise-level incentives for adjustment and restructuring, without giving rise to renewed shortage.

Finally, in the light of these conclusions, what might be the most effective way forward for the transition economies still experiencing widespread soft budgets? Several answers can be given, albeit briefly: (1) the growth of the *de novo* private sector, comprising firms not wedded to the old economic system and its procedures, and increasingly dominating supply in new markets; (2) the physical decline of many existing SOEs, as their capital wears out and they lose their traditional markets—eventually, governments will simply be unable to subsidize them (directly or indirectly) enough to keep them going any longer; (3) the stimulus of competition from both imports and other domestic producers (forcing improvements in quality, technology, marketing, and so on), highlighting the potential importance of an effective commercial and competition policy; (4) the development of new, increasingly profit-oriented banks and other financial institutions that will simply not provide additional credit on the former terms; and (5) the gradual disengagement of the government and its agencies from the bulk of production and economic activity, except for a continuing role in setting and maintaining the “rules of the game” for a market-type economy. Among other things, this last condition takes for granted that corporate governance will also gradually be strengthened.

None of these conditions is easy to fulfill, and most are only imperfectly or incompletely in effect even in the most developed market economy. However, substantial progress in these directions is needed in order to overcome soft budget constraints.

Notes

I am grateful to András Simonovits, Mark Schaffer, and participants in seminars at Heriot-Watt and St Andrews Universities for suggestions that have enabled me to improve the paper considerably; remaining errors are my own.

1. Some very interesting theoretical models as well as empirical analysis have been developed using this approach. See Kornai and Martos 1981; Lackó 1989; Kapitány et al. 1989. For a model of a growing economy possessing a dynamic equilibrium with shortage, see Kornai (1982).

2. I restrict attention to functions here. Much of the literature also covers the more general case of an excess demand correspondence.

3. Hence the income terms m_h are defined as follows: $m_h = p \cdot \bar{x}_h + \sum s_{hf} \cdot (p \cdot y_f)$, that is, all profits generated by firms are distributed to households according to their respective shareholdings.
4. Those experiencing excess demand clearly have an incentive to do so; those experiencing excess supply have an incentive to buy more than they need in order to operate an informal/secondary market to supply those less well informed or less well connected. Such situations were not uncommon under central planning.
5. In both cases, only the goods markets and the labor market are mentioned here. Most interesting models in the literature also, of course, include a "market" for money, because to develop a useful model there has to be at least one market in which agents are not constrained. Then in a shortage economy, goods are in short supply, labor supply is less than under a Walrasian equilibrium and there is excess demand for labor, and households accumulate "excess" money balances.
6. It should be noted, however, that Malinvaud's interesting model does not satisfy Walras's Law, inasmuch as profits are not distributed to households; on the other hand, in the single period of the model there is an exogenous component of demand, namely the government's demand for goods and services.
7. Households were supposed to behave more or less in the standard neoclassical manner, as utility maximizers subject to (hard) budget constraints, and possibly also subject to additional transaction constraints as indicated.
8. To define such a normal degree of shortage implies that we can measure shortage. Kornai devotes a great deal of effort to this task, explaining that it is wrong to measure it in a one-dimensional, oversimplified manner and insisting on multidimensional, complex approaches.
9. Qian (1994) uses the notion of soft budgets to explain shortage in a model where investors compete with households for certain inputs. The point is that a bad project may be able to secure funding, but in the presence of shortage it might not be able to obtain all the inputs it needs: hence at least some bad projects might be discouraged, improving the average quality of investment.
10. This suggests an interesting contrast: under socialism the planners are always trying to cut back on enterprises' excessive demands for investment resources, whereas under capitalism it is more usual for the authorities to be seeking ways of stimulating investment demand.
11. Strictly, this point would have to be amended a little if the planners had adopted priorities that favored certain sectors or activities over others.
12. McKinnon himself suggested that state-owned firms should be denied access to any additional credit. He justified this tough approach by arguing that in practice, given the distortions that characterized the latter years of central planning, it would not be easy to discriminate between those firms that were truly creditworthy and those that were not.
13. This step was already taken in Hungary as long ago as 1968, in the sense that central plans were not thereafter decomposed into compulsory plans for each state-owned enterprise. This form of planning also faded away in Poland during the 1980s but was in force everywhere else right until the collapse of communism itself.
14. One might wonder, in this regard, why Russia apparently finds it politically easier to accept further cuts in an already reduced public spending program rather than to take effective measures to enforce tax discipline.

References

- Alfandari, Gilles, and Mark Schaffer (1996). Arrears in the Russian Enterprise Sector. CERT Discussion Paper 96/8. Edinburgh: Heriot-Watt University.
- Amsden, A. H., J. Kochanowicz, and L. Taylor (1994). *The Market Meets Its Match*. Cambridge, Mass.: Harvard University Press.
- Arrow, K., and F. Hahn (1973). *General Competitive Analysis*. New York: Holden Day.

- Aslund, A., P. Boone, and S. Johnson (1996). How to Stabilize: Lessons from Post-Communist Countries. Mimeo.
- Barro, R. J., and H. I. Grossman (1974). Suppressed inflation and the supply multiplier. *Review of Economic Studies* 42: 87–104.
- Barro, R. J., and H. I. Grossman (1976). *Money, Employment and Inflation*. Cambridge: Cambridge University Press.
- Benassy, J. P. (1982). *The Economics of Market Disequilibrium*. New York: Academic Press.
- Carlin, W., J. van Reenan, and T. Wolfe (1995). Enterprise restructuring in early transition: The case study evidence from Central and Eastern Europe. *Economics of Transition* 3(4): 427–458.
- Davis, Christopher, and Wojciech Charemza, eds. (1989). *Models of Disequilibrium and Shortage in Centrally Planned Economies*. London: Chapman and Hall.
- EBRD. (1995). *Transition Report 1995*. London: European Bank for Reconstruction and Development.
- EBRD. (1996). *Transition Report 1996*. London: European Bank for Reconstruction and Development.
- Fischer, S., R. Sahay, and C. A. Végh (1996a). Stabilization and growth in transition economies: the early experience. *Journal of Economic Perspectives* 10(2): 45–66.
- Fischer, S., R. Sahay, and C. A. Végh (1996b). Economies in transition: the beginnings of growth. *American Economic Review* 86(2): Papers and Proceedings, 229–233.
- Gomulka, Stanislaw (1986). *Growth, Innovation and Reform in Eastern Europe*. Brighton: Wheatsheaf Books.
- Grosfeld, I., and G. Roland (1995). Defensive and Strategic Restructuring in Central European Enterprises. CEPR Discussion Paper No. 1135. London: CEPR.
- Hare, Paul (1982). Economics of shortage and non-price control (review article). *Journal of Comparative Economics* 6: 406–425.
- Hare, Paul (1989). The economics of shortage in the centrally planned economies. In Davis and Charemza (1989).
- Hare, Paul (1997). Investment, Growth and Industrial Renewal in the Transition Economies. CERT Discussion Paper 97/1. Edinburgh: Heriot-Watt University.
- Johnson, S., and G. Loveman (1995). *Starting Over in Eastern Europe: Entrepreneurship and Economic Renewal*. Cambridge, Mass.: Harvard Business School Press.
- Kapitány, Zsuzsa, János Kornai, and Judit Szabó (1989). Reproduction of shortage in the Hungarian car market. In Davis and Charemza (1989).
- Kornai, János (1980). *The Economics of Shortage*. Two vols. Amsterdam: North Holland.
- Kornai, János (1982). *Growth, Shortage and Efficiency*. Oxford: Basil Blackwell.
- Kornai, János (1986a). The Hungarian reform process: visions, hopes, and reality. *Journal of Economic Literature* 24: 1687–1737.
- Kornai, János (1986b). The soft budget constraint. *Kyklos* 39(1): 3–30.
- Kornai, János (1992). *The Socialist Economic System*. Oxford: Oxford University Press.
- Kornai, János (1995). Eliminating the shortage economy: A general analysis and examination of the developments in Hungary. Parts 1 and 2. *Economics of Transition* 3(1 and 2): 13–37, 149–168.
- Kornai, János, and Béla Martos, eds. (1981). *Non-Price Control*. Budapest: Akadémiai Kiadó.
- Lackó, Mária (1989). Sectoral shortage models in Hungary. In Davis and Charemza (1989).
- McKinnon, R. (1991). *The Order of Economic Liberalization*. Baltimore: Johns Hopkins University Press.
- Malinvaud, E. (1977). *The Theory of Unemployment Reconsidered*. Oxford: Basil Blackwell.
- Newbery, David, ed. (1995). *Tax and Benefit Reform in Central and Eastern Europe*. London: CEPR.
- Portes, R. (1989). The theory and measurement of macroeconomic disequilibrium in centrally planned economies. In Davis and Charemza (1989).

Portes, R. (1991). Introduction to *The Path of Reform in Central and Eastern Europe, European Economy*. Brussels: Commission of the European Communities.

Qian, Yingyi (1994). A theory of shortage in socialist economies based on the "soft budget constraint." *American Economic Review* 84(1): 145–156.

Sachs, J. D., and A. M. Warner (1996). Achieving Rapid Growth in the Transition Economies of Central Europe. Mimeo, Harvard Institute for International Development.

Schaffer, Mark (1989). The credible-commitment problem in the center-enterprise relationship. *Journal of Comparative Economics* 13: 359–382.

Stiglitz, J. E. (1994). *Whither Socialism?* Cambridge, Mass.: MIT Press.

World Bank (1996). *From Plan to Market*. World Development Report 1996. Washington, D.C.: World Bank.

12 The Kornai Effect Revisited

Richard E. Quandt

1 Introduction

János Kornai is unequalled in his contributions to our understanding of socialist and centrally planned economies. Among many other descriptive as well as analytical accounts of their functioning, he introduced the observation (1979, 1980, 1982, 1986) that if the firm's budget constraint is soft, that is, if the firm is bailed out every time that its profits are negative, then the firm's demand for inputs tends to be larger than would normally be the case. The reason for this is that firms learn to anticipate the bailouts and adjust their behavior accordingly. This is what we call the "Kornai effect."

Goldfeld and Quandt (1988, 1990a, 1990b, 1992, 1993) formalized this notion in a series of simple models. All the models are characterized by output (or price) uncertainty: the manager must make his input decisions before the output (or price) uncertainty is resolved. The various models differ among themselves in terms of distributional assumptions concerning the random variable representing the uncertainty (continuous or discrete) and whether bailout is assumed to be complete or only partial. In Goldfeld and Quandt (1993), the model is the following.

Output is produced with a single input x according to the production function x^β , $0 < \beta < 1$, but the output that is actually produced is $x^\beta \theta_1$, where θ_1 is a random variable that attains either of the two values $\underline{\theta}_1$ and $\bar{\theta}_1$, $\underline{\theta}_1 < \bar{\theta}_1$. Costs are given by $w x$, where w is the unit cost of the input, but the costs that actually accrue are $w x \theta_2$, where θ_2 can attain the values $\underline{\theta}_2$ or $\bar{\theta}_2$, $\underline{\theta}_2 < \bar{\theta}_2$. It is also assumed that the θ_i are symmetrically placed around unity, so that $\underline{\theta}_i = 2 - \bar{\theta}_i$, that there is given a joint frequency distribution of the θ s, specified by the four numbers $\{p_{ll}, p_{lh}, p_{hl}, p_{hh}\}$, and that the expected value of each of the θ s is equal to unity. Finally, the model assumes that bailout is complete, that is, if operating profit were negative, the actual profit that accrues to the firm is zero. This model is generalized by Magee (1993); she relaxes the assumption of complete bailouts, replacing it with the assumption that if a negative operating profit π is generated, the firm actually experiences $\psi\pi$, $\psi < 1$. In addition, she also introduces taxes, so that if profits are positive, the firm experiences $\tau\pi$ rather than π , $\tau < 1$. Both papers find that expected profit maximization by the firm leads to the Kornai result, but not uniformly; generally the greater the uncertainty and the smaller the correlation between the output and cost uncertainties, the more likely the Kornai effect is to hold.

In the present chapter I both simplify the model and complicate it in certain ways so as to be able to draw additional conclusions. I first examine a simplified version

of the symmetric model, but with no cost uncertainty. This will allow us, *inter alia*, to plot the mean-variance frontier for output as a function of the bailout coefficient ψ , so that the central authorities, which may have a utility function over output mean and variance, can set ψ optimally. Second, we introduce a simple asymmetric case, which permits an examination of the effect of higher moments on the prevalence of the Kornai effect. Finally, we consider the possibility that the manager can secure improved treatment at the hands of the bailout agency by spending some effort and expense on persuading the bailout agency of the worthiness of the firm's cause ("whining").

2 The Symmetric Model

Basic Considerations

As in the first section, output will be given by $x^\beta \theta$, where $\beta < 1$, $\theta = \underline{\theta}$ with probability 0.5 and $\theta = \bar{\theta}$ with probability 0.5; $\underline{\theta} + \bar{\theta} = 2$, which insures that expected output is equal to 1.0. Costs are given by $w x$. The firm is an expected profit-maximizer. If operating profit is positive, the firm is permitted to keep the entire profit. If operating profit is negative, a portion ψ accrues to the firm, with the portion $1 - \psi$ being reimbursed to the firm by the central authorities. Operating profit is positive in the first state of nature if

$$0 < x < \left(\frac{\underline{\theta}}{w} \right)^{1/(1-\beta)} \quad (12.1)$$

and positive in the second if

$$0 < x < \left(\frac{\bar{\theta}}{w} \right)^{1/(1-\beta)}. \quad (12.2)$$

It follows immediately that the expected profit function is

$$E(\pi) = \begin{cases} x^\beta - wx, & \text{if } x < \left(\frac{\underline{\theta}}{w} \right)^{1/(1-\beta)}; \\ 0.5(x^\beta \bar{\theta} - wx) + 0.5\psi(x^\beta \underline{\theta} - wx), & \text{if } \left(\frac{\underline{\theta}}{w} \right)^{1/(1-\beta)} \leq x < \left(\frac{\bar{\theta}}{w} \right)^{1/(1-\beta)}; \\ \psi(x^\beta - wx), & \text{if } x \geq \left(\frac{\bar{\theta}}{w} \right)^{1/(1-\beta)}. \end{cases} \quad (12.3)$$

The optimal solution in the first and third branches of $E(\pi)$ is

$$x^{(1)} = x^{(3)} = \left(\frac{\beta}{w} \right)^{1/(1-\beta)}, \quad (12.4)$$

and in the second branch it is

$$x^{(2)} = \left(\frac{\beta}{w} \right)^{1/(1-\beta)} \left(\frac{\bar{\theta} + \psi \underline{\theta}}{1 + \psi} \right)^{1/(1-\beta)}. \quad (12.5)$$

Coherence requires that each $x^{(i)}$ satisfy the region in which the branch from which it is derived is valid. It follows immediately from comparing $x^{(1)}$ and $x^{(2)}$ in (12.4) and (12.5) with the bounds in (12.3) respectively that $x^{(1)}$ will be a valid solution if

$$\beta < \underline{\theta}, \quad (12.6)$$

and that $x^{(2)}$ will be a valid solution if

$$\underline{\theta} < \beta \left(\frac{\bar{\theta} + \psi \underline{\theta}}{1 + \psi} \right) < \bar{\theta}. \quad (12.7)$$

It follows immediately that

1. $x^{(3)}$ can never be a valid solution, for it would require $\beta > \bar{\theta}$, which is a contradiction.
2. The relative magnitudes of the interior solutions obey $x^{(1)} < x^{(2)}$, because $(\bar{\theta} + \psi_0 \underline{\theta}) / (1 + \psi_0) > 1$.
3. Either region 1 or region 2 must contain a valid interior solution, for the right-hand inequality on (12.7) always holds and asserting the opposite of the proposition then yields

$$\beta > \underline{\theta} \quad \text{and} \quad \underline{\theta} > \frac{(\bar{\theta} + \psi \underline{\theta})\beta}{1 + \psi},$$

which is a contradiction.

4. It then further follows that the Kornai effect will hold if either $x^{(1)}$ is not a valid solution, or if $x^{(1)}$ and $x^{(2)}$ are both valid solutions, the latter yields a higher expected profit.

We next plot the regions where the two solutions are valid. Letting $\underline{\theta}$ be on the horizontal axis and β on the vertical, $x^{(1)}$ is valid everywhere below the 45° line and $x^{(2)}$ is valid everywhere above the curve given by $\beta = \underline{\theta}(1 + \psi) / (\bar{\theta} + \psi \underline{\theta})$, denoted by **B**. In between, both solutions are “valid,” and the one with the higher expected profit will be chosen. It is easy to verify that the two expected profits are

$$E(\pi_1 | x^{(1)}) = \left(\frac{\beta}{w}\right)^{1/(1-\beta)} \left(\frac{w}{\beta}\right)(1-\beta)$$

and denoting $(\bar{\theta} + \psi\theta)/(1+\psi)$ by ϕ ,

$$E(\pi_2 | x^{(2)}) = 0.5 \left(\frac{\beta}{w}\right)^{1/(1-\beta)} \phi^{1/(1-\beta)} \left[\left(\frac{\beta}{w}\right)^{-1} (1+\psi) - w(1+\psi) \right].$$

It follows that $E(\pi_1 | x^{(1)}) - E(\pi_2 | x^{(2)}) > 0$ if and only if

$$\beta < 1 - \frac{\log \phi}{\log 2 - \log(1+\psi)}, \quad (12.8)$$

The curve between the 45° line and the line marked **B** plots the right-hand side of (12.8). Figures 12.1 and 12.2 depict the situation for values of ψ 0.5 and 0.05 respectively. It is clear that (1) for any value of β , the smaller is θ , and hence the greater the variance of the production uncertainty, the greater the likelihood that the Kornai effect will hold, (2) the smaller is ψ , the larger is the region in which the Kornai effect holds.

The Mean-Variance Locus

We now introduce the explicit assumption that θ is distributed uniformly on $(0, 1)$. $x^{(1)}$ is chosen if (12.8) is satisfied, in which case the level of output does not depend

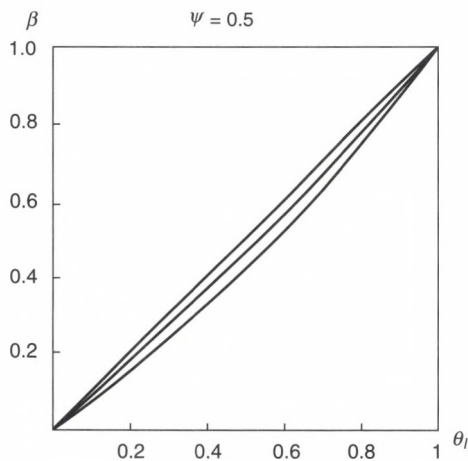


Figure 12.1

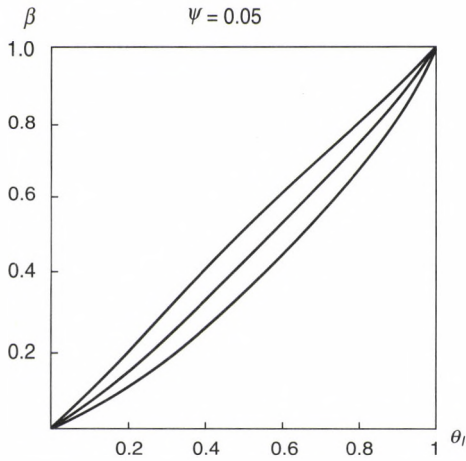


Figure 12.2

on $\underline{\theta}$ any more. $x^{(2)}$ is chosen in the reverse case, but the level of output in this event is a function of $\underline{\theta}$ and can be written as $x^{(2)}(\underline{\theta})$. Denote further the right hand side of (12.8) by $1 - k(\underline{\theta})$; then the expected value of output can be written as

$$E(x) = \left(\frac{\beta}{w}\right)^{1/(1-\beta)} \text{Prob}(\beta < 1 - k(\underline{\theta})) + \int_{\beta > 1 - k(\underline{\theta})} \left(\frac{\beta}{w}\right)^{1/(1-\beta)} \phi(\underline{\theta})^{1/(1-\beta)} U(\underline{\theta}) d\underline{\theta}, \quad (12.10)$$

where $U(\underline{\theta})$ is the density function of $\underline{\theta}$. We proceed by the following steps.

1. Because $\phi = (2 - \underline{\theta}(1 - \psi))/(1 + \psi)$ and $\underline{\theta}$ is uniform on $(0, 1)$, we have

$$\underline{\theta} = \frac{2 - (1 + \psi)}{1 - \psi}$$

$$\frac{d\underline{\theta}}{d\phi} = -\frac{1 + \psi}{1 - \psi}$$

and hence the density of ϕ is

$$f(\phi) = \frac{1 + \psi}{1 - \psi}$$

over the range from 1 to $2/(1 + \psi)$.

2. We next need the density of $\log \phi$. Let $z = \log \phi$; then the required density is

$$g(z) = \frac{1+\psi}{1-\psi} e^z$$

over the range from 0 to $\log(2/(1+\psi))$.

3. We next obtain the density of $v = z/A$, where $A = \log(2/(1+\psi))$. It is easily seen to be

$$h(v) = \frac{1+\psi}{1-\psi} A e^{Av}$$

over the $(0, 1)$ interval.

4. Integrating this density yields the probability in (12.10) as

$$\text{Prob}(\beta < 1 - k(\underline{\theta})) = \frac{1+\psi}{1-\psi} \left[\left(\frac{2}{1+\psi} \right)^{1-\beta} - 1 \right]. \quad (12.11)$$

Performing the integration in (12.10) yields for the second term

$$\left(\frac{\beta}{w} \right)^{1/(1-\beta)} \left(\frac{1+\psi}{1-\psi} \right) \left(\frac{1-\beta}{2-\beta} \right) \left(\frac{2}{1+\psi} \right)^{2-\beta/(1-\beta)} \left[1 - \left(\frac{2}{1+\psi} \right)^{\beta(\beta-2)/(1-\beta)} \right]. \quad (12.12)$$

Expected output therefore is

$$E(x) = \left(\frac{\beta}{w} \right)^{1/(1-\beta)} \frac{1+\psi}{1-\psi} \left[\left(\frac{2}{1+\psi} \right)^{1-\beta} - 1 \right] + \left(\frac{\beta}{w} \right)^{1/(1-\beta)} \left(\frac{1+\psi}{1-\psi} \right) \left(\frac{1-\beta}{2-\beta} \right) \left(\frac{2}{1+\psi} \right)^{2-\beta/(1-\beta)} \left[1 - \left(\frac{2}{1+\psi} \right)^{\beta(\beta-2)/(1-\beta)} \right]. \quad (12.13)$$

In completely analogous fashion one may derive the second moment of output, which is

$$E(x^2) = \left(\frac{\beta}{w} \right)^{2/(1-\beta)} \left(\frac{1+\psi}{1-\psi} \right) \left(\frac{1-\beta}{3-\beta} \right) \left[\left(\frac{2}{1+\psi} \right)^{3-\beta/(1-\beta)} - \left(\frac{2}{1+\psi} \right)^{3-\beta} \right]. \quad (12.14)$$

The variance of output is then

$$\text{var}(x) = E(x^2) - E(x)^2. \quad (12.15)$$

Figure 12.3 displays the mean-variance locus for $w = 0.1$ and for $\beta = 0.5$. In figure 12.4 we superimpose on this the locus for $\beta = 0.52$. We see immediately that a higher β (more productive firm) leads to a rightward shift of the locus, thus yielding higher expected output for any given level of variance. In this figure we also superimpose

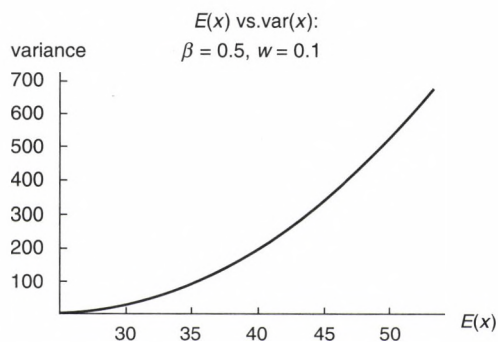


Figure 12.3

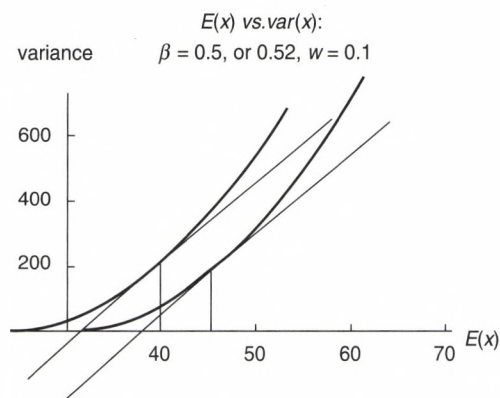


Figure 12.4

two hypothetical parallel linear indifference curves of the central authorities. Because every point on a locus gives the mean-variance combination for a particular value of ψ , the central authorities would want to set the value of ψ so as to maximize their utility. We can determine the ψ values corresponding to the points of tangency. It is, of course, quite to be expected that the optimal mean (optimal from the point of view of the central authorities) for the more productive firm will be higher and the variance lower. What is interesting is that the point of tangency for the more productive firm corresponds to a higher ψ than for the low-productivity firm: approximately 0.38 vs. 0.25. Hence, optimal bailouts will diminish when firms' productivity is high.

3 An Asymmetric Model

Let the firm's production function be, as before, $x^\beta \theta$, where θ is an error term given by

$$\theta = \begin{cases} 1 - \gamma_l \\ 1 + \gamma_h \end{cases} \quad (12.16)$$

γ_l and γ_h are positive and sufficiently small so that $1 - \gamma_l$ is positive. Obviously, if $\gamma_l \neq \gamma_h$, the error term is not symmetric around 1, which is a departure from the case we considered previously.

We still wish to require that $E(\theta) = 1$; hence, if the smaller and larger values of θ occur with probabilities p and $1 - p$, we have

$$E(\theta) = p(1 - \gamma_l) + (1 - p)(1 + \gamma_h) = 1 \quad (12.17)$$

from which it follows that

$$p = \frac{\gamma_h}{\gamma_l + \gamma_h}. \quad (12.18)$$

We also determine the variance, which is

$$\sigma^2 = \frac{\gamma_h}{\gamma_l + \gamma_h} (1 - \gamma_l)^2 + \left(1 - \frac{\gamma_h}{\gamma_l + \gamma_h}\right) (1 + \gamma_h)^2 - 1 = \gamma_l \gamma_h. \quad (12.19)$$

We now assume the usual simple bailout mechanism with a subsidy rate of $1 - \psi$ when operating profit is negative. The profit function is, as before,

$$\pi = \begin{cases} x^\beta \theta - wx, & \text{if } x^\beta \theta - wx \geq 0; \\ \psi(x^\beta \theta - wx) & \text{otherwise,} \end{cases} \quad (12.20)$$

where $0 < \psi < 1$. The value of x can, in principle, fall into one of the following regions:

$$R_1: \quad 0 < x \leq \left(\frac{1 - \gamma_l}{w} \right)^{1/(1-\beta)}, \quad (12.21)$$

$$R_2: \quad \left(\frac{1 - \gamma_l}{w} \right)^{1/(1-\beta)} \leq x \leq \left(\frac{1 + \gamma_h}{w} \right)^{1/(1-\beta)} \quad (12.22)$$

and

$$R_3: \quad \left(\frac{1 + \gamma_h}{w} \right)^{1/(1-\beta)} \leq x. \quad (12.23)$$

Accordingly, the expected profit function is

$$E(\pi) = \begin{cases} x^\beta - wx, & \text{if } x \in R_1; \\ (1-p)[x^\beta(1+\gamma_h) - wx] + p\psi[x^\beta(1-\gamma_l) - wx], & \text{if } x \in R_2; \\ \psi(x^\beta - wx) & \text{if } x \in R_3, \end{cases} \quad (12.24)$$

The optimal interior solutions in the three regions are

$$\begin{aligned} x^{(1)} &= \left(\frac{\beta}{w}\right)^{1/(1-\beta)} \\ x^{(2)} &= \left(\frac{\beta}{w} \left[1 + \frac{\gamma_l \gamma_h (1-\psi)}{\gamma_l + \gamma_h \psi}\right]\right)^{1/(1-\beta)} \\ x^{(3)} &= \left(\frac{\beta}{w}\right)^{1/(1-\beta)}. \end{aligned} \quad (12.25)$$

The solutions in the three regions are valid solutions if and only if they satisfy the boundaries that define the regions. Hence, $x^{(1)}$ is a valid solution if and only if

$$\frac{\beta}{w} \leq \frac{1-\gamma_l}{w} \quad (12.26)$$

or if

$$\beta \leq 1 - \gamma_l. \quad (12.27)$$

The solution value $x^{(2)}$ is valid if and only if

$$1 - \gamma_l \leq \beta \left(1 + \frac{\gamma_l \gamma_h (1-\psi)}{\gamma_l + \gamma_h \psi}\right) \leq 1 + \gamma_h. \quad (12.28)$$

The following conclusions can be verified. (1) The right-hand inequality in (12.28) is always true. (2) There is no solution in R_3 , for that would require $\beta > 1 + \gamma_h$, which is a contradiction. (3) If there exists no solution in R_1 , there does exist a solution in R_2 and conversely. This is easily seen as follows. If there is no solution in R_1 , we must have $\beta > 1 - \gamma_l$, but that implies that the left-hand inequality of (12.28) holds. Conversely, if there is no solution in R_2 , the left-hand inequality of (12.28) must be violated (because the right-hand inequality always holds), and hence we must have $\beta < 1 - \gamma_l$.

We already know that the greater the variance of the uncertainty, the likelier it is that the Kornai effect will hold. We now examine the effect of departures from symmetry for a constant level of variance. We also note that the third moment about

the mean is $\gamma\gamma_h(\gamma - \gamma_h)$, so that if $\gamma_l > \gamma_h$, the distribution has positive skewness and conversely. A large value of γ means that there is a small probability of a large unfavorable departure from the mean and a large probability of a small positive departure. (It is tempting to attribute any effects to "skewness," but as we change γ , holding the mean of θ at 1 and its variance at a constant v , all higher moments change, not only the third.) We therefore fix $\gamma\gamma_h$ at the value v , and replace γ_h by v/γ_l . It is easy to see that as v increases, so does the region in which $x^{(2)}$ is a valid solution, as does the actual value of $x^{(2)}$. Focussing for the moment on the way in which $x^{(2)}$ changes with γ , we obtain

$$\frac{dx^{(2)}}{d\gamma_l} = \left(\frac{\beta}{w}\right)^{1/(1-\beta)} \left(\frac{1}{1-\beta}\right) \left[1 + \frac{v(1-\psi)}{\gamma_l + v\psi/\gamma_l}\right]^{\beta/(1-\beta)} (-1) \frac{v(1-\psi)}{(\gamma_l + v\psi/\gamma_l)^2} \left(1 - \frac{v\psi}{\gamma_l^2}\right)$$

which shows that changes in γ do not have an unambiguous effect on $x^{(2)}$.

It is also possible for $x^{(1)}$ and $x^{(2)}$ both to be valid solutions. In that case the correct solution is the one that yields higher expected profit. In the intersection of R_1 and R_2 , the boundary is given by

$$0 = E(\pi_1 | x^{(1)}) - E(\pi_2 | x^{(2)}) = \frac{1}{\beta} - 1 - \phi^{1/(1-\beta)} \left[\frac{1}{\beta\phi} \lambda_1 - \lambda_2 \right], \quad (12.29)$$

where

$$\lambda_1 = \frac{\gamma_l + v + \psi v/\gamma_l - \psi v}{\gamma_l + v/\gamma_l}$$

$$\lambda_2 = \frac{\gamma_l + \psi v/\gamma_l}{\gamma_l + v/\gamma_l}.$$

The boundaries are shown in figures 12.5, 12.6, 12.7, and 12.8 for different values of the parameters. In each diagram the following is the case: (1) All points below the highest boundary line are valid solutions in R_1 ; (2) All point above the lowest boundary are valid solutions in R_2 ; The middle boundary has the property that below it $E(\pi_2 | x^{(2)}) < E(\pi_1 | x^{(1)})$ and above it the converse holds. For $\psi = 0.5$, the shapes of these regions are rather like in the symmetric case (except, of course, that an increase in γ diminishes the output shock, and hence the boundaries have (predominantly) negative slope. But for small values of ψ the shape of R_2 is quite different from that in the symmetric case. It is easy to see that there are some values β , say 0.2 in figure 12.6, which have the property that at high values of γ only $x^{(2)}$ is a valid solution, for somewhat smaller values of γ $x^{(1)}$ also becomes a valid solution but $E(\pi_2 | x^{(2)})$ is

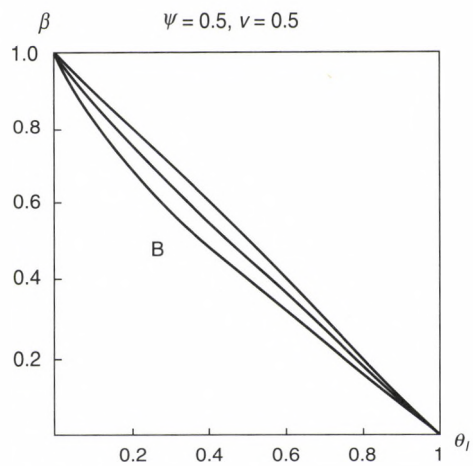


Figure 12.5

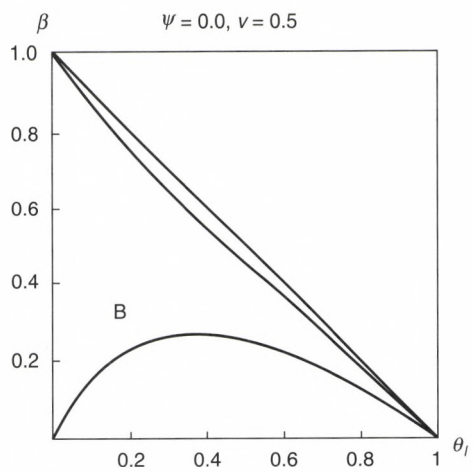


Figure 12.6

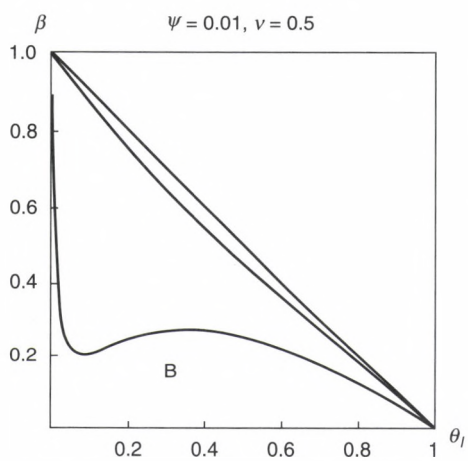


Figure 12.7

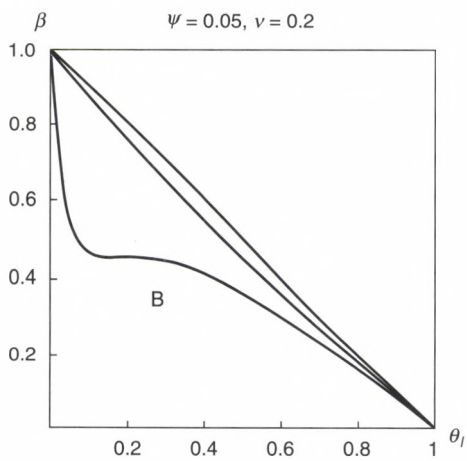


Figure 12.8

larger than $E(\pi_1|x^{(1)})$, for still smaller values of γ this profit relationship is reversed, for still smaller values of γ $x^{(2)}$ is no longer a valid solution, and for even smaller values of γ $x^{(2)}$ again becomes a valid solution. A comparison with the symmetric case is also instructive. If one were to pick random values of β and θ in the symmetric case or random values of β and γ in the asymmetric case, it would turn out in both cases that half of all cases obtained are ones in which $x^{(1)}$ is a valid solution. In the symmetric case, with $\psi = 0$ (the most favorable case for generating valid $x^{(2)}$ solutions), slightly more than 50 percent of the cases would lead to a valid solution. But in the asymmetric case, substantially greater number of cases generate valid $x^{(2)}$ s.

A final comparison is between the output levels corresponding to different γ values and several values of ψ and of the variance v . These are shown in figures 12.9 and 12.10. In each case the horizontal line corresponds to the output level $x^{(1)}$. The greater the variance and the smaller the value of ψ , the higher is output. But output is not monotone in γ ; it increases at first, reaches a maximum, and for even higher values of γ it declines.

Finally, we examine the mean-variance trade-off as in the second section. In the present instance it does not appear to be possible to obtain closed-form expressions for the relevant densities and for the mean and the variance. Hence, the relevant loci were determined by simulation. For each combination of β , v , and w , we selected values of ψ from 0.0 to 0.99 in increments of 0.01, and for each of these 100 values of ψ we generated 20,000 random γ s; for each of these we determined the region

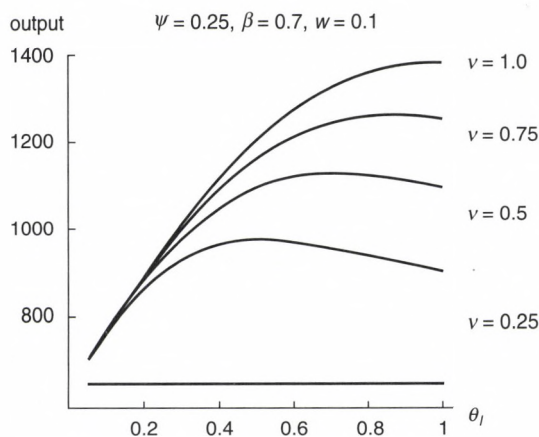


Figure 12.9

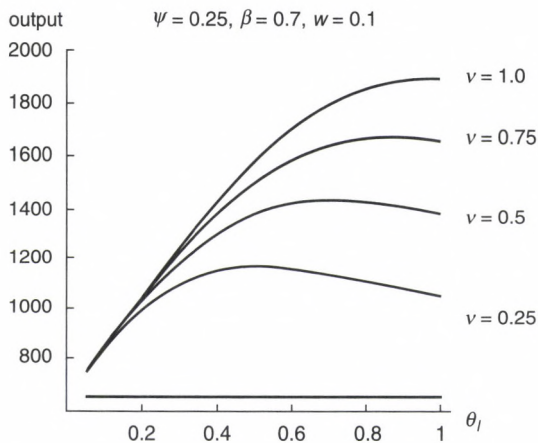


Figure 12.10

in which the solution lies and the value of the solution. From these we computed the sample mean and variance of output for each of the ψ -values. A different ψ corresponds to every point on these loci shown in figures 12.11 and 12.12. It needs to be pointed out that these are not as smooth as the corresponding curves in the second section, because of sampling variation. In both figures we display the mean-variance locus for $\beta = 0.50$ and 0.52 respectively. In figure 12.11, $v = 0.2$ and in figure 12.12 $v = 0.5$. We again approximately superimpose the central authorities' straight-line indifference curves tangent to the loci. It turns out from examining the ψ -values corresponding to the tangencies that they are nearly the same for the relatively higher and lower productivity firms. Although lower productivity always implies a lower expected output, it may, however, be accompanied by either lower or higher output variance.

4 The Whining Model

The Basic Model

We now return to the model of the second section, with the additional possibility that the manager can hire one unit of specialized managerial labor, at a cost of c . The purpose of this specialized labor is to influence the central authorities so as to secure more favorable bailout treatment for the firm; hence we call this the case of "whining." If no such labor is hired, the value of the "bailout parameter" is ψ_0 ; if

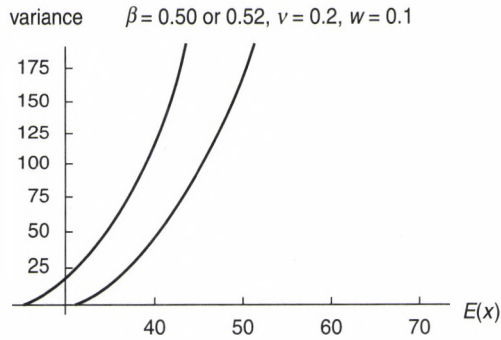


Figure 12.11

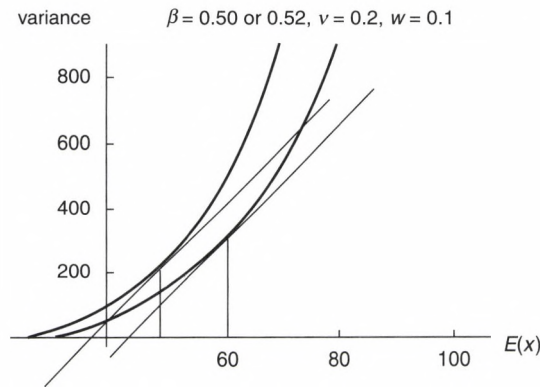


Figure 12.12

such labor is hired, its value is $\psi_1 < \psi_0$. If the firm hires no specialized labor, its expected-profit function is (12.3); if it does hire such labor, it is

$$E(\pi) = \begin{cases} x^\beta - wx - c, & \text{if } x^\beta \underline{\theta} - wx - c \geq 0; \\ & \text{if } x^\beta \bar{\theta} - wx - c \geq 0 \\ 0.5(x^\beta \bar{\theta} - wx - c) + 0.5\psi_1(x^\beta \underline{\theta} - wx - c) \text{ and} & \\ \psi_1(x^\beta - wx - c). & \text{if } x^\beta \underline{\theta} - wx - c < 0; \\ & \text{if } x^\beta \bar{\theta} - wx - c < 0. \end{cases} \quad (12.30)$$

This formulation assumes that the manager must commit to hiring specialized labor *before* production uncertainty is resolved. The interior solutions to the expected profit maximizing problem are obviously the same as for (12.3) (with ψ —or in the present case ψ_0 —replaced by ψ_1). Although in the model of the second section there cannot be a valid solution in R_3 , this is not the case here, although it is the case that there cannot be an interior solution simultaneously in regions 1 and 3. There is also the further possibility that neither region 1 nor region 3 contains a valid solution. Consider the simplified case in which $w = \beta$. Then the region 1 and region 3 solutions are both equal to 1 and the region 1 solution is not valid if $\underline{\theta} - \beta - c < 0$, whereas the region 3 solution is not valid if $\bar{\theta} - \beta - c > 0$. Both will hold simultaneously when $\underline{\theta} < \beta + c < \bar{\theta}$, which can occur for suitable parameter values. Two of the boundaries are given by

$$\underline{\theta} - \beta - c = 0 \quad (12.31)$$

and

$$2 - \underline{\theta} - \beta - c = 0, \quad (12.32)$$

while the boundaries corresponding to region 2 are obtained by substituting (12.5) (in which ψ_1 has replaced ψ) into the middle inequalities in (12.30) and are

$$\left(\frac{2 - \underline{\theta} + \psi_1 \underline{\theta}}{1 + \psi_1} \right)^{\beta/(1-\beta)} (2 - \underline{\theta}) + \beta \left(\frac{2 - \underline{\theta} + \psi_1 \underline{\theta}}{1 + \psi_1} \right)^{1/(1-\beta)} - c = 0 \quad (12.33)$$

and

$$\left(\frac{2 - \underline{\theta} + \psi_1 \underline{\theta}}{1 + \psi_1} \right)^{\beta/(1-\beta)} \underline{\theta} + \beta \left(\frac{2 - \underline{\theta} + \psi_1 \underline{\theta}}{1 + \psi_1} \right)^{1/(1-\beta)} - c = 0 \quad (12.34)$$

The regions in parameter space in which the various types of valid solutions hold are considerably more complicated than in the previous cases. This is illustrated in figure 12.13, which displays the curves $x^\beta \underline{\theta}$, $x^\beta \bar{\theta}$ and the line $wx + c$. It follows immediately from (12.30) that the interval between the points of intersection of $wx + c$ with $x^\beta \underline{\theta}$ give region 1, the two intervals given by, first, the leftmost intersections of $wx + c$ with $x^\beta \bar{\theta}$ and $x^\beta \underline{\theta}$, and second, by the rightmost intersections, give region 2, and finally the regions outside the leftmost and rightmost intersections give region 3.

We should note that in Goldfeld and Quandt (1993) it was a convenient simplification to assume throughout the analysis that $w = \beta$. Although this assumption clearly affects the solution values in the various regimes, it was proved that

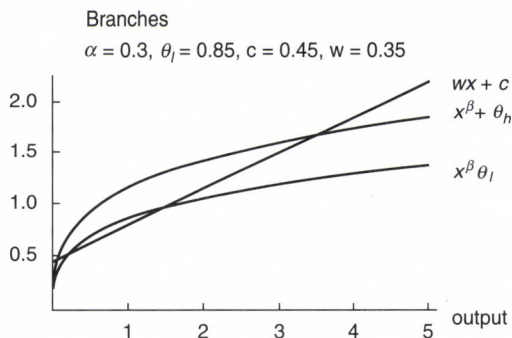


Figure 12.13

changes of w relative to β do not affect which interior solution is valid in the various regimes and which one corresponds to the highest expected profit level. Hence, statements concerning the relative frequency of the Kornai effect hold irrespective of the value of w . It is clear that a change in w , which alters the slope of the straight line, can affect which of these regions exists at all; thus, as w increases, first region 1 disappears and eventually region 2 disappears. Hence, except for illustrative purposes, w needs to be varied independently in the analysis.

Computational Results

For various values of β , θ , ψ_0 , ψ_1 , c , and w , we computed the optimal solution to the model. This involved first obtaining the optimum when no specialized labor is employed (equation (12.3)), then when such specialized labor is employed (equation (12.30)), and finally choosing as the solution the one yielding higher expected profit. The ranges of parameter values over which the computations were performed in the basic computations were as follows: θ : (0.1–0.99); β : (0.1–0.99); c : (0.1–1.0); ψ_0 : (0.1–0.9); w : (0.1–0.99). Each of these ranges was further subdivided and computations were performed for ten equidistant values of the parameters in subinterval. Hence, in this basic set of computations 10^5 separate optimality calculations were performed. The value of ψ_1 was set alternately at $0.2\psi_0$, $0.5\psi_0$, and $0.8\psi_0$. One additional case was examined in which the range of c was (1.1–2.0) and the value of $\psi_1 = 0.2\psi_0$. The overall outcomes are shown in table 12.1.

Clearly, the nonoccurrence of the Kornai effect with no z -type labor being used is invariant with respect to variations in c and in ψ_1/ψ_0 . For the (0.1–1.0) range of c , there are some instances in which there is no Kornai effect, yet some z is being employed. In general, however, the lower is ψ_1/ψ_0 , the more prevalent the Kornai

Table 12.1
Overall results

	$c: (0.1 - 1.0)$ $\psi_1 = 0.2\psi_0$	$c: (0.1 - 1.0)$ $\psi_1 = 0.5\psi_0$	$c: (0.1 - 1.0)$ $\psi_1 = 0.8\psi_0$	$c: (1.1 - 2.0)$ $\psi_1 = 0.2\psi_0$
KE yes, z yes	6,427	2,988	789	2,528
KE yes, z no	48,313	51,895	54,191	51,282
KE no, z yes	640	497	400	1,570
KE no, z no	44,620	44,620	44,620	44,620

effect is, as one would expect, because that makes the investment of z -type labor more profitable. The upward shift in the range of c reduces the prevalence of the Kornai effect slightly.

The influence of the various parameters on the prevalence of the Kornai effect on z -usage can be examined most easily in terms of the fraction of cases in which the Kornai effect holds or in which z -usage is positive for given values of some pair of parameters. Corresponding to fixing a pair of parameters such as β and θ at particular values gives 1000 cases to be examined, due to the variations of the other three parameters; one can then compute the fraction of such cases in which the Kornai effect occurs or in which z -usage is positive. For the case in which β and θ have particular alternative values, the corresponding fractions are displayed in tables 12.2 and 12.3.

It is clear from table 12.2 that as the value of θ increases, i.e., as the variance of θ decreases, the prevalence of the Kornai effect generally becomes less pronounced and as β increases, that is, the intrinsic productivity of the procurement process increases, so does generally the prevalence of the Kornai effect. However, these effects are not completely monotonic and some slight reversals can be noted near the bottom of table 12.2. Table 12.3 shows similar effects with respect to the usage of z , with again one reversal of the pattern in the last row.

Analogous tables could be exhibited in which we could show for particular values of parameters other than β and θ the results of aggregating over the variation in the remaining parameters. The general conclusion of examining such tables is that the prevalence of the Kornai effect does not depend significantly on variations in c , w , and ψ_0 (within the range of variation examined), but the employment of the z -type labor does. Thus, for example, holding β constant, varying c , and aggregating over the remaining parameter variations shows that the employment of z falls with increasing c ; the proportionate rate of decline being smaller for larger values of β . This is to be expected, because high productivity tends to enhance the option value

Table 12.2

Prevalence of the Kornai effect

$\theta \rightarrow$	0.199	0.397	0.594	0.792	0.990
β					
0.199	1.000	0.0	0.0	0.0	0.0
0.397	1.000	1.000	0.090	0.0	0.0
0.594	1.000	1.000	1.000	0.090	0.0
0.792	0.990	0.993	1.000	1.000	0.0
0.990	0.980	0.960	0.940	0.920	1.000

Table 12.3Prevalence of z -usage

$\theta \rightarrow$	0.199	0.397	0.594	0.792	0.990
β					
0.199	0.0	0.0	0.0	0.0	0.0
0.397	0.025	0.0	0.0	0.0	0.0
0.594	0.113	0.039	0.0	0.0	0.0
0.792	0.366	0.227	0.057	0.0	0.0
0.990	0.290	0.250	0.210	0.160	0.0

of being bailed out. Increases in ψ_0 increase the employment of z for given β and ψ_1/ψ_0 values, in that the absolute discrepancy between ψ_1 and ψ_0 increases with increasing ψ_0 , making the employment of more z profitable. Holding β constant and increasing w first increases the employment of z and eventually decreases it, with the peak utilization of z occurring at higher and higher w values as β becomes larger and larger. The intuition here is the following: as w rises, production becomes less profitable, and so for a while it makes sense to take advantage of the lower ψ -value that the employment of z guarantees. However, eventually the higher bailout cannot adequately compensate for the rising w , and z -usage shrinks. It is also clear that the greater the intrinsic productivity of the process, the higher will be the w value at which z -usage must begin to shrink.

We finally examine some results as the ψ_0/ψ_1 ratio changes. It can be seen from table 12.4 that there is a substantial sensitivity to the values of that ratio and the employment of z declines drastically as the benefit of employing such labor is reduced.

Table 12.4
Prevalence of z -usage for $\theta = 0.1$

$\psi_1/\psi_0 \rightarrow$	0.2	0.5	0.8
β			
0.199	0.001	0.0	0.0
0.397	0.044	0.010	0.0
0.594	0.160	0.046	0.0
0.792	0.412	0.200	0.022
0.990	0.300	0.220	0.120

5 Conclusions

We have investigated several simple models in which revenue uncertainty exists and bailouts are obtained by the firm when losses occur. The bailouts are proportional to the magnitude of the loss, but in some cases the firm can secure more favorable treatment (i.e., receive a higher reimbursement for losses) by expending some managerial resources on lobbying.

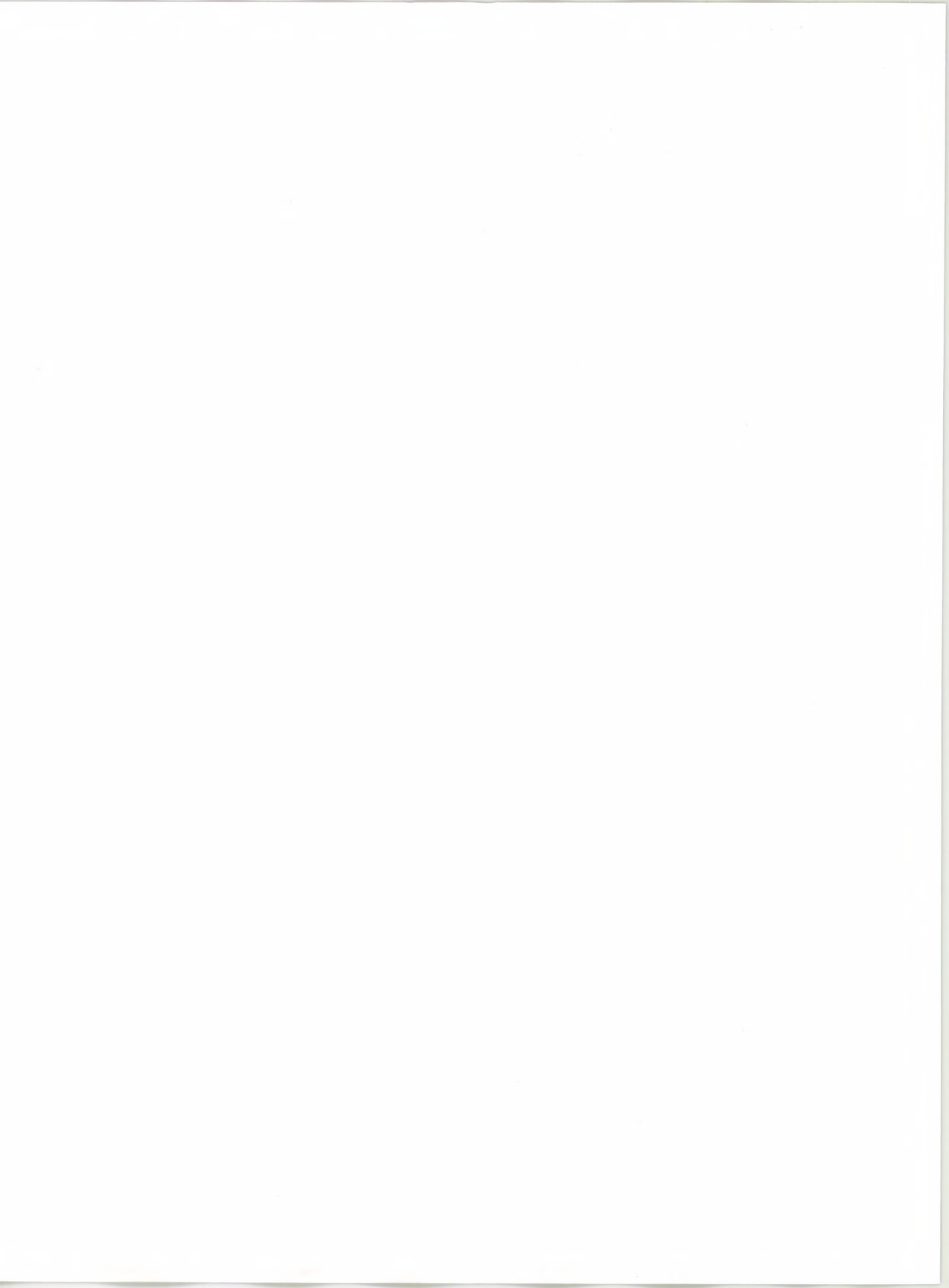
In the case of both symmetric and asymmetric error distributions, we can derive the mean-variance locus of output, which permits optimal bailout rates to be chosen by the central authorities. The parameter regions in which the Kornai effect occurs are much more complicated when (1) the distribution is asymmetric, and (2) whining is allowed.

The model exhibits a diversity of solutions, including multiple equilibria. The Kornai effect holds only in a subset of possible parameter values. Generally, the Kornai effect holds when the revenue uncertainty is large and the productivity of the production process is high. These are the circumstances in which the option value of having bailouts available is sizeable. Interestingly, the relative frequency with which the "unproductive" lobbying labor is employed is relatively small, which stands in contrast to the case in which θ is a continuous random variable, as in Goldfeld and Quandt (1988). Also interestingly, the prevalence of the Kornai effect is almost completely unaffected by *how much more favorable* the firm's treatment is after lobbying, that is, by the ψ_1/ψ_0 ratio; this ratio affects predominantly only the relative frequency with which lobbying labor will be employed, with the expected result that the more this ratio departs from unity, the more such labor will be used.

References

- Goldfeld, S. M., and R. E. Quandt (1988). Budget constraints, bailouts and the firm under central planning. *Journal of Comparative Economics* 12: 502–520.
- Goldfeld, S. M., and R. E. Quandt (1990a). Output targets, the soft budget constraint, and the firm under central planning. *Journal of Economic Behavior and Optimization* 14: 205–222.
- Goldfeld, S. M., and R. E. Quandt (1990b). Rationing, defective inputs and Bayesian updates under central planning. *Economics of Planning* 23: 161–174.
- Goldfeld, S. M., and R. E. Quandt (1992). Effects of bailouts, taxes, and risk aversion on the enterprise. *Journal of Comparative Economics* 16: 150–167.
- Goldfeld, S. M., and R. E. Quandt (1993). Uncertainty, bailouts, and the Kornai effect. *Economics Letters* 41: 113–119.
- Kornai, J. (1979). Resource-constrained versus demand-constrained. *Econometrica* 47: 801–820.
- Kornai, J. (1980). *The Economics of Shortage*. Amsterdam: North-Holland.
- Kornai, J. (1982). *Growth, Shortage and Efficiency*. Oxford: Blackwell.
- Kornai, J. (1986). The soft budget constraint. *Kyklos* 39: 3–30.
- Magee, K. L. (1993). Partial Subsidies, Taxes and the Kornai Effect. BA thesis, Princeton University.

III THE TRANSNATIONAL EXPERIENCE



13 **Determinants of Productivity in Central Europe during Economic Transition**

Sharad Bhandari and David M. Kemme

1 Introduction

The economic transition in Central Europe is critically dependent upon the development of small and medium-sized private enterprises.¹ In the Czech and Slovak Federal Republic (CSFR), Hungary, and Poland, the private small and medium-sized enterprise (SME) sector has grown from an insignificant share of the economy to become a major contributor to economic performance.² However, the behavior and performance of the SME sector during transition is only gradually being understood, inasmuch as the process of transition from socialist to market-based economies is a relatively recent phenomenon. In this chapter we examine a sample of private SMEs from a carefully defined population of manufacturing firms, and assess the causes of productivity differences in the three Central European countries early in the transition.

In the next section, we set out hypotheses on why productivity may vary among firms within countries in our sample. The hypotheses account for some unique characteristics of transition economies that may affect firms' performance, and are also based on previous research on the productivity of manufacturing firms. In the third section we briefly describe the data, sampling process, and the sample properties. In the fourth section we specify the empirical production function that we estimate for SMEs in each country and discuss the estimation procedures and results. The last section presents a summary with policy implications.

2 Hypotheses

We analyze the relationship between total factor productivity of firms and their characteristics.³ As a maintained hypothesis, we assume that each firm attempts to maximize the level of output given a set of inputs and the underlying technology for three reasons. First, this is less stringent than assuming profit maximization, which requires that firms strive for allocative and scale efficiency in addition to technical efficiency. Second, small businesses are typically reluctant to give accurate responses to questions regarding profits. And third, there is also some evidence that maximization of profits was not the only, or even the most important, goal of firms in our sample. Only 23 percent of the Polish entrepreneurs who took part in the survey cited profit-making as their main objective. Most of them listed independence (39 percent) and achievement (24 percent) as their primary goals (Webster

1993b, p. 14). The motivation of entrepreneurs in Hungary and the CSFR was similar (Webster 1993a, p. 16; Webster and Swanson 1993, p. 17).⁴

Thus, the underlying technology of a representative firm in each country can be approximated by the production function:

$$y = F(f(x), z), \quad (13.1)$$

where y is the output of the firm, x is a vector of factor inputs, with $f(x)$ specified in the fourth section as a translog production function, and z is a vector of other characteristics of the sample firms which may influence productivity.

Numerous factors may account for differences in productivity among firms operating in similar environments. Explanations by researchers have emphasized organizational norms, access to and use of technology, as well as structural and institutional factors.⁵ Entrepreneurs' perception of their goals, their personal characteristics, and a description of sample firms' constraints and potential have already been examined in previous studies.⁶ We focus on those determinants of productivity that may shed some light on the behavior of firms specifically in transition economies, and others that have been the subject of past analyses of manufacturing firms, particularly in developing countries. These determinants are: whether the firm was independent when it started production, the background of the entrepreneur (that is, whether the entrepreneur had prior business experience, whether he was an "insider" or an "outsider" with neither prior experience nor privileged access to inputs, customers, and so on), whether the firm's main customer was the state sector or other private enterprises, and whether the firm's main competitor was the state sector or other private firms.

Private firms in transition economies may be broadly dichotomized as: (1) those that are offshoots of former state enterprises or cooperatives, and (2) those that were created new, and independent of any other organizational form. We distinguish between these two types based on firms' responses to questions regarding their ownership history and the date on which they started production. To be classified as new and independent, firms had to meet two criteria. First, they had to be fully privately owned and legally unconnected to a cooperative and/or a state-owned entity since startup. Second, they had to have started production after the liberalization law permitting private sector activity was implemented.⁷ This second criterion excludes private firms that operated in the socialist era when the constraints and opportunities facing them were significantly different from those during transition.

The performance of new and independent firms is important because they are most likely to characterize the new wave of entrepreneurship during transition. Furthermore, they may perform differently from privatized state enterprises or descen-

dants of cooperatives. The difference can arise from the relatively wider *ex ante* range of choice in terms of product, technology and inputs that independent startups possess relative to privatized state firms that are, at least in the short run, dependent on their existing structures. This may be particularly salient during the initial period of the transition process. Webster (1993, p. 23) contends that entrepreneurs of firms that started out fresh, with no ownership ties to the state sector, were able to employ more appropriate technology and mix of inputs after prices were liberalized and thus were spared the potential restructuring due to large changes in relative prices. We would expect independent startups to be smaller, less dependent on any one customer for their markets, and more productive than offshoots of former state firms and cooperatives during the initial stages of economic transition.

Another attribute of transition economies which may account for variations in firms' productivity is the background of the entrepreneur. We classify this into three categories: pre-reform entrepreneurs, insiders, and outsiders.⁸ Pre-reform entrepreneurs are defined as those who derived a significant portion of their income from private business prior to the reforms. We would expect a positive association between prior business experience and productivity.⁹ An insider is defined as anyone who held a high-level position in the party or the government. For Poland, it also includes anyone who had access to long-term financing. In Hungary and the CSFR, it also encompasses those entrepreneurs whose firms were created in toto from a large state-owned company, provided that the newly created firm had more than 25 employees at the start of production. These people may be the managerial elite and have the contacts and the skills and talents necessary to manage small and medium businesses. Thus their firms should be more productive. Outsiders are defined as those with neither prior business experience nor privileged access to finance and inputs.

The effect of having state owned enterprises (SOEs) as main customers on firms' productivity could be either positive or negative. Webster (1993) suggests that having SOEs as major customers is advantageous in that sales contracts tend to be large, but disadvantageous in that payments by SOEs tend to be late and, further, contracts would decline as the state sector is dismantled. A firm's productivity also depends on its main competitor. If, for example, SOEs are a firm's main competitors, the pressure to improve performance may be weaker than if the firm were competing with private or foreign firms.

In order to isolate these characteristics from other relevant factors, we control for the size of firms, the level of education of the managers, the age of equipment, the number of competitors reported by firms in their main product market, and the market-orientation of firms. Not only do these variables serve as controls, but they

are also interesting in their own right. Because independent startups tend to be smaller than offshoots of former state firms or cooperatives, the differences in observed productivity could be the result of differences in the size of firms. In particular, larger firms could benefit from scale economies in production and distribution, and thus be able to realize a higher level of output for given levels of inputs.¹⁰ Similarly, educated managers are better able to organize production, seek out new markets, and increase productivity.¹¹ Also, although we have represented the technology of firms by equation (13.1), in reality, firms use different vintages of technology. Firms employing newer technology should exhibit higher levels of productivity. The number of competitors in a firm's product market is also relevant for its productivity. For a given size of the market, a larger number of competitors would exert more intense pressure on a firm to improve its productivity. Similarly, holding other factors constant, firms producing for export may be more efficient than those producing solely for domestic markets because of international competitive pressures.¹² Arguments may also be made for differences in productivity across industries within countries. Such variations may arise from, for example, differences in the ease of entry of firms and in technology, or because of the country's comparative advantage in one sector rather than another.¹³

3 The Data and Sample Properties

We employ micro-level data to estimate the production functions for each country. Data for each country are taken from three firm-level surveys conducted by World Bank researchers in May 1991 in Poland, in September 1991 in Hungary, and in February 1992 in the CSFR. The data set consists of a randomly selected sample from a population of firms which met certain criteria.¹⁴ The firms selected cover a broad spectrum of manufacturing activities and are quite dispersed geographically. At the time of the surveys, the samples represented about 15 percent of the population of registered, domestically owned private manufacturers, excluding sole proprietorships, for the CSFR, 1.7 percent for Hungary, and 1.6 percent for Poland.

All three countries started extensive reform of legal and economic policies at least 18 months prior to the surveys. As a result, some effects of policy changes were expected to be present in the questionnaire responses. It should be emphasized that the surveys provide a snapshot of the operation of firms at a particular point in time during the transition, and repeat surveys would be necessary to reach a better understanding of their performance.

In the empirical analysis, we use total sales per month in U.S. dollars for output Y .¹⁵ Inputs include: (1) total labor, X_1 , in full-time equivalents; (2) cost of raw mate-

rials and intermediate inputs per month in U.S. dollars, X_2 ; (3) floor space in square meters, X_3 ; and (4) cost of energy per month in U.S. dollars, X_4 .¹⁶

The sample mean, median, and standard deviation for each variable are reported for each country in table 13.1. There are clear differences among countries with respect to overall output and usage of factor inputs. Hungarian firms produce much larger levels of output with more employees, raw materials, and energy. They use significantly less floor space, however.¹⁷ In terms of the average measures of output and inputs, the firms in the CSFR lie between Polish and Hungarian firms.

In the regression analysis, we control for sector-specific fixed effects by including a vector of dummy variables, D_{sj} . These are $D_{s1} = 1$ for knitting, $D_{s2} = 1$ for clothing, $D_{s3} = 1$ for metal products, $D_{s4} = 1$ for machinery, $D_{s5} = 1$ for the "other" branches of industry.¹⁸ The proportion of sample firms in each of the six industries is given by the mean of the corresponding dummy variable. Among the five well-defined industries, metal products accounted for a large number of firms in all three countries. The other important industries, in terms of the number of firms in the sample, were clothing and machinery in the CSFR, machinery and plastics in Hungary, and knitting and clothing in Poland.

Characteristics which may be important in transition economies are also proxied by a vector of dummy variables, with $D_{i1} = 1$ if the firm is a new and independent start-up, $D_{i2} = 1$ if the entrepreneur was an insider, as defined earlier, $D_{i3} = 1$ if the owner/manager was a pretransition entrepreneur, $D_{i4} = 1$ if the firm's main customers were SOEs, and $D_{i5} = 1$ if the firm's main competitors were SOEs. Table 13.1 indicates that the proportion of independent startups in the sample is significantly larger in the CSFR (42 percent) and in Poland (39 percent) than in Hungary (17 percent). The independent start-ups tended to be smaller. The average number of workers in these firms was 22 in the CSFR, 26 in Hungary, and 24 in Poland. The corresponding averages in firms that were offshoots of state firms or cooperatives were 45 in the CSFR, 50 in Hungary, and 41 in Poland.

The differences across countries in terms of the background of sample entrepreneurs are large. Whereas the proportion of insiders ranged only between 19 percent in Hungary and 27 percent in the CSFR, the proportion of entrepreneurs with pre-reform business experience varied between 10 percent in the CSFR and 57 percent in Hungary. Conversely, the proportion of outsiders was much larger in the CSFR (64 percent) than in Poland (40 percent) or in Hungary (24 percent). The proportion of firms with SOEs as main customers was similar in all three countries. It ranged from 43 percent in the CSFR to 44 percent in Hungary and Poland. But the proportion of firms with SOEs as their main competitors ranged from 25 percent in Poland to 58 percent in the CSFR.

Table 13.1
Sample statistics

	CSFR	Hungary	Poland	Pooled
Output and inputs				
Sales, monthly (US\$)	60,190.06 22,962.96 (114,929.60)	118,763.40 42,567.57 (366,322.59)	49,093.56 21,482.28 (79,788.30)	77,839.33 27,027.03 (235,468.53)
X_1 : Labor (full time equivalent)	35.27 21.13 (42.24)	45.54 29.00 (63.36)	34.49 18.50 (49.45)	38.74 21.75 (53.02)
X_2 : Materials (US\$)	44,825.32 8,888.89 (174,564.00)	129,994.51 20,506.76 (348,569.31)	40,675.84 9,485.65 (88,151.45)	74,382.45 11,486.49 (239,709.58)
X_3 : Floor space (m ²)	2,529.61 800.00 (6,312.36)	1,210.86 580.00 (1,758.11)	3,705.19 500.00 (12,528.03)	2,433.02 580.00 (8,065.82)
X_4 : Energy (US\$)	1,655.96 740.74 (2,946.49)	1,860.29 783.78 (3,517.00)	456.33 120.60 (1,139.52)	1,340.35 385.91 (2,818.40)
Sectoral breakdown				
D_{s1} : knitting	0.095	0.058	0.169	0.105
D_{s2} : clothing	0.162	0.116	0.169	0.148
D_{s3} : metal products	0.162	0.221	0.182	0.190
D_{s4} : machinery	0.122	0.174	0.143	0.148
D_{s5} : other	0.351	0.267	0.260	0.291
D_{s6} : plastics	0.108	0.163	0.078	0.118
Firm characteristics				
D_{n1} : new, independent startups	0.419	0.174	0.390	0.321
D_{n2} : insider	0.270	0.186	0.208	0.219
outsider	0.635	0.244	0.402	0.418
D_{n3} : prior entrepreneurship	0.095	0.570	0.390	0.363
D_{n4} : SOE customers	0.432	0.442	0.442	0.439
D_{n5} : SOE competition	0.581	0.314	0.247	0.376
D_{d1} : small firm (≤ 10 employees)	0.189	0.128	0.260	0.190
medium firm (11–100 employees)	0.757	0.791	0.688	0.747
D_{d2} : large firm (> 100 employees)	0.054	0.081	0.052	0.063
D_{d3} : college education	0.378	0.593	0.740	0.574
D_{d4} : age of equipment (3 years or less)	0.284	0.209	0.234	0.241
D_{d5} : < 10 competitors	0.405	0.465	0.429	0.435
PEXP: exports (%)	39.77 42.50 (31.99)	46.27 40.00 (34.88)	48.55 50.00 (34.56)	44.78 45.00 (33.69)
No. of exporters	26	37	20	83
N : number of firms	74	86	77	237

Note: For continuous variables the first entry is the mean, the second is the median, and the standard deviation is in parentheses. The mean is given for dummy variables.

The size distribution of firms in the sample, measured by the number of full-time employees, is not too different across countries: 5–8 percent of the sample firms had more than 100 employees, 13–26 percent of the firms had fewer than 11 employees, and the bulk of the firms in each country had between 11 and 100 employees. Furthermore, a significantly larger proportion of Polish owner/managers had more education, and a somewhat smaller percentage of Hungarian firms had relatively new equipment. The proportion of firms with fewer than 10 competitors—41 percent in the CSFR, 43 percent in Poland, and 47 percent in Hungary—did not vary much across countries. There were more Hungarian firms engaged in exporting, but the percent of sales exported was highest for Polish firms. These attributes are represented by D_{ds} , where $D_{d1} = 1$ if the firm had fewer than 11 employees, $D_{d2} = 1$ if the firm had more than 100 employees, $D_{d3} = 1$ if the entrepreneur had at least the equivalent of an undergraduate college degree, $D_{d4} = 1$ if the average age of equipment was less than three years old, $D_{d5} = 1$ if the firm had fewer than 10 competitors, and by PEXP, the percentage of sales exported.

4 Estimation Results

Separate production functions for the three countries were estimated in order to allow for probable differences in the environments of firms with respect to product demand, input supply, and access to technology. Because Hungary was generally regarded as somewhat further along in the reform process at the time the data were collected, Hungarian firms may perform somewhat differently than, for instance, their CSFR counterparts. Similarly, firms in the CSFR, where the transition had just begun, may function differently than their Polish counterparts that operated in an environment of shock therapy.¹⁹ Furthermore, if there were any biases in the exchange rates used to convert the local-currency value of output and inputs into U.S. dollar equivalents, results from regressions using a pooled sample across countries, by industry, for example, could be biased.

For each country, we first estimate a transcendental logarithmic production function which imposes relatively few a priori restrictions on the underlying technology.²⁰ The stochastic specification is:

$$\ln(Y) = \ln(\alpha_0) + \sum_{i=1}^k \beta_i \ln(X_i) + \frac{1}{2} \sum_{i=1}^k \sum_{j=1}^k \gamma_{ij} \ln(X_i) \ln(X_j) + \delta Z + \varepsilon \dots \quad (13.2)$$

where Y is output of a firm, X_i are inputs, Z is a vector of firm characteristics, and ε is a random error term. The parameters, $\ln(\alpha_0)$, which denote the average level of productivity, β_i , γ_{ij} ($= \gamma_{ji}$), and the vector δ are to be estimated.

In each case, we checked for the presence of heteroscedasticity using White's (1980) test. The results from these tests, presented in table 13.2, showed that in all three countries, the null hypothesis of homoskedastic residuals could not be rejected. This is particularly reassuring since the null hypothesis maintains not only that the errors are homoskedastic, but also that the model is correctly specified, that is, the errors are independent of the regressors (White 1980, p. 823). We then tested the null hypothesis that the underlying technology can be represented by the more parsimonious Cobb-Douglas specification, which in this case is equivalent to testing $H1_0: \gamma_{ij} = 0 \forall i, j$. This hypothesis could not be rejected for the three countries. The F -statistic from this test is also presented in table 13.2.

The parameter estimates for the Cobb-Douglas specification are presented in table 13.3. We also report White's χ^2 statistic, the adjusted R^2 , and the F -statistic for the null hypothesis, $H2_0$, that the production functions exhibit constant returns to scale. Based on White's χ^2 test, there is no evidence of heteroskedasticity. The adjusted R^2 statistics reveal that the independent variables explain between 65 and 73 percent of the variation in output in the three countries. These are reasonably high proportions. Furthermore, in all three countries, the null hypothesis that firms are operating under constant returns to scale within the range of sample output cannot be rejected.

The parameter estimates for the four factor inputs are all within economically meaningful ranges. In all three countries, the coefficients for labor and raw materials are statistically significant. In Hungary and the CSFR, the coefficient estimate for floor space is also significant. The estimates indicate that Polish and the CSFR firms are considerably more labor-intensive than their Hungarian counterparts. A doubling of the labor input leads to an increase in output of 64 percent in Poland, 60 percent in the CSFR, and 36 percent in Hungary. The differences in labor-intensity across countries are likely the result of differences in wage rates. Average

Table 13.2
Translog specification diagnostic tests

	CSFR	Hungary	Poland
White's χ^2	66.0127	84.7523	75.8293
(p -value)	(0.7061)	(0.4872)	(0.4839)
d.f.	73	85	76
F -test $H1_0: \gamma_{ij} = 0$	1.1952	1.3598	1.5947
(p -value)	(0.3209)	(0.2235)	(0.1384)
d.f.	10, 43	10, 55	10, 46

monthly wages in the sample firms were substantially higher in Hungary (\$222) than in Poland (\$154) and in the CSFR (\$150). The coefficient estimate for raw materials is similar in the three countries, ranging from 25 percent in Hungary to 28 percent in Poland and the CSFR.

Table 13.3 also shows that the origin of firms—whether they started up new and independent or were derived from former SOEs or cooperatives—does not affect productivity after accounting for other relevant variables. The coefficient estimates for new firms are positive in all three countries, but they are not statistically significant at conventional levels of significance. The effect of being an insider, on the other hand, is negative and statistically significant for firms in the CSFR. Insiders, who may have had certain advantages in the capital market, apparently are not more competent when it comes to production. With respect to the advantages of prior experience, (D_{i3}), working in partnerships (VGMKs and GMKs) and agricultural cooperatives that, Webster (1993a) suggests, would contribute to increased productivity in Hungary were nonexistent. This was also true of the other two countries. It appears that business experience gained in the pre-reform era is anachronous. However, firms that had SOEs as their main customers seem to have benefited in Hungary. They were 35 percent more productive than those which sold their products mainly to other private firms. In the other two countries, the differences were not statistically significant. The type of competitor does not seem to make a difference to firms' productivity in all three countries.

Table 13.3 also indicates that the size of the firm, (D_{d1} and D_{d2}), has no effect on productivity. Neither large nor small firms are more productive than firms with 11 to 100 employees. However, productivity is positively associated with the level of education of the manager/owner, (D_{d3}) in Poland. Webster (1993b) reports that an unusually large number of Polish entrepreneurs were well educated, and 74 percent of the owner/managers in this sample had at least an undergraduate degree. Part of the explanation of the large, positive, and statistically significant coefficient estimate for education in Poland may lie in the difference in the proportion of educated Polish managers who went abroad for training and that of less educated entrepreneurs who did so. That is, educated entrepreneurs in Poland appear to be more productive because a substantially larger proportion of them, compared to the less educated ones, received foreign training. In our sample, about 19 percent of the educated Polish entrepreneurs reported that they received some kind of training abroad, whereas only 5 percent of those with less than a college degree did. In Hungary and the CSFR, 18 percent of college educated entrepreneurs went abroad for training but 12 and 10 percent, respectively, of those with less than a college degree also did so.

Table 13.3
Production function estimates by country

Specification	CSFR Cobb-Douglas	Hungary Cobb-Douglas	Poland Cobb-Douglas
Variable	Parameter estimate (<i>p</i> -value)	Parameter estimate (<i>p</i> -value)	Parameter estimate (<i>p</i> -value)
$\ln \infty_0$ (intercept)	4.210 (0.0001)	4.0947 (0.0001)	4.3650 (0.0001)
$\ln X_1$ (labor)	0.595 (0.0055)	0.3598 (0.0765)	0.6418 (0.0059)
$\ln X_2$ (raw material)	0.281 (0.0001)	0.2464 (0.0001)	0.2800 (0.0001)
$\ln X_3$ (floor space, m^2)	0.2119 (0.0106)	0.2396 (0.0997)	0.0485 (0.4666)
$\ln X_4$ (energy)	0.0032 (0.9632)	0.1218 (0.1439)	0.0614 (0.5002)
D_{s1} (knitting)	0.5334 (0.1440)	-0.2804 (0.5148)	-0.1245 (0.7820)
D_{s2} (clothing)	-0.1694 (0.6428)	-0.4650 (0.1668)	0.0309 (0.9457)
D_{s3} (metal working)	-0.2095 (0.5704)	-0.0375 (0.8910)	0.1489 (0.7124)
D_{s4} (machinery)	0.1301 (0.7125)	-0.0539 (0.8607)	0.3893 (0.3618)
D_{s5} (other)	0.0751 (0.8131)	-0.3732 (0.1661)	0.3515 (0.3733)
D_{f1} (new, independent firm)	0.1502 (0.4859)	0.3596 (0.1955)	0.2789 (0.1478)
D_{f2} (insider)	-0.3403 (0.0982)	-0.2163 (0.4517)	-0.0837 (0.7559)
D_{f3} (prior entrepreneurship)	-0.1406 (0.6504)	0.2432 (0.3129)	0.0007 (0.9973)
D_{f4} (SOE customers)	0.0276 (0.8770)	0.3448 (0.0441)	0.0321 (0.8839)
D_{f5} (SOE competition)	0.1809 (0.2913)	0.0565 (0.7719)	0.0684 (0.7655)
D_{d1} (size-small firm)	-0.4887 (0.1780)	0.2240 (0.4816)	0.0749 (0.8353)
D_{d2} (size-large firm)	-0.2837 (0.5508)	0.0338 (0.9389)	-0.0373 (0.9427)
D_{d3} (college education)	-0.0182 (0.9126)	0.2254 (0.2796)	0.4302 (0.0565)

Table 13.3 (continued)

Specification	CSFR Cobb-Douglas	Hungary Cobb-Douglas	Poland Cobb-Douglas
Variable	Parameter estimate (<i>p</i> -value)	Parameter estimate (<i>p</i> -value)	Parameter estimate (<i>p</i> -value)
D_{d4} (age of equipment)	0.1275 (0.5489)	0.5615 (0.0204)	-0.1044 (0.6352)
D_{d5} (<10 competitors)	-0.0753 (0.6728)	-0.0313 (0.8646)	0.0142 (0.9479)
PEXP (% of sales exported)	0.0064 (0.0636)	0.0058 (0.0385)	-0.0051 (0.1740)
Diagnostic statistics	CSFR	Hungary	Poland
White's χ^2 (<i>p</i> -value)	69.3087 (0.6008)	80.1475 (0.6286)	76.9744 (0.4472)
d.f.	73	85	76
Adjusted R^2	0.7269	0.6538	0.6526
<i>F</i> -test H_{20} (<i>p</i> -value)	0.2857 (0.5952)	0.0368 (0.8484)	0.0261 (0.8721)
d.f.	1, 53	1, 65	1, 56
<i>N</i>	74	86	77

Use of equipment less than three years old (D_{d4}) has a positive effect on productivity in Hungary, but not elsewhere. Equipment of younger vintage could have been more productive in Hungary if most of it was obtained from the West. However, this does not seem to be the reason for the differential effect of age of equipment in Hungary versus that in the CSFR and Poland. Almost as large a proportion of firms in Hungary using older equipment (41 percent) obtained it from the West as that using younger equipment (47 percent). In the CSFR, the correlation between age of equipment and its origin was much stronger. Only 8 percent of the firms using older equipment reported that it was Western, whereas 39 percent of the firms using younger equipment did so. In Poland, a higher proportion (27 percent) of the firms with older equipment said they obtained it from the West whereas only 18 percent of those using newer equipment did so. Therefore, if the positive effect on productivity of equipment of younger vintage in Hungary was the consequence of its originating in the West, this effect should have been even stronger for the CSFR, but it was not.

The coefficient estimates of the degree of competition, proxied by the number of competitors, is statistically insignificant, but the market orientation of firms does

seem to influence firms' productivity. Firms that export a higher proportion of sales (PEXP) tend to be more productive in the CSFR and Hungary. The coefficient estimate on exports was small but statistically significant. In Poland, exports do not seem to have any impact on productivity. Whatever the reason for this difference among the three countries, it is not the result of different geographic orientation of exports. If Hungarian and CSFR exporters sold in Western markets, for example, while Polish ones sold in the traditional Eastern countries, then this could provide a plausible explanation for the results—we would expect firms selling in mature Western markets, where product standards are generally higher, to be more competitive. However, among our sample firms, 99 percent of the exports from the CSFR and Hungary, and all of the exports from Poland were sold in Western markets.

5 Summary

Productivity may be a proxy for success in transition, and policies that enhance the productivity of firms in transition economies should be considered. Therefore, we estimate production functions in order to ascertain the determinants of total factor productivity of small and medium private manufacturing firms. Controlling for industry-specific fixed effects, the importance of several firm characteristics was determined. The effects on total factor productivity of these characteristics varied somewhat from country to country, but several results were quite robust.

The results with respect to those variables unique to transition economies are particularly interesting. First, the type of customer had an effect on firms' productivity only in Hungary. Hungarian firms that sold their products mainly to SOEs appear to have benefited from large sales contracts. Second, being an insider—having particular advantages from the pretransition period—had a negative effect on the level of total factor productivity, if any effect at all. Although we may have expected privileged access to capital or connections to government officials to have a positive effect, higher levels of productivity had not yet resulted. Third, experience as a manager/owner operating a private enterprise in the pre-transition period had no effect on the level of total factor productivity for enterprises owned or managed by the same individuals in the transition period. That is, the new environment is so different that past managerial experience from the pre-reform period is irrelevant.

Fourth, firm size offered no advantage in terms of the level of total factor productivity. Economies of scale are not significant enough to confer substantial advantage to larger firms. Very small firms were just as likely to be successful as large firms.²¹ The strategy to promote the indigenous private sector adopted in all three

countries seems to be a worthy approach for economic development in the transition economy. Fifth, the level of education appears to increase the level of total factor productivity significantly in Poland. The level of education of the population as a whole in each country is quite high. But there is some evidence that the proportionately greater tendency among educated Polish entrepreneurs to have received foreign training relative to those in the CSFR and Hungary enhanced their productivity. Whether or not these entrepreneurs can use their educational background to maintain the level of productivity or continuously bring improved products to the market place is an open question. Sixth, if the firm had a strong export orientation, the level of total factor productivity was higher. Competition from foreign firms and the emphasis on quality found in foreign markets influenced the domestic production process. Seventh, the newness of equipment had a positive effect on productivity of Hungarian firms, and this does not seem to be the consequence of different sources—Western, Eastern, or domestic—of equipment.

It is important to note that our results reflect the activities of these small and medium enterprises at a particular point in time during the transition process. As the process continues to unfold all of the sample firms will have to adjust to continuing changes in the legal environment, the level of competition and changes in demand. The small and medium enterprise sector is one of the most dynamic in all three countries and will certainly attract further study.

Notes

We would like to thank Phil Hersch for helpful suggestions and comments on an earlier draft of this chapter. All errors remain our responsibility.

1. Brock and Evans (1989) provide a review of the critical role small business provides in a market economy. Jeffrey Sachs and others have long emphasized the important role of the private sector during the transition period. See, for example, Berg and Sachs (1992), Lipton and Sachs (1990), Roman (1989), Sachs (1992).

2. McDermott and Mejstrik (1992) report that the average employment for the smallest 5 percent of firms (based on employment) in the CSFR is 900 employees, much larger than what would normally be considered a small enterprise. Basically, small firms did not exist during the communist period. Note also that our analysis uses data collected in 1992, prior to the formal dissolution of the Czech and Slovak Federation.

3. There are many reasons why profitable and productive firms may still fail during the transition period including the fact that profits may indeed be transitory. The market for some products will disappear as consumers' tastes change or the domestic comparative advantage of producers is eroded due to foreign trade. Further, financial constraints may be sufficient to bankrupt a firm during the transition period. For each country, constraints and prospects are discussed in detail in Webster (1993a, 1993b) and Webster and Swanson (1993).

4. Independence was defined as "freedom to work on one's own," and achievement as "the desire to make the best use of one's skills" (Webster 1993b, p. 14). The three criteria—independence, achievement,

and profit-making—may not be mutually exclusive, but the tendency among entrepreneurs to choose the first two over profit-making indicates that the motivation to make profits was tempered by other personal goals.

5. See Nelson (1981) for a survey of issues related to productivity.

6. See Webster (1993a, 1993b) and Webster and Swanson (1993) for qualitative analyses of the survey. Our quantitative analysis allows us to estimate the effect on productivity of each factor, after netting out the influence of other relevant factors.

7. These laws were implemented in January 1989 in Hungary and Poland, and in April 1990 in the CSFR.

8. This classification is based on Kaufman (1992).

9. Webster (1993a) also claims that pre-reform entrepreneurs in Hungary were more productive than those in the CSFR or Poland. Although these entrepreneurs tended to be craftsmen in all three countries, in Hungary, the craftsmen tended to have larger operations. In addition, the Hungarian entrepreneurs of the 1980s also included managers and members of economic working partnerships (VGMKs and GMKs) and agricultural cooperatives, organizations that did not exist, or were less prominent, in the CSFR and Poland.

10. See, for example, Page (1984), Pitt and Lee (1981) and Tyler (1979). It should be noted that the firms in our sample are still SMEs, the largest having less than 434 full-time equivalent workers.

11. See, in particular, Pack (1987), chapters 3 and 6.

12. See, for example, Chen and Tang (1987).

13. Ideally, we would have liked to estimate production functions for each industry within each country, but the small size of the sample precluded us from doing so.

14. A large sample of firms was drawn at random from official registration lists of the central statistical office in each country. Firms in this sample were contacted to verify their principal activity and their willingness to participate in the survey. Then firms, with domestic private ownership of at least 51 percent, and with seven or more employees (six for the CSFR), were selected randomly from this list. For each country, the overall sample was supplemented by a stratified sample of firms from the same population in five specific branches of industry: knitting (ISIC 321), clothing (ISIC 322), metal working (ISIC 381), machinery (ISIC 382 and 383) and plastic products (ISIC 356). The survey methodology ensures that the data are comparable across countries insofar as the firms are responding, in most cases, to similar questions. There were a total of 320 firms interviewed. These make up the complete database. Of these, 83 had missing values for one of the independent variables or for output. Also, by the time the surveys were done, some of the firms had fewer than six employees. We selected firms with at least two employees, assuming that firms with fewer than two employees cannot sustain a manufacturing activity. Thus, the total sample we employ in the empirical analysis consists of 237 firms. See Webster and Swanson (1993), p. 14, Webster (1993a), p. 16, and Webster (1993b), p. 14.

15. The U.S. dollar equivalents of inputs and sales were calculated using exchange rates prevailing at the time of the surveys. The exchange rates used were 27 korunas, 74 forints, and 9310 zloty per dollar. Insofar as these exchange rates are biased, any analyses of data, pooled across countries, could also be biased. This is one reason why we opt for a country-specific analysis.

16. The sample included data on the replacement cost of equipment as estimated by the managers. After preliminary analyses using these data, we judged them to be unreliable proxies for the measurement of capital input. Discussions with the interviewers who actually visited the enterprises confirmed the subjective nature of these data. Thus, we use energy consumption because it is a reasonably good proxy for capital utilization; see, for example, Foss (1963), Heathfield (1972), Levy (1981).

17. Nevertheless, floor space did not seem to be a constraint on Hungarian firms. When asked if their land and buildings were sufficient for their needs, 56 percent of the Hungarian entrepreneurs in the sample responded that they were, compared to 52 percent and 38 percent, respectively, in Poland and the CSFR.

18. The plastics sector was taken as the base sector.

19. For a description of the conceptual issues involved in transition strategies in each country see Kemme (1991), and Kiss (1993). For country specific discussions see Brada (1991) on the CSFR; Gomulka (1992), Lipton and Sachs (1990), Sachs (1992), and Wellisz (1991) on Poland; and Hare (1991) and Kádár (1989) on Hungary.
20. See Denny and Fuss (1977).
21. Recall these are large firms within the SME classification; the largest had 434 full time equivalent employees.

References

- Berg, Andrew, and Jeffrey Sachs (1992). Structural adjustment and international trade in Eastern Europe: the case of Poland. *Economic Policy* (April 1992): 118–173.
- Brada, Josef C. (1989). Technical progress and factor utilization in East European economic growth. *Economica* 56: 433–448.
- Brada, Josef C. (1991). The economic transition of Czechoslovakia from plan to market. *Journal of Economic Perspectives* 5(4): 171–177.
- Brock, William A., and David S. Evans (1989). Small business economics. *Small Business Economics* 1: 7–20.
- Chen, Tain-ji, and De-piao Tang (1987). Comparing technical efficiency between import-substitution-oriented and export-oriented foreign firms in a developing economy. *Journal of Development Economics* 26: 277–289.
- Denny, Michael, and Melvyn Fuss (1977). The use of approximation analysis to test for separability and the existence of consistent aggregates. *American Economic Review* 67(3): 404–418.
- Foss, F. Murray (1963). The utilization of capital equipment. *Survey of Current Business* 43: 8–16.
- Gomulka, Stanislaw (1992). Polish economic reform program, 1990–1991: principles, policies and outcomes. *Cambridge Journal of Economics* 16: 355–372.
- Hare, Paul (1991). Hungary: in transition to a market economy. *Journal of Economic Perspectives* 5(4): 195–201.
- Heathfield, David F. (1972). The Measurement of capital usage using electricity consumption data for the UK. *Journal of Royal Statistical Society* 135: 208–220.
- Kádár, Béla (1989). Market economy—Hungarian way. *Acta Oeconomica* 40(3–4): 235–238.
- Kaufman, Wallace (1992). What Role do Social Conditions Play in Selecting Successful Entrepreneurs? Mimeo.
- Kemme, David M. (1991). *Economic Transition in Eastern Europe and the Soviet Union: Issues and Strategies*. New York: Institute for East-West Studies, 1991.
- Kiss, Károly (1993). Western Prescriptions for Eastern Transition: A Comparative Analysis of the Different Schools and Issues. Working Paper of the Institute for World Economics, Hungarian Academy of Sciences, No. 18.
- Levy, Victor (1981). On estimating efficiency differentials between the public and private sectors in a developing economy—Iraq. *Journal of Comparative Economics* 5: 235–250.
- Lipton, David, and Jeffrey Sachs (1990). Creating a market economy in Eastern Europe: the case of Poland. *Brookings Papers on Economic Activity* 1: 75–133.
- McDermott, Gerald A., and Michal Mejstrik (1992). The role of small firms in the industrial development and transformation of Czechoslovakia. *Small Business Economics* 4: 179–200.
- Nelson, Richard (1981). Research on productivity growth and productivity differences: dead ends and new departures. *Journal of Economic Literature* 19: 1029–1064.

- Pack, Howard (1987). *Productivity, Technology and Industrial Development*. New York: Oxford University Press.
- Page, John M. (1984). Firm size and technical efficiency. *Journal of Development Economics* 16: 129–152.
- Pitt, Mark M., and Lung-Fei Lee (1981). The measurement and sources of technical efficiency in the Indonesian weaving industry. *Journal of Development Economics* 9: 43–64.
- Román, Zoltán (1989). The size of the small-firm sector in Hungary. *Small Business Economics* 1: 303–307.
- Rumer, Boris (1991). New capitalist in the USSR. *Challenge* (May–June): 19–22.
- Sachs, Jeffrey (1992). The economic transformation of Eastern Europe: the case of Poland. *Economics of Planning* 25(1): 5–20.
- Terrell, Katherine (1992). Productivity of western and domestic capital in Polish industry. *Journal of Comparative Economics* 16(3): 494–514.
- Tyler, William G. (1979). Technical efficiency in production in a developing country: an empirical examination of the Brazilian plastics and steel industries. *Oxford Economic Papers* 3: 477–495.
- Webster, Leila (1993a). The Emergence of Private Sector Manufacturing in Hungary: A Survey of Firms. World Bank Technical Paper 229. Washington, D.C.: World Bank.
- Webster, Leila (1993b). The Emergence of Private Sector Manufacturing in Poland: A Survey of Firms. World Bank Technical Paper 237. Washington, D.C.: World Bank.
- Webster, Leila, and Dan Swanson (1993). The Emergence of Private Sector Manufacturing in the Czech and Slovak Federal Republic: A Survey of Firms. World Bank Technical Paper 230. Washington, D.C.: World Bank.
- Wellisz, Stanislaw (1991). Poland under “Solidarity” rule. *Journal of Economic Perspectives* 5(4): 211–217.
- White, H. (1980). A heteroscedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity. *Econometrica* 48: 817–838.

14 **The Evolutionary Path away from Socialism: The Chinese Experience**

Bernard Chavance

Socialism and Capitalism in Theoretical and Historical Perspective

Understanding the shift from socialism to capitalism implies some general framework of analysis for both systems. János Kornai (1992) has proposed such a framework in a remarkable and classical book. I will suggest here another possible way of conceptualization, inspired by his problematic.

Capitalist and socialist systems share some basic features that raise problems for their comparative analysis. In both types of systems, sometimes under different names, we find exchange, prices, money, wages, capital, and accumulation of capital. Economic theories developed on each side have generally insisted on the differences or the contrast between the role of these categories in the two systems.¹ In broad perspective, the genuine contrasts may be characterized more precisely within a framework stressing the common elements that distinguish both capitalism and socialism from other historical economic systems.

The most general, abstract definition of capitalism given by Marx is based on a kind of “Marxian cross.” In his theory, the horizontal relation of commodity production is combined with the vertical wage-labor relation, and consequently dialectically transformed by it. In this approach, the system is based on two separations: (1) the reciprocal separation of producers or enterprises, who are nevertheless interdependent through the division of labor, (2) the separation of workers from the means of production, and consequently from the means of subsistence.² The first separation explains the categories of value and money, the second, based on the first one, the categories of surplus value and capital. These categories represent mediations or “forms of movement” of the double, systemic separation. From this abstract definition, Marx develops more concrete characterizations of capitalism, and discusses various institutional forms and their evolution, analyzed mainly through the English case of the mid-nineteenth century.

Let us leave aside Marx’s moral condemnation of capitalism, the connection he posits between his theory of capitalism and his communist utopia, and the link he establishes in his works between the most general definition of the capitalist system and more concrete historical institutions. He has thus given an abstract characterization of an economic system that is distinguished from other historical systems by the simultaneous presence of money, wages, profit, and capital—and the underlying dual separation. Let us call such a system, not capitalism, but a “monetary—wage labor system” (MWLS). If we use an image from natural sciences, that system could be defined as a genus, and capitalism and socialism as two species within the genus.

Including socialist systems as a second species in the genus of MLWS is of course controversial. But Marx's abstract definition does apply to them. Reciprocal separation of interdependent enterprises exists in socialism as in capitalism, under different *forms*. Reproduction of the general category of *money* in socialist systems is based on such separation, that centralized planning founded on state ownership of capital was not able to abolish.³ Separation of workers from production means is also reproduced, giving to socialist systems, on the basis of their monetary character, their wage-labor dimension. Comparative analysis of capitalist and socialist systems has to take into account both common features and different characteristics; such dual approach is important to theoretize the transformation of national systems from one species to the other.

Within the genus of MWLS, socialist systems are distinguished from capitalist ones by a specific institutional base: the combination of a one-party regime and of the domination of state ownership of capital, both pillars of the base being cemented by communist ideology.⁴

In its developed form, capitalism has existed for more than two centuries. It has experienced a high diversity of forms as national economic systems evolving through historical time. Socialism, conversely, lasted for seventy years in a smaller number of national types (actually four decades in all but one). National capitalist systems based on the Marxian cross have evolved through a great variety of historical institutional forms in five articulated dimensions: monetary regime, wage-labor relation, competition forms, role of the state, and integration with the current international regime.⁵ A diversity of "development modes"⁶ has also been observed in economic history of capitalist systems worldwide. The socialist species has equally been characterized by a diversity of evolving institutional forms and of development modes in national systems.

Nevertheless this diversity is more restricted for two reasons. First, the life span of the species has been shorter than for capitalism, and the number of individual types, that is, national economic systems, more limited. Second, the institutional base defining the *differentia specifica* of the species sets some stricter limits on the variety of possible institutional configurations of intermediate levels, in comparison to the more diversified capitalist species.⁷

Institutional and developmental diversity in capitalist and socialist economies remain circumscribed by broad contrasts, as in the opposition between demand-constrained and resource-constrained systems (Kornai 1992). Diversity develops at a more concrete level determined by specific historical and national forms and trajectories within each species. A task of comparative analysis of economic systems lies in articulating various levels of abstraction: the most abstract level of common

features of monetary-wage labor systems, the general level of contrasted or simply different characteristics of species, and the more concrete level of diverse national forms and trajectories within each species.⁸

Socialist and Postsocialist Transformation and Evolution

The history of formation, evolution, and eventual transformation of various types of socialist systems shows the distinctive role of the institutional base, this apparently heterogeneous institutional mix of an original political system and of an ownership regime. Erection and dislocation of the institutional base signify entry into and exit from the socialist species. Other institutional and developmental forms that emerge or evolve in the framework of the institutional base do not, as such, imply intersystemic change. A wide variety of such forms has been observed in the history of economic reforms in socialist economies (Chavance 1994a), but only dismantling the base implies intersystemic change. In this sense, a peculiar inverted Marxian base/superstructure relation seems to obtain in socialist systems.⁹

Erection of the institutional base is achieved when the imposition of the one party regime is combined with wide-scale nationalization (for example, 1918 in Soviet Union or 1956 in China). The process and institutions of centralized planning, or direct bureaucratic coordination, can only develop once the institutional base has been erected. The latter is the actual condition of “megahierarchy” or the general control hierarchy of the state sector (Chavance 1995). In various forms of reformed socialism, the megahierarchy is weakened or partially fragmented, allowing indirect bureaucratic coordination to evolve. Partial changes take place in the ownership sphere and even in the political system and in their modes of interaction, but without questioning the one-party regime and the domination of state ownership, nor their combination. Consequently, the institutional base remains fundamentally untouched. The dislocation of the institutional base signifying exit from the species can take place through dismantling of one of the two pillars (one-party regime or state ownership). Two possible paths of exit are so defined, depending on the pillar first destroyed or eroded.

Erection of the institutional base, under the influence of communist ideology, has in most cases been the result of a political revolution or of radical military and geopolitical changes caused by outside forces. Dislocation of the institutional base, on the contrary, has in general taken place in a context of a structural crisis of the system that lasted for a decade—except in the case of China. The process of dismantling resulted, in Central Europe and in the former Soviet Union, from the

collapse of the formal institutions of the one-party regime. The second pillar of the base, state ownership, had in radically reformed economies been partially eroded, particularly during the last decade, through the extension of the second economy, and in a limited number of countries (Hungary and Poland) through the initiating of a reform of state ownership. But the collapse of the political pillar, in a kind of chain reaction taking place in a context of significant geopolitical changes resulting from the Gorbachev government's *perestroika* in the USSR, was the unequivocal cause of the dismantling of the institutional base. It allows a precise datation of the start of the process of postsocialist systemic transformation, even though many trends in this transformation emerged before the political break.

Although the hardships, strains, and problems of postsocialist transformation are correctly stressed in many analyses, on a general and historical-systemic level the relative easiness of the change can be—paradoxically—underlined. At such a level, the shift from one species to the other, within the common genus of CWLS, may explain such easiness. Most changes are changes of institutions, formal and informal,¹⁰ but the dual economic separation found in both species remains. Changes in political regime, ownership forms, coordination mechanisms, are overwhelming. But they do not establish or create categories of money, prices, wages, profit, and capital; they modify the intermediate institutional forms in which these categories are produced and reproduced. This is ignored or understated in the conventional view of the “transition to market economy.” But the transformation process actually means a radical change in *forms*, or *metamorphosis*: let us only mention the monetary and banking system, and various aspects of the wage-labor nexus (wage determination, employment regime, industrial relations, work organization). The shift from (partially) passive to active money, from seller's to buyer's markets, from shortage of labor to unemployment and more generally from shortage economy to surplus economy (Kornai 1995a, 1995b) is a shift in the forms of the monetary and wage-labor dimensions of the economic system, not their qualitative emergence. This change of forms had sometimes, in some fields, been initiated under the reformed or the late socialist system.¹¹ To a certain extent, it is not more radical than the changes observed in various formal and informal institutions of different capitalist economies through historical evolution or following big crises, wars, or large social movements.

The Chinese Path

In the Chinese experience of economic reforms can be found both general features common to reform paths in other socialist countries and specific, original Chinese

characteristics. The latter are partially related to peculiarities of the institutional, organizational, and developmental features of the Chinese system, within the socialist species, on the eve of reforms. Beyond the weight of agriculture and the legacy of a low development level, specific features were people's communes that mixed rural administration, collectivized agriculture, local industry, and various social functions; the strict constraints on population movements between countryside and cities; the direct role of party organs in the management of industrial enterprises; administrative allocation of labor and lifetime employment for industrial workers; a limited role of central planning with little "tautness" of plans (Granick 1987); and control by regional or local authorities of a significant fraction of state or "collective" industry.

The main objective of early institutional changes, starting in 1979, was an ambitious "modernization" strategy of communist leaders with a clear nationalistic motivation. Institutional reforms were approached in a cautious, instrumental, and pragmatic way, without any clear blueprint at the beginning. The process of agricultural decollectivization illustrated the complex interplay between the party leadership's attitude of "groping for stones to cross the river" and the social mobilization of peasants, which overcame the initial resistance of rural cadres. The legitimation of this extraordinary break with the Maoist legacy resulted from the remarkable and unexpected boom of agricultural production and from the growth of peasants' incomes in the first half of the 1980s. Such a growth impulse triggered the cumulative unbalanced process of the next decade.

"Urban reforms" launched in 1984 were actually a compromise between various views about desirable changes in state industry (Lin 1989), based on the dual-track system of management and prices for all enterprises. The dual system is typical of the gradual and experimental approach, which allows in principle a reversal of the process if deemed necessary. Initially inspired by agricultural reforms, it is the only experience of progressive and evolutionary deplanning in the history of socialist economies that succeeded.¹²

In the growth context stimulated by agricultural expansion and the local industry boom that followed, and accompanying drawbacks as corruption and speculation notwithstanding, the dual system induced the state industry to "grow out of the plan" (Naughton 1995*b*) in a few years. This result was not totally expected even by reformist leaders, but it soon appeared to be irreversible. The "open door policy" in the 1980s, also experimental and gradual, gave an important contribution to the overall process of change in this period. When a new impulse to economic reforms was given by Deng Xiaoping in 1992, after the conservative backlash following to the 1989 Tiananmen Square massacre, the degree of systemic change effected since

the end of the 1970s was already high in comparison to other experiences of reform in socialist systems. It did not result from a global project of "transition to the market economy" but from a cumulative process where changes in agriculture, the rural economy in general, industry at various levels, trade and services, and relations with foreign partners appeared as mutually reinforcing. The development mode of the Chinese system had been significantly altered.

The Chinese experience shows some broad similarities with other reform processes in socialist economies. We see here a concatenation of reform types or sequences accompanied by corresponding ideological accommodation that were, in other economies, scattered over various historical periods or over different countries. *Systemic adjustment*, mainly characterized by deconcentration processes within the megahierarchy and increased autonomy of enterprises and localities, was initiated for industry during the first half of the 1980s. *Radical reform 1*, which started with agricultural decollectivization, followed, with the extension of various "non-state" sectors and enterprises, the dual-track regime in state industry, introduction of temporary hiring for industrial workers, and the extension of early experiences of opening (special economic zones, opened regions or cities, foreign investment). Then came *Radical reform 2* in the first half of 1990s, marked by banking reform, tax-system reform (uniform corporate tax and VAT), and reform of state enterprises (for example, corporatization, a variety of "shareholding systems"), radical reform 1, and radical reform 2 correspond to reform experiences that took place in other socialist economies.¹³ But in the 1990s, some institutional changes introduced in China go beyond such previous attempts, reaching *Radical reform 3*, that is, measures that were introduced in Eastern Europe only after the systemic shift, as opening of stock exchanges (1990), beginning the externalization of social protection from state enterprises (for example, privatization of housing, forming social welfare funds), moving to external convertibility of the yuan.

The four types of reforms were not strictly sequential but overlapped, and the frontiers were sometimes blurred by the conflicting political reformist cycles that marked the whole process and still mark it today. No coherent reform doctrine was elaborated in the process, the dominant pragmatic mood often leading to expost rationalization of actual changes.

The ideological evolution of the communist regime has been remarkable indeed. As in the history of official reformist doctrines in other countries, the general trend has been a gradual flexibilization of socialist ideology, evolving into a process of increasing euphemization of the postulated contrasts between socialism and capitalism. The successive stages of this evolution are: admitting a plurality of ownership forms in socialism (that is, a limited space for nonstate enterprises), introducing

a controlled market in the planned economy, admitting a regulated market as the main form in the socialist economy (the socialist market economy called for in 1993). The process of “hollowing out” of the official socialist ideology has in some respects gone farther in China than it did in Poland or Hungary in the 1980s; the change in mentalities and attitudes in society at large is also striking.

Sustained Growth and Systemic Change

Systemic sliding in China following institutional reforms and consequent behavioral changes has been obviously path-dependent and contingent on specific Chinese conditions (Ash and Kueh 1996). It is a historical case of circular and cumulative causation, if we combine Veblenian and Kaldorian insights by understanding the concept as a combination of a continuous process of accelerated organizational and institutional change on one hand and of a progressive growth regime on the other.

The Chinese systemic slide reveals a specific mixture of intention and spontaneity. The irregular but continuous movement away from inherited rules and regularities has built on increasing returns of institutional and behavioral change in a context where reserves of entrepreneurship at different levels of the economy were significant and where various agents' capacities of learning in a rapidly changing environment appeared remarkable. Changes in formal and informal institutions have interacted, modifying behavior and anticipations and eventually the whole development mode. In a complex and conflictual social process, a cumulative causal chain between institutional change, policy shifts, supply restructuring, and continuous growth of demand, has been formed on a middle-term historical perspective.

The high growth context has played a fundamental role in the Chinese experience. By contrast, in some Eastern European countries *radical reforms* were introduced in the 1980s when the economy was plunged in a structural crisis, with low growth and high foreign indebtedness, while most radical institutional changes were effected during the early severe postsocialist depression, creating many strains and legitimation problems. In China high growth was an explicit objective of the “modernization” strategy,¹⁴ but reformist leaders had no clear initial views about the means they would use or accept to this end. Although such growth created tensions of its own, as inflation and overheating, leading to a cyclical pattern of economic policy and reform, it provided the main legitimation of the reform process in the eyes of various social groups, including numerous layers of the state and party apparatus. It has made the wide-scale redistributive effects of systemic change between social groups, regions and localities, and levels of state administration

acceptable, and most strata in the countryside or in the cities have obtained an increased standard of living in the process.

Original Forms and Unconventional Ways

Emerging institutional and organizational forms, marked by the Chinese socialist heritage, often have a hybrid character (Nee 1993) or are disturbingly nonconformist (Rawski 1994; McMillan 1995) in the eyes of the "transition orthodoxy."

This is an element of the debate opposing standard views stressing the irrelevance of the Chinese experience to the general transition debate (Sachs and Woo 1994, 1996), to China specialists or evolutionary critics of the standard approach (McMillan and Naughton 1992; Murrell 1993; Rawski 1995; Naughton 1995a, Walder 1995b). Although the Chinese lessons have no interest for the present transformation strategy of postsocialist countries of Europe, inasmuch as the problems and conditions (especially political) are so different, they remain of great value as a challenge to conventional views about the irreformability of socialist systems, to standard peremptory prescriptions about transition policy in general, and to more fundamental theses of economic theory. The future evolution of the Chinese economy remains uncertain, as the process of change is highly unbalanced, full of internal tensions and threats. Nevertheless, and whatever happens in the future, almost twenty years of endogeneous evolution and transformation of the economic system in China have to be taken into account in any general dynamic and historical theory of socialist systems.¹⁵

In agriculture, land ownership has remained formally "collective" with land use attributed to peasant species for periods up to thirty or fifty years. Leases can be rented or transmitted to children, but land cannot be bought or sold. The system is a kind of generalized family farming with state "rent." In other fields, the most remarkable feature of Chinese reforms is the extension of the so-called non-state sector. This broad and fuzzy category includes collective enterprises, urban and rural (the latter consisting mainly of the famous township and village enterprises), "individual" enterprises in the cities or countryside, and "other" forms (that is, "private" and mixed Chinese and foreign firms). The wide "collective" category is difficult to define: are these enterprises public or semipublic,¹⁶ quasi-private, mixed or hybrid, or "vaguely defined cooperatives" (Walder 1995a; Weitzman and Xu 1994)? In any case, they are not dependent of a higher hierarchical line, and although local authorities' interference is widespread, TVEs have been the most dynamic sector in the whole reform process.

The state sector underwent little change in formal ownership in the 1980s. Evolution was confined to a wide-scale process of devolution of control to regional or local authorities, strengthening a Maoist legacy and contributing to the remarkable decline in the weight of the central budget (Wang 1995). The result was a further fragmentation of the megahierarchy, in a period when the dual-track system led to a shrinking sphere for central (or regional) planning. During the 1990s a diversification of formal categories of industrial ownership within or around the state sector has taken place, with a growing variety of substantive property-rights configurations within a similar formal category. Simultaneously, a crisis and progressive mutation of the state industrial sector develops along with a progressive hardening of the budget constraint, with some similarities with the first phase of postsocialist transformation in European countries: an increasing number of loss-making enterprises, growth of interenterprise debt and bad loans, and so on. As state enterprises remain the principal source of budgetary revenues at various levels, this has become a growing source of concern for Chinese officials.

Although the notion of privatization has remained taboo, reform or "restructuring of state assets" has accelerated since 1993. The measures in favor of corporatization have given rise to a number of initiatives of enterprises or advanced regions. Many changes have taken place in the formal structure of ownership, in actual distribution of property rights, and in organizational forms. The analysis is made difficult by the importance of semiformal or informal changes, by the fuzzy character of statistical categories, and by the lack of a clear interpretive framework among Chinese scholars themselves, linked to the uncertainty of the official ideological schemes.

A part of state ownership is transferred to collective, private, or foreign ownership through a number of ways: leasing, spinoffs, joint ventures. Various models of the "shareholding system" are developed in diverse localities (Ma 1995). Cross-ownership of different kinds is extended, reminiscent of "recombinant property" (Stark 1996) or "interwoven ownership" (Chavance and Magnin 1997) typical of postsocialist European economies, but, unlike them, without bank involvement. Some large "state-owned" enterprises in a city like Shanghai can take the form of a holding company, with shareholding participations distributed between the municipal administration bureau of state-owned assets, other state enterprises, collective or private enterprises, employees and managers; within the enterprise can be hosted a number of joint ventures (Yuan Zhigang 1996).

In the ownership sphere, main developments in China have thus been the extension of the non-state sector and the simultaneous reduction of the state sector proper to a minority of industrial production (corresponding to large enterprises

and more concentrated in traditional or heavy sectors),¹⁷ and a mutation process of state ownership resulting in a growing diversity of configurations about formal and informal property rights. A majority of the non-state sector could be termed non-private as well (Putterman 1995), while dispersed "institutional" shareholding and managerial power often characterize the evolving "public" ownership.

If Chinese reforms of the 1980s led to the emergence of a "socialist mixed economy" with a growing institutional and organizational diversity (Nee and Stark 1989), the continuing systemic sliding has evolved into a specific postsocialist mixed economy where original Chinese innovations are combined with similar features found in Eastern Europe, as a dynamic non-state sector, a growing multiplicity and heterogeneity of owners for large enterprises, fuzzy mixtures of private and public forms, interwoven ownership within or around the public sector, and high managerial power (Chavance and Magnin 1997).

Evolution of the Coordination Mix

The change of the coordination setup in evolution and transformation of socialist economic systems cannot be reduced to a simple transition from plan to market. Both species of capitalist and socialist orders have complex coordination systems where various modes or mechanisms are combined in specific ways: they have a different "coordination mix."¹⁸

Coordination modes are complex social and institutionalized processes that coordinate the division of labour in society or in organizations as firms. Considering both social and technical division of labor, monetary-wage labor systems generally combine various modes of coordination: state, markets, networks and microhierarchies.¹⁹ Markets are here defined abstractly by the mediation of monetary exchange between production and consumption. Specific mechanisms determining prices and quantities, and various types of market power determine diverse *forms* of markets, at a more concrete degree of analysis. In this approach, markets are present in all monetary-wage labor systems, but in great variety, or on a continuum of different forms.

Processes of radical reforms of socialist systems or of postsocialist transformation, are marked by significant changes in the coordination mix, that evolves toward a configuration that is closer to the generic mix of the capitalist species.²⁰ Some modes of coordination can go through qualitative and rapid transformation, as with state coordination, through the abolition of megahierarchy, or with the switch of some markets from the domination of seller's to the domination of buyers (Kornai 1995a,

1995b); types of competition that represent important mediations in the coordination of the social division of labor may be transformed significantly. But evolution and transformation of the coordination mix is a complex process where no clear dividing line separates systemic types.

The Chinese path of coordination change is consequently more complicated than a simple process of growing out of the plan into the market. It mixes elements of socialist reform changes and postsocialist transformation. The *state* microregulation has shrunk to a small minority of industrial production; the megahierarchy has been fragmented in the process of devolution of control to lower levels of state administration, and its significance has been reduced by the gradual extension and the generalization of horizontal links for state enterprises. The activity of the central *state* has generally shifted from microregulation to "indirect bureaucratic coordination," and a rather polycentric state administration has partially replaced the megahierarchy. The decrease in the degree of redistribution has implied a hardening of the budget constraint even for state enterprises, although the constraint is generally harder in the non-state sector. Social and industrial *networks*, with stabilized individual or organizational relations based on trust and reciprocity, played an important role in Chinese socialism. They have been transformed, restructured, and sometimes extended in the process of change of the coordination setup. The decrease of vertical links does not imply the only extension of horizontal ones, but also frequently an activation of networks, as a coordination mode that represents neither market nor (mega)hierarchy, or runs across both state and markets. The networks character of emerging Chinese capitalism mixes elements of Chinese tradition of "bureaucratic capitalism" of the 1920s and 1930s (Bergère 1986), postsocialist features of intrication between state and non-state sectors, and cross-ownership ties of various kinds (Stark and Bruszt 1994), and characteristics typical of Asian forms of capitalism, in which business networks and family groups play an important role (Orru, Biggart, and Hamilton 1991).

The extension of *markets*, or "marketization" proper, has been a strong tendency in the Chinese economy during the last two decades in the sense that new goods and new services have been subjected to monetary exchange and that new types of markets have emerged (as housing market or financial markets). Extension of inter-regional trade chains and interindustrial exchanges have altered the "cellular" form of economy inherited from the Maoist autarkic strategy. Marketization in this sense should not be confused with the metamorphosis of markets that existed in the socialist economy as markets for consumption goods, production goods, and "labor."

Such change of forms has been radical mainly for product markets, where the gradual freeing of a majority of prices, through the dual-track system, has spread to

the whole economy in less than a decade. Entry and competition of the non-state sector have accelerated the process (Qian and Xu 1993; Naughton 1995*b*) contributing to the shift from seller's to buyer's markets. Changes in the labor market remained limited in the 1980s but have accelerated in the 1990s (Chan 1996), with a growing segmentation between workers hired on a temporary or a permanent basis in state industry, and between state or non-state industry in general. An increasing diversity of "wage-labor nexus" forms (including work organization and wage formation) is present in Chinese industry and services, ranging from protoindustrial types to lifetime employment. Urban unemployment has grown but is difficult to measure. Most agricultural production remains based on family, excluding wage labor. But significant internal and temporary migrations are now observed between rural and urban regions with different levels of development and growth: in the latter they form the "floating population."

As for *microhierarchies*, a general diversification has been observed, with the creation of many new types of small and medium enterprises experiencing a rather hard budget constraint, and the diversification of state enterprises following various strategies of survival, restructuring, opening to foreign investors, expansion, and transformation. The general tendency has been a hardening of the hierarchical constraint for employees in industrial enterprises.

A Second Way toward Postsocialist Transformation

Following our approach of the institutional base, the *differentia specifica* of socialist systems is given by the combination of the one-party regime and of the dominant role of state ownership.²¹ This combination is compatible with a variety of other institutions, although within a certain range illustrated by history of these systems and especially experiences of economic reforms. As a consequence the dislocation of the institutional base, which signifies exit from the socialist species and integration in the capitalist species,²² can take place through two different ways: dismantling of the political pillar or erosion of the state ownership pillar.

The first way is illustrated by Central European and former Soviet countries. The second way corresponds to the Chinese (and possibly the Vietnamese) case. Gradual erosion of the domination of state ownership has progressively dismantled the institutional base, its relative coherence, and produced a qualitative transformation in the economic system. But the piecemeal hollowing out of communist ideology about the contrast between socialism and capitalism was also instrumental in such dismantling. Crumbling away of the institutional base's cement affects its stability and

coherence, even before its second pillar—predominance of state ownership—starts cracking. The Chinese experience, as noted by W. Brus, suggests that domination of communist ideology should be distinguished from the one-party system.²³

There is an asymmetry between the formation of the socialist system from capitalism or protocapitalism, and the postsocialist transformation. In the first case, the imposition of the one-party regime leads in all cases to the building of the second pillar of the institutional base, state ownership, and consequently to the effects in terms of coordination, behavioral regularities and development mode. The role of ideology is fundamental in this process of ownership transformation; J. Kornai (1992) writes about the “genetic program” based in the political bloc (one party-regime plus ideology) that transmits in a quasi-spontaneous process the characteristics of the system to every cell or single social relationship. The Kornaian “main line of causality” so explains principally the process of formation of the socialist system, its extension from one national economy to other ones, and its reproduction during periods of relative institutional stability. But in the process of systemic sliding resulting from sustained radical economic reforms, the main line can be modified, reversed or broken. Gradual deprogramming takes place and becomes irreversible. Changes in elements or subsystems interact with the institutional base. Evolution in the coordination mix and organizational diversification influence the ownership sphere.²⁴

The political system itself is modified by institutional and societal changes (Jie 1995). When the communist party has abandoned fundamental theses of its old ideology, has been converted to the “market economy,” and has accepted the limitation of state ownership to restricted parts of the economy, it falls in this category of monocratic regimes compatible with a capitalist economic system, and even promoting it. Such a “quiet revolution from within” (Walder 1995*d*) was maturing in Poland (and especially in Hungary in the late 1980s) before the regime’s demise; it has progressed much further in China during the 1990s.²⁵

In the Chinese evolutionary way out of the socialist species, we see a progressive move from radical reforms to postsocialist transformation. A Hegelian development of quantitative into qualitative change has taken place, a crossing of the systemic Rubicon.²⁶ Dating the qualitative change in this case is more difficult to evaluate, and more open to controversy. Actually no clear-cut line divides systemic orders in the coordination sphere or in the ownership sphere, where shifts of dominance are effected gradually. This is also the case in the first way of postsocialist transformation. In China, my hypothesis is that the Rubicon was crossed in the first half of the 1990s, following the dismantling of the institutional base through ownership changes and ideological transformation.

The Chinese transformation is a process where weakened systemic coherence and increased heterogeneity and diversity, rather than accelerating a structural crisis as happened in Poland or Hungary, appear as vectors toward sequential but continuous and irreversible change. J. Kornai's thesis of the inherent instability of "market socialism" is thus confirmed, but with an argument that goes beyond the negative experiences of European socialist systems.

Other historical examples of high and sustained growth, extending over one or two decades have generally been observed in contexts of relative institutional and organizational stability.²⁷ This was the case of the favorable, and historically exceptional, period of the "golden age" of advanced capitalist economies after World War II until the early 1970s. A gradual destabilization of the interaction between the institutional setup, the accumulation regime and the regulation mode, led to the end of that period. The exceptional character of the phase of high growth in China comes not only from the unprecedented level and duration of that growth, but also from the context of continuous change of institutions and systemic drift that both prompted it and was made possible by it, and finally from the role of hybrid and nonconformist forms of ownership and coordination that were created in the process.

The theoretical challenge resulting from the Chinese experience is that high growth has been the consequence of gradual systemic change in a context of growing heterogeneity, lack of affinity, and extension of "fuzzy," nonconformist, or unconventional forms. Cumulative and circular causation, as we know from history, is a contingent and fragile process that does not last in a long-term perspective. But lessons of such experiences do matter for our understanding of economic systems and economic history.

Notes

1. In the "political economy of socialism" developed during the late Stalin period, the capitalist "form" of economic categories was opposed to their socialist "content" that was postulated to result from the abolition of private property and the planning of production. This opposition had been first formulated by Bukharin in his *Economics of the Transition Period* (1920) (Chavance 1980, 1998).

2. I follow here the interpretation of C. Bettelheim (1976).

3. The "political economy of socialism," the official doctrine in socialist countries, after denying the presence of money in the socialist economy, eventually admitted it but was never able to explain it convincingly (Chavance 1980, 1998). Economic theories in the West have often characterized socialist economies as nonmonetary or semimonetized (for the latter approach see Kornai 1992), on the basis of the identification of money in general and "active money" in particular. But money in general, as the expression and mediation of the separation of interdependent producers, should be distinguished of the various forms it can assume in diverse institutional and historical configurations. Consequently socialist systems

can be analyzed as a specific (and somewhat extreme) form of monetary system rather than as non-monetary or semimonetary system.

4. This proposition is close to János Kornai's (1992) "main line of causality" where political structure and official ideology form bloc 1, state ownership bloc 2, and bureaucratic coordination, behavioral features, and typical economic phenomena blocs 3, 4 and 5 respectively. The concept of institutional base stresses the specific role of the connection between the first two Kornai's blocs in the definition of socialist systems within the genus of MWLS.

5. These represent the five "institutional forms" or intermediate categories discussed in the French "regulation theory" (see Boyer 1990).

6. The development mode includes an "accumulation regime" and a "regulation mode" (Boyer 1990).

7. Institutional and organizational innovations from below in traditional socialist systems are thus more strictly constrained than in capitalist systems, or they remain confined to the informal sphere. But they are not totally absent, as history testifies.

8. The contribution of Kornai's (1992) work is mainly developed at the general level of comparison of capitalism and socialism, using stylized representations of both historical systems. A remarkable achievement of this book is the combination of theoretical and historical analysis of the socialist system, at the abstraction level of the "family" of national systems. The Marxian and Schumpeterian inspirations are manifest here. Who will write a twin book with a similar Kornaian approach, *The Capitalist System*?

9. This image should not be pushed too far, as in theories in terms of the "domination of politics" (see Chavance 1994b), which confuse the actual crucial role of the political regime in defining socialism as an economic system and the (illusory) mastery of the economy by the apparatus of the party-state.

10. D. North (1994) states that changes in informal institutions generally take more time than in formal institutions, but swift transformation of the former can also be observed. Creation and inertia interact in complex and diversified ways. Many tensions and frictions typical of postsocialist transformation result from different temporalities at various levels of institutional change. The same was true at the early times of formation of the socialist systems.

11. The shift to capitalism from reformed socialism illustrates the lack of a clear-cut institutional frontier in various fields of change. But even the shift from traditional or classical socialism (as in the GDR or Czechoslovakia) also reveals qualitative changes in intermediate institutional forms, not a qualitative change in the dual nature (monetary and wage-labor) of the abstract system.

12. The Gorbachev-Ryzhkov economic reform of 1987 was another similar attempt to introduce, for the sake of tactical compromise, a dual system in state industry. In this case it not only failed, but it was one also of the causes of the eventual collapse of the Soviet system.

13. Typical examples of systemic adjustment were the East German reform of 1963 or the Kosygin package of 1965 in the USSR. The Czechoslovak reform of 1966–1968 or the Hungarian New Economic Mechanism of 1968 were radical reforms 1, based on the notion of combining a flexibilized plan with a controlled market. Radical reforms 2, which included an extension of the private sector, banking and tax reform, and which began transforming state ownership, coming close to the notion of privatization, took place in Hungary and Poland during the 1980s, especially the second half of the decade (Chavance 1994a).

14. In 1981 Deng Xiaoping set the goal of quadrupling the level of per capita GNP by 2000. The Chinese growth record at that time was good in comparative perspective, if we except giant fluctuations that accompanied Maoist campaigns (the Great Leap Forward and Cultural Revolution). But demographic trends and the example of fast-growing Asian neighbors prompted the "acceleration" strategy.

15. Atypical cases or evolutions have often been underestimated or relegated to specific boxes of proposed typologies in the field of "comparative analysis of economic systems," which suffered from a static and somewhat descriptive bias. This remains the case of the Yugoslav experience, which has seldom been fully integrated in theories of socialist systems. The Japanese case stimulated a wide debate on the various forms of capitalism only since the 1980s, at a time when its challenge to American supremacy was openly felt.

16. TVEs are formally owned by the township or village, and their management is controlled by the corresponding local authorities.
17. In 1965, 90 percent of industrial gross output value came from state enterprises and 10 percent from collective ones. The proportion was 78 and 22 percent in 1978. The state sector's share has continuously decreased since, and it has been less than half since 1992 (48 percent). In 1994 it represented 34 percent of IGOV, collective enterprises 41 percent, individual enterprises 11 percent, and "other ownership enterprises" 14 percent (*China Statistical Yearbook* 1995).
18. See Chavance and Magnin (1997). The concept of a structured combination of coordination mechanism is developed in J. Kornai (1992), where he distinguishes five main types: bureaucratic coordination, market coordination, self-governing coordination, ethical coordination, and family coordination.
19. Microhierarchy within the enterprise is a mode of coordination of the *technical* division of labor. In socialist systems, where state ownership dominates, a megahierarchy or general control hierarchy plays an important role in the coordination of the *social* division of labor (along with other modes as networks and markets). Although the principle of verticality is common to both types of hierarchy, they should be clearly distinguished (Chavance 1995).
20. Such a generic mix is an abstract category encompassing a great variety of national types in economic history.
21. J. Kornai gives a broad definition of the capitalist system based on two criteria, a positive one (the predominance of the capitalist forms of private property) and a negative one (the absence of a party in undivided possession of power committed to eliminating such predominance) (Kornai 1992, pp. 89–90). This suggests a general characterization of capitalism by symmetry with the two first blocs defining socialism. The line of argument proposed in this chapter is convergent on this point.
22. As stated before, socialist systems belong to the same genus of MWLS as do capitalist systems, the distinction coming from the institutional base. If the latter is dismantled, socialist systems are "naturally" transformed into capitalist systems, or in a middle-term perspective and at a lower level of abstraction, into a subspecies of "postsocialist capitalist systems." Abstractly, transformation into a third species of monetary-wage labor systems could be imagined, but no such systems are, as Keynes would have said, "in sight." Let us note that various historical "market socialist" blueprints, whether old or new, would—if they hypothetically had been tested in practice—have belonged to the MWLS genus. As for possible blueprints for nonmonetary and nonwage labor systems for postcapitalist (or postsocialist) systems, they have very few followers today. The Marxian communist blueprint is neither feasible nor desirable (Nove 1983; Chavance 1994b).
23. Deng Xiaoping persuaded the Chinese leadership "that in order to preserve the "undivided power of the Communist Party" it had to embark upon another "long march," this times toward a full-fledged market economy, perhaps even capitalist economy" (Brus 1993b, p. 180). In this case, such undivided power is not any more an "unsurmountable obstacle to profound change in the economic system" (Ibid.). See also Brus (1993a).
24. No clear and direct correlation or "affinity" between ownership forms and coordination modes appears in a short-term perspective (Rawski 1994), illustrating the different temporalities of various levels of change. In postsocialist Europe, too, the change in the coordination mix toward depaternalization and modification of market forms has exhibited a different, shorter temporality than transformation in the ownership system.
25. In the Chinese experience, enhanced autonomy of local or regional levels of state or party apparatus was essential in the process of implicit political change (Shirk 1993; Lieberthal 1995). A. Walder (1995c, p. 16) notes that instead of a process of autonomization of society vis-à-vis the state, we rather see here the "lower reaches of a state structure gaining increased autonomy from higher reaches and forging new kinds of political and economic ties with individuals and enterprises outside." See also J. Oi (1992, 1995); J. Unger and A. Chan (1995).
26. Kornai used the image of crossing the Rubicon to depict the change from reform to "revolution," but he limited the latter to the case of breaking the monopoly of power of the Communist party (Kornai

1992). I submit that there is a second way to effect such a change, that does not imply dismantling of the one-party regime—at least in a middle-term historical perspective.

27. But in a number of such cases, significant waves of institutional change had taken place before.

References

- Ash, R., and Y. Kueh (1996). *The Chinese Economy under Deng Xiaoping*. Oxford: Clarendon Press.
- Bergère, M.-C. (1986). *L'âge d'or de la bourgeoisie chinoise 1911–1937*. Paris: Flammarion.
- Bettelheim, C. (1976). *Economic Calculation and Forms of Property*. New York: Monthly Review Press.
- Boyer, R. (1990). *Regulation Theory: A Critical Assessment*. New York: Columbia University Press.
- Brus, W. (1993a). The politics and economics of reform: reflections on János Kornai's new book. *Acta Oeconomica* 45(1–2): 173–182.
- Brus, W. (1993b). Marketization and democratization: The Sino-Soviet divergence. *Cambridge Journal of Economics* 17(4): 423–440.
- Chan, A. (1996). Chinese enterprise reforms: convergence with the Japanese model?. In B. McCormick and J. Unger (1996), 181–202.
- Chavance, B. (1980). *Le capital socialiste. Histoire critique de l'économie politique du socialisme (1917–1954)*. Paris: Le Sycomore.
- Chavance, B. (1994a). *The Transformation of Communist Systems: Economic Reform since the 1950s*. Boulder, Colo.: Westview Press.
- Chavance, B. (1994b). *La fin des systèmes socialistes: crise, réforme, transformation*. Paris: L'Harmattan.
- Chavance, B. (1995). Hierarchical forms and coordination problems in socialist systems. *Industrial and Corporate Change* 4(1): 271–291.
- Chavance, B. (1998). La théorie de l'économie socialiste dans les pays de l'Est entre 1917 et 1989. In A. Béraud and G. Faccarello, eds., *Nouvelle histoire de la pensée économique*, vol. 2. Paris: La Découverte.
- Chavance, B., and E. Magnin (1997). Emergence of various path-dependent mixed economies in post-socialist Central Europe. In A. Amin and J. Hausner, eds., *Beyond Market and Hierarchy*. Aldershot: Edward Elgar, 196–232.
- China Statistical Yearbook*. (1995). Beijing: China Statistical Publishing House.
- Granick, D. (1987). The industrial environment in China and the CMEA countries. In G. Tidrick and Chen Jiyuan, eds., *China's Industrial Reform*. New York: Oxford University Press.
- Jie, Chen (1995). The impact of reform on the party and ideology in China. *The Journal of Contemporary China* 9: 22–34.
- Kornai, J. (1992). *The Socialist System: The Political Economy of Communism*. Oxford: Clarendon Press.
- Kornai, J. (1995a, 1995b). Eliminating the Shortage Economy: A General Analysis and Examination of the Developments in Hungary. *Economics of Transition* 3(1): 13–37 and 3(2): 139–168.
- Lieberthal, K. (1995). *Governing China: From Revolution through Reform*. New York: W. W. Norton.
- Lin, C. (1989). Open-ended economic reform in China. In V. Nee and D. Stark, eds. (1989), 95–136.
- Ma, Shu-Yun (1995). Shareholding system reform: the Chinese way of privatization. *Communist Economies and Economic Transformation* 7(2): 159–174.
- McCormick, B., and J. Unger (1996). *China after Socialism: In the Footsteps of Eastern Europe or East Asia?* Armonk, N.Y.: M. E. Sharpe.
- McMillan, J. (1995). China's nonconformist reforms. In E. Lazear, ed., *Economic Transition in Eastern Europe and Russia*. Stanford: Hoover Institution Press, 419–433.

- McMillan, J., and B. Naughton (1992). How to reform a planned economy: lessons from China. *Oxford Review of Economic Policy* 8: 130–143.
- Murrell, P. (1993). What is shock therapy? What did it do in Poland and Russia? *Post-Soviet Affairs* 9(2): 111–140.
- Naughton, B. (1994a). What is distinctive about China's economic transition? State enterprise reform and overall system transformation. *Journal of Comparative Economics* 18(3): 470–490.
- Naughton, B. (1994b). Chinese institutional innovation and privatization from below. *American Economic Review* 84(2): 266–270.
- Naughton, B. (1995a). China's economic success: effective reform policies or unique conditions? In K. Poznanski, ed., *The Evolutionary Transition to Capitalism*, pp. 135–153. Boulder, Colo.: Westview.
- Naughton, B. (1995b). *Growing out of the Plan: Chinese Economic Reform, 1978–1993*. New York: Cambridge University Press.
- Nee, V. (1992). Organizational dynamics of market transition: hybrid forms, property rights and mixed economy in China. *Administrative Science Quarterly* 37(1): 1–27.
- Nee, V., and D. Stark (1989). *Remaking the Economic Institutions of Socialism: Eastern Europe and China*. Stanford: Stanford University Press.
- North, D. (1994). Economic performance through time. *American Economic Review* 84(3): 359–358.
- Nove, A. (1983). *The Economics of Feasible Socialism*. London: Allen & Unwin.
- Oi, J. (1992). Fiscal reform and the economic foundations of local state corporatism in China. *World Politics* 45: 99–126.
- Oi, J. (1995). The role of the local state in China's transitional economy. *China Quarterly* 144: 1132–1149.
- Orru, M., N. Biggart, and G. Hamilton (1991). Organizational isomorphism in East Asia. In W. Powell and P. DiMaggio, eds., *The New Institutionalism in Organizational Analysis*, pp. 361–389. Chicago: University of Chicago Press.
- Putterman, L. (1995). The role of ownership and property rights in China's economic transition. *China Quarterly* 144: 1047–1064.
- Qian, Yingyi, and Xu Chenggang (1993). Why China's economic reforms differ: The M-form hierarchy and entry/expansion of the non-state sector. *Economics of Transition* 1(2): 135–170.
- Rawski, T. (1994). Progress without privatization: the reform of China's state industries. In V. Milor, ed., *Changing Political Economies: Privatization in Post-Communist and Reforming Communist States*, pp. 27–52. Boulder, Colo.: Lynne Reiner.
- Rawski, T. (1995). Implications of China's reform experience. *China Quarterly* 144: 1150–1173.
- Sachs, J., and W. T. Woo (1994). Structural factors in the economic reforms of China, Eastern Europe and the Former Soviet Union. *Economic Policy* 18: 101–145.
- Sachs, J., and W. T. Woo (1996). China's transition experience, reexamined. *Transition* 7(2–3): 1–5.
- Shirk, S. (1993). *The Political Logic of Economic Reform in China*. Berkeley: University of California Press.
- Stark, D. (1996). Recombinant property in East European capitalism. *American Journal of Sociology* 101(4): 993–1027.
- Stark, D., and L. Bruszt (1994). Restructuring Networks in the Postsocialist Economic Transformation. Cornell Working Papers on Transition from State Socialism, 94/9.
- Unger, J., and A. Chan (1995). China, corporatism, and the East Asian model. *Australian Journal of Chinese Affairs* 33: 29–53.
- Walder, A. (1995a). Local governments as industrial firms: an organizational analysis of China's transitional economy. *American Journal of Sociology* 101(2): 263–301.
- Walder, A. (1995b). China's Transitional Economy: Interpreting its Significance. *China Quarterly* 144: 963–979.

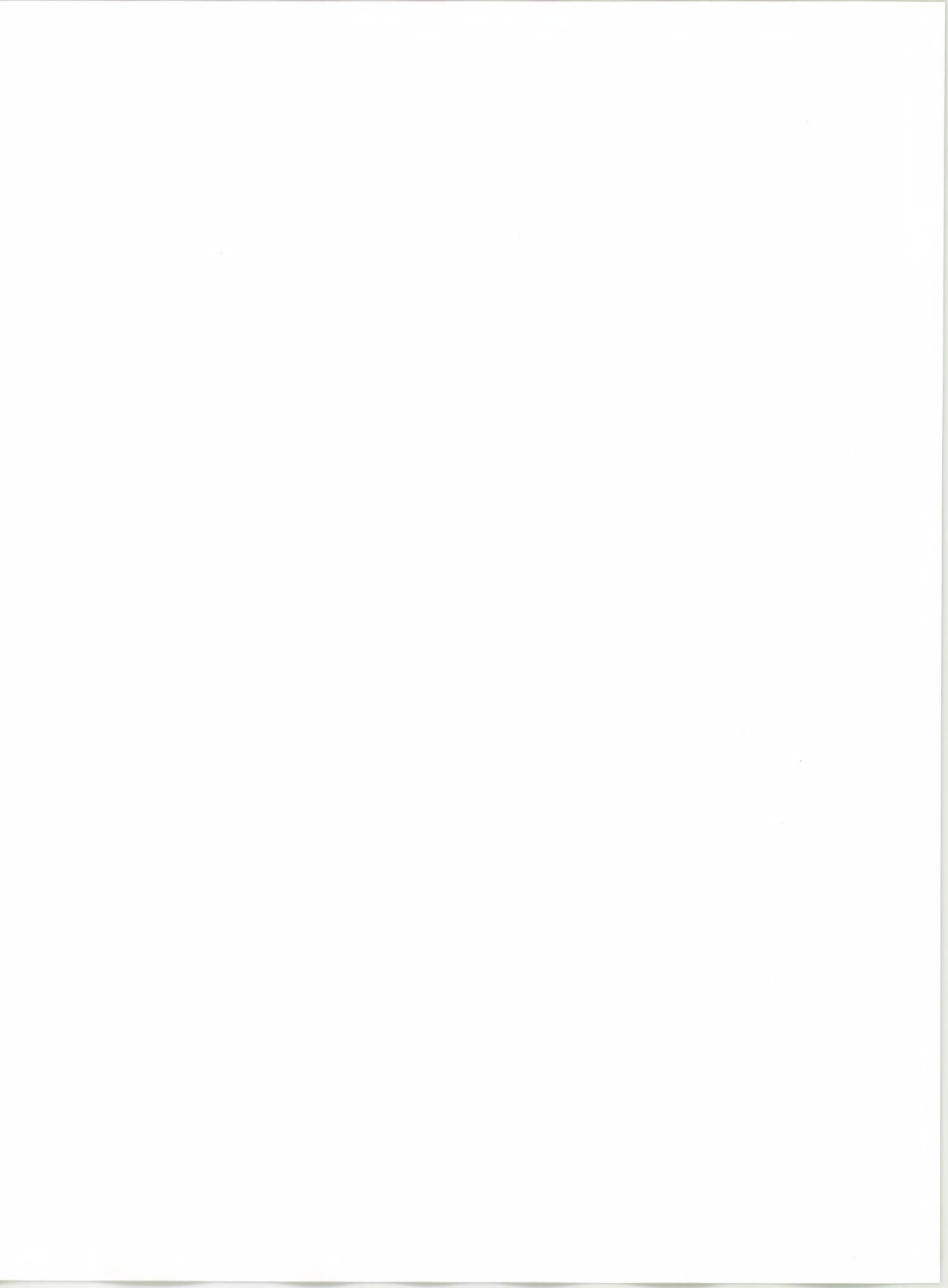
Walder, A., ed. (1995c). *The Waning of the Communist State: Economic Origin of Political Decline in China and Hungary*. Berkeley: University of California Press.

Walder, A., ed. (1995d). The quiet revolution from within: economic reform as a source of political decline. In A. Walder, ed. (1995c), pp. 1–24.

Wang, Shuogang (1995). The rise of the regions: fiscal reform and the decline of central state capacity in China. In A. Walder, ed. (1995c).

Weitzman, M., and Xu Chenggang (1994). Chinese township-village enterprises as vaguely defined cooperatives. *Journal of Comparative Economics* 18(2): 121–145.

Yuan, Zhigang (1996). La reconstitution des actifs de l'État et la contrainte du chômage en Chine. Working paper, CEMI, École des Hautes Études en Sciences Sociales, Paris.



15 **Perestroika Economics from the Inside**

Michael Ellman

1 Introduction

Perestroika economics was very strange by the standards of Soviet economic history in the post–World War II period, of standard macroeconomics, and of transition economics. In seven years a relatively stable macroeconomic situation was replaced by acute shortages and monetary and fiscal crisis, open inflation exploded, the net foreign debt grew rapidly, and the country embarked on a deep depression. At the time it was difficult to understand why and how the policies that led to these results were conceived and implemented. Now, however, there is a new source, the abundant memoir literature, which throws new light on this topic. This chapter discusses some of the topics on which the memoirs have provided new information.

2 Macroeconomics

During the Brezhnev period, the strict control on monetary emission (i.e., financing the budget deficit by increasing cash in circulation) of the late Stalin and Khrushchev periods was relaxed (Sitnin 1994, p. 47). Nevertheless, even in the Brezhnev period the Politburo devoted considerable attention to balancing the retail market. Each year the plan and the budget for the following year were carefully considered at a meeting of the Politburo. These meetings were not just rubber stamps—they sometimes took two days (Kunaev 1994, p. 259). In these annual discussions in the Politburo on the plan and budget for the following year most attention was given to “the imperative that there be no deficits in the balance of money incomes and expenditures of the population, that is, that money incomes should correspond to the retail trade plan.¹ If there was a discrepancy of 8 to 10 billion rubles, or even 5 billion, it became a huge problem for N. K. Baibakov (then chairman of Gosplan), for Kosygin (later replaced by Tikhonov), and for Brezhnev himself. Year after year, new plans for emergency measures were born, including price increases on various goods in order to balance the plan and the budget. . . . This was better than the budgetary anarchy unleashed later by Prime Minister Nikolai Ryzhkov and Valentin Pavlov.”² (Arbatov 1991 p. 249, and 1992, pp. 214–215.) Why was this attention to financial control at the highest political level replaced by the “budgetary anarchy” of the Gorbachev years? There seem to have been a number of factors at work.

One reason has to do with Gorbachev’s struggle, in 1980–1982, to obtain approval for his Food Program. This was an expensive program, combining increases in

procurement prices (paid by the state to the farms) with substantially increased government investment in the agro-industrial sector. The Ministry of Finance strongly opposed raising agricultural procurement prices in the absence of noninflationary financing for this (Gorbachev 1995a, p. 197). Similarly, Gosplan opposed the increase in investment in agriculture that Gorbachev advocated (Zoteev 1998). Gosplan argued—quite rightly—that it would upset macroeconomic equilibrium. Gorbachev, then the Politburo member responsible for agriculture and the author of the program, found these attitudes extremely unhelpful. When he came to power, this negative experience may have influenced his approach to financial caution and to the allocation of the national income. His “solution” in 1982 to the problem of finding noninflationary financing for the Food Program (Gorbachev 1995a, p. 205) was to spend openly as agricultural expenditure the money already being spent, but disguised as “loans” made by the banking system to the state and collective farms and never repaid. Gorbachev was quite right to think that these unrecovered loans were equivalent to government expenditure, but this source of financing would not have worsened inflationary pressures only if the practice of making loans to agriculture that were subsequently written off had ended. Because it did not, this “solution” to the problem of finding resources for the Food Program was an illusory one. The leadership also proposed in 1982 to raise revenue by retail price increases for food, but this idea was abandoned prior to implementation.

As far as Gosplan’s opposition in 1982 to increasing investment in agriculture is concerned, Gorbachev in 1982—like Dzerzhinsky in 1926 (Ellman 1992, pp. 110–111)—seems to have lacked understanding of the macroeconomic effects of excessive investment. This had unfortunate macroeconomic consequences after Gorbachev attained supreme power and launched his ill-fated “acceleration” policy, the core of which was an increase in investment in engineering. Instead of paying attention to the sensible arguments of Gosplan, Gorbachev (1995a, p. 336) “hoped to overcome stagnation by using ‘the advantages of socialism,’ such as plan-mobilization methods, organizational work, the consciousness and activity of the working people.” As a former Gosplan economist put it (Zoteev 1998), Gorbachev’s “political mentality at that time was typically Soviet—political will transcended common sense.”

Another decision that reflected Gorbachev’s cavalier attitude to financial constraints concerns grain imports (Gorbachev 1995a, pp. 234–235). In the winter of 1982–1983 the leadership decided on substantial grain imports, despite warnings from the government that there was no money to finance this.

These experiences as head of a major spending lobby (agriculture) struggling with a Ministry of Finance worried about inflationary expenditures and a

Gosplan worried about excessive investment, worries that he rejected as needless obstructionism, seem to have colored Gorbachev's attitudes after he became general secretary. This had very adverse consequences for consumer welfare and the economic system. For example, when he set out his initial program at the April (1985) Plenum, he spoke about modernizing the engineering sector and accelerating technical progress, but had nothing to say about the need to strengthen financial control, even though by then he had probably been briefed about the real state of the budget.³ In retrospect, he regards his illusion in 1985 as being that radical reform would be possible with the inherited officials (Gorbachev 1995a, p. 283). An economist, on the other hand, would argue that his illusion was the idea that radical reform was possible while ignoring the need for financial control (McKinnon 1993). In 1987, during the preparations for the June Plenum on economic reform, there was a discussion in the Politburo about the financial situation of the country—the first time the Politburo had discussed the budget deficit (Gorbachev 1995a, p. 351). The minister of finance, B. I. Gostev, warned about the dangers of increasing the financial autonomy of enterprises, but his advice was rejected. Gorbachev himself strongly argued against reducing the budget deficit by reducing food subsidies (Gorbachev 1995a, p. 352). Yakovlev, too, had a cavalier attitude to the need for financial control. At the end of 1985, in a long letter to Gorbachev he expressed the view that the Ministry of Finance needed to be “curbed” in view of its propensity to restrict the freedom of the enterprises (Yakovlev 1994, p. 211).

The decline in the share of enterprise profit paid into the state budget in 1987–1989, and the huge increase in the enterprise incentive funds that followed from Gorbachev's “reforms,” were major destabilizing factors that substantially increased inflation and shortages in the last years of the USSR. The responsibility for this lies with Gorbachev and Yakovlev. The “conservative” Ministry of Finance warned at the time of the dangers of launching such policies.⁴

Another factor that caused the neglect of stabilization policy, in particular in 1987–1989, was Gorbachev's often repeated view that perestroika was a revolution. In revolutions things are bound to be destabilized, and calls for stabilization he deemed to be “conservative” and “bureaucratic resistance to perestroika.” For example, his close aide Cherniaev (1997, p. 20) summarized Gorbachev's attitude in August–September 1989 to stabilization as follows: “He spoke out sharply against stability: what do we need stability for, we are experiencing a revolution. If there is stability, that will be the end of perestroika; stability is stagnation.”⁵ Even in October 1991, when the country was in an acute economic crisis and was disintegrating both economically and politically, Gorbachev told a foreign visitor (Cherniaev 1997,

p. 221), "there is no need to dramatize the situation. We are experiencing a revolution, and under such circumstances complete order is impossible."

Gorbachev's views and actions in 1987–1990 can be compared to Stalin's in 1929–1930. Both leaders rejected macroeconomic equilibrium. Stalin expressed this with his rejection of "equilibrium,"⁶ Gorbachev with his rejection of "stability." Both succeeded in destroying the economic system they had inherited (the NEP in Stalin's case; socialist planning/the administrative economy/the command economy in Gorbachev's case). In both cases, the transition from one economic system to another caused much suffering, although in the former case the suffering (dekulakization, famine, terror) was much greater than in the latter case. Stalin succeeded in creating a system that lasted for half a century, but that system turned out not to be viable in the long run. Gorbachev's role was mainly destructive and he himself did not succeed in creating a new economic system. However, Gorbachev played a key role in making possible the subsequent introduction throughout the Former Soviet Union of local variants of an economic system that had established its viability and dynamism on a global scale.

Defending himself and his policies at the 28th Party Congress (July 1990), Gorbachev (1995a, p. 545) argued that the miserable situation at that time with respect to food supplies was a result of the poor state of agriculture and the food processing industry which he had inherited in 1985. This entirely ignored the role of his irresponsible macroeconomic policies in generating the shortages (Cook 1992).

3 Price Reform

One way of overcoming macroeconomic disequilibrium, of reducing the budget deficit and restoring equilibrium on the retail market, was to raise prices. This policy was several times implemented in Eastern Europe. A clear theoretical exposition of the usefulness of price increases, *ceteris paribus*, in restoring equilibrium on the retail market has been given by Peebles (1991, chapter 4). Two former Politburo members, Medvedev (1994, pp. 54–56) and Ryzhkov (1992, pp. 243–244; 1995, p. 249), whose views on many questions differ, both argue that the failure to introduce price increases ("price reform," in the euphemism of the perestroika period) in January 1988 was a major policy error that played a very important role in ensuring the failure of the economic reform adopted with such fanfare in June 1987. According to Medvedev, the responsibility for the failure to act decisively when it was still possible was a failure of political will resulting from lack of firm support from leading economists, opposition in the media, and lack of a firm position by the government.

Ryzhkov blames a failure of political will among politicians and economists, resulting from concern about possible public response. Similarly, Shevardnadze (1991, p. 318) states that the need to readjust prices was obvious but that "we were afraid of popular dissatisfaction, the loss of wide support for perestroika."⁷

It should be noted that price increases of the type discussed in the USSR in the mid-1980s (i.e., increases in the price of basic food products) are not necessarily effective under state socialist conditions in reducing shortages. There are three reasons for this. First, workers may obtain compensating (or more than compensating) wage increases by, for instance, organizing strikes. The importance of this phenomenon was demonstrated several times in Poland. Second, as Kornai (1980, chapter 19; 1985) pointed out, planners may respond to a reduction in consumer demand by measures that maintain the "normal" level of shortages. Third, as Podkaminer (1982, 1986, 1987, 1988) pointed out, the fundamental cause of disequilibrium on the consumer-goods market was not the low price of food but the low price and limited availability of durables and services. According to this last line of argument, what was necessary to restore equilibrium on the Soviet consumer-goods market in the mid-1980s was primarily an increase in housing rents and in the price and availability of consumer durables and services. Podkaminer's argument has been confirmed by relative price movements in Poland since price liberalization in 1989–1990 (Bell and Rostowski 1995). On the other hand, prices increases which reduce subsidies or increase the taxes or profit paid to the state budget will improve the fiscal situation (unless they are completely counterbalanced by compensation payments to the public, which themselves are financed out of the budget).

Detailed light on the failure to increase prices until 1991 has been shed by Pavlov who was directly involved in this area, not at the rarified level of the Politburo but at the practical level of detailed price policy. According to Pavlov, price increases were seriously considered several times in 1982–1990, but each time were deferred by Gorbachev, who bears personal responsibility for this major policy failure. In July 1982 Brezhnev signed a decree about increases in the price of bread and other goods (with compensation for the public), scheduled to go into effect on January 15, 1983. In November 1982, however, after the death of Brezhnev and the succession of Andropov, Gorbachev (who while Brezhnev had been alive had strongly backed the introduction of the new prices) personally persuaded Andropov to defer the introduction of the new prices on the grounds that it would be a serious political error to begin his rule with such an unpopular decision (Pavlov 1995, pp. 70–71). Although initially only deferred by one month, this decision in fact amounted to a cancellation of the proposed price increases. Gorbachev (1995a, p. 234) confirms that the

planned bread and other price increases were canceled after Brezhnev's death on his advice, although his account differs in some details from that of Pavlov.⁸

In 1984 the "conservative" Pavlov, then head of a department in Gosplan, sent a report to the central committee secretary for economic affairs, Ryzhkov, in which he explained the difficult financial-monetary situation (Pavlov 1995, pp. 200–201). He stated that unless appropriate measures were taken, by 1989 there would inevitably be a collapse of the financial-monetary system. In particular, he pointed out—quite correctly—that with the existing system of prices, neither agricultural production increases nor food imports could possibly solve the food problem.

In 1985, when Gorbachev himself came to power (Pavlov 1995, p. 72), "he became aware of the necessity to alter the price structure" (this terminology suggests that he was briefed on this matter by financial officials and either accepted or gave the impression of accepting, their arguments). Nevertheless, he took no action on this matter. In 1987, in the light of the decisions of the June Plenum about economic reform, new wholesale prices were calculated by the State Committee on Prices and distributed to the localities for introduction from January 1, 1988. At the end of 1987, however, the top political leadership decided not to introduce them (Pavlov 1995, pp. 78–85).⁹ The political leadership was afraid of the political consequences that might follow the increase in retail prices, which was regarded as a logical consequence of the higher wholesale prices. Neither Gorbachev nor Ryzhkov was prepared to take the responsibility for what was expected to be an unpopular decision.

Ryzhkov renewed his interest in price reform in 1989, which ultimately led to the fiasco of May 1990, when the Supreme Soviet rejected the government's proposal to raise retail food prices. At the same time there was a run on the shops to take advantage of the existing prices, leading to a worsening of shortages and an increase in sociopolitical tension. Pavlov (1995, p. 63) explains the long delay in raising retail prices, which, according to him, was absolutely necessary already at the beginning of the 1980s, by the fact that "Party and state officials just as previously [i.e., when Kosygin had declined to increase retail prices] were afraid of accepting the responsibility for the possible social and political consequences of raising retail prices." He refers repeatedly and ironically to their "Novocherkassk syndrome," which paralyzed action in this area. (Novocherkassk is a town in the North Caucasus where Khrushchev's 1962 food price increases led to riots that were suppressed with the use of military force and repression.) No doubt developments in Poland in the 1970s and 1980s¹⁰ were also relevant.

Although the information provided by Pavlov is interesting and probably largely correct, it is incomplete. He neglects to mention (it would spoil his negative evalu-

ation of Gorbachev) the December 4, 1986, session of the Politburo at which the Council of Ministers' proposal to increase prices was discussed (Cherniaev 1993, pp. 122–123). At this session Gorbachev himself backed the proposal (as did Ryzhkov and others) but it was strongly opposed by Ligachev, Vorotnikov, and Shevardnadze.¹¹ As a result of this strong disagreement, the decision to raise prices was deferred. In fact, Gorbachev was not the consistent opponent of price reform described by Pavlov. Gorbachev's attitude to price reform fluctuated. He argued for it at the December 4, 1986, meeting of the Politburo and fully accepted its importance at another Politburo meeting on April 30, 1987, where he even said (Vorotnikov 1995, p. 142), "Prices decide everything".¹² On the other hand, after price reform had been rejected (at the end of 1987) he argued quite differently. In March 1989, at a central committee plenum, he argued, "Reform of prices is necessary, but now is not the time. It is necessary to increase production (for 2–3 years) and then decide" (Vorotnikov 1995, p. 253). In the same year, in response to the observation by U.S. Secretary of State Baker that it was necessary to begin economic reform with a reform of prices, Gorbachev said, "I think that if we had done that, people would very soon have lost faith in that policy which begins by hitting them in the pocket. . . . In any case we are twenty years late with price reform. Another two or three years makes no difference" (Cherniaev 1993, p. 294).

4 The Appeal for Western Money

The normal Soviet practice was to avoid significant debts to the West. The long-serving former Gosplan chairman Baibakov (1993, p. 255) quotes Kosygin as once telling him, "Nikolai, look for resources where you like, but under no circumstances will I allow us to be debtors of Western countries." Although some money was borrowed from the West in the Brezhnev period, the amounts were quite modest. Under Gorbachev, policy changed radically.

A characteristic feature of 1990–1991 was the appeal to the West for money. Pavlov (1995, pp. 276–333) has described how he (then minister of finance) and Gerashchenko (then chairman of Gosbank, i.e., the Soviet monobank) visited the United States in the spring of 1990 to investigate the possibility of securing large-scale U.S. financing for the restructuring of the Soviet economy. They hoped, in particular, to secure \$25 billion to finance the transition to the market. They were disappointed to discover that the Bush administration, Congress, and private business were all unwilling to finance the USSR on any significant scale. Pavlov's reaction was to reorient his search for foreign loans from the United States to Europe.

The purpose for which the leading Soviet politicians sought this money was to soften the blow to the public (with higher prices, reduced production, and unemployment) resulting from the transition to the market. As Shevardnadze put it at a discussion of economic policy in April 1990, when some speakers stressed the social costs involved (Vorotnikov 1995, p. 373), "Is it possible for us to rely on foreign help? I propose to make more use of direct foreign investment in our economy." In this way he hoped to make the Ryzhkov program socially acceptable.

Traditionally, Soviet leaders had regarded the West as a threat to the USSR, as an ideological enemy and a possible source of military aggression. This had been an important factor determining their economic system and economic policies. Gorbachev, with his "new thinking," saw the West as a model and a source of assistance. The Soviet foreign trade economist Kurierov (1998) has noted, with respect to the reform of the foreign trade sector, that "it was inspired not so much by the immediate economic needs of that system [i.e., the Soviet economic system], as by abstract ideas, based on the experience of a fundamentally different economic system, of how foreign economic relations should be organized in a contemporary 'civilized' state." When his policies led the USSR into a deep crisis, instead of changing the policies, Gorbachev turned to the West for help. At a meeting in Madrid in October 1991 (Cherniaev 1997, p. 245), Gorbachev asked Bush for \$3.5 billion in agricultural export credits and \$370 million in immediate cash. He also asked Bush to press Saudi Arabia and South Korea for a loan of \$1 billion.

Abalkin (1992, chapter II.2) has described how as an adviser to the president in 1991 he prepared for Gorbachev's meeting with the G7 in London in July 1991. Abalkin's group considered that the idea of obtaining billions of dollars from the leaders of the G7 as a gift was illusory, and it put forward the idea of the inclusion of the USSR in the world economy as a mutually beneficial development that would require cooperation on various levels, including the financial.

Cherniaev (1993, p. 445) has explained how in 1991, prior to August, Western help formed one of the three prongs of Gorbachev's policy for dealing with the situation in the country. The other two were, sociopolitical agreement (e.g., by means of an agreement between the center and the republics) and economic stabilization (for example by means of a government economic stabilization program). After August 1991, Gorbachev's press secretary for the last few months of 1991, Grachev (1994, chapter 2) has described how foreign assistance and domestic public opinion were the twin instruments with the help of which Gorbachev hoped to save the Soviet Union. Gorbachev (1995*b*, pp. 607–613) himself has given an account of his efforts in September–December 1991 to secure Western economic assistance. In November 1991 the external payments situation had grown so acute, that

Gorbachev sent Major—as coordinator of the G7—a letter asking for \$1.5 billion, of which \$320 million was requested within two weeks (Cherniaev 1993, p. 499).

Western leaders gave some encouragement to the notion that perestroika could count on Western economic backing. In December 1989 at the Malta summit, President Bush, discussing U.S. economic relations with the USSR (Gorbachev 1995*b*, pp. 143–144), offered not “help” but “cooperation.” He stated that the administration intended to initiate measures leading to extending most favoured nation treatment to Soviet exports and making credits to the USSR possible and granting the USSR observer status in GATT. At the summit meeting in the United States in May–June 1990, Bush said that he had decided to sign the trade agreement with the USSR. Gorbachev (1995*b*, p. 186) valued this, not so much for the trade agreement per se, but as a sign that the United States was moving “from verbal support for perestroika to real deeds.” Seventeen months later, in their October 1991 Madrid meeting with Gorbachev, the U.S. leaders went farther (Cherniaev 1997, p. 247). Secretary of State Baker offered food gifts of \$250 million, \$250 million of immediate agricultural export credits, and an additional \$1 billion of agricultural credits in two months. Baker also drew attention to U.S. assistance in obtaining a loan from Saudi Arabia for the USSR. Although these measures were less than Gorbachev had asked for, they were at any rate something. European countries also provided tied credits. In May 1991 the Soviet Security Council approved Pavlov’s proposal for Soviet membership of the IMF, with only KGB head Kriuchkov opposing it (Cherniaev 1997, p. 141).¹³ In October 1991, in connection with the country’s entry into the IMF, the managing director of the IMF, M. Camdessus, told Gorbachev (Cherniaev 1997, p. 224) that if the right policies were pursued, “the USSR, not some time or other in the indefinite future, but within a few years, will become an economic superpower.” This idea, supported by the West, was a complete illusion. The USSR never became an economic superpower. After the USSR’s collapse, the FSU countries experienced high inflation, depression, criminalization, and state collapse.

5 The Evolution of Gorbachev’s Thinking

Although committed to radical change already before his accession, Gorbachev initially focused on change within the traditional Soviet economic model (Ellman and Kontorovich 1997, pp. 269–270). His 1985 economic policies were intended (Cherniaev 1997, pp. 319–320) “to make the system work. The system itself was not subjected to any doubts. . . . In his first years it did not occur to Gorbachev to think about the rejection of socialism, of the Soviet system, of the power of the CPSU, of

the monopoly of state ownership and the dictatorship of the state plan." This was the intellectual ambience in which "acceleration" (1985) and the 12th Five Year Plan (1986) were launched.

By 1986, Gorbachev's views had radicalized. He now rejected the entire post-1929 history of socialist construction and was searching for a Leninist alternative to Stalinism (Tsipko 1998). This was the ambience in which the "radical reform" announced at the June (1987) Plenum was formulated.

By 1990–1991 his views had radicalized further, and in principle he supported the transition to a market economy. He devoted great efforts to trying to turn the CPSU into a social-democratic party. It was in this ambience that he jointly (with Yeltsin) commissioned Shatalin and others to work out what became the "500 days program" (Gorbachev 1995a, chapter 18).

Gorbachev was one of those unusual politicians, like Gladstone, who get more radical as they get older. Nevertheless, Gorbachev never adopted the "transition orthodoxy" of Balcerowicz, Kohl (in the former GDR), and Gaidar. As late as May 1991, his wife—probably reflecting a family consensus—told Burlatsky (1996, p. 270), "The people are not prepared for the market, are not prepared for the rejection of state ownership. People value what socialism has given them—free education, medical care, and housing. Socialism should be our orientation, despite all reforms."

The only constants in Gorbachev's economic policy ideas were his belief in illusions and quick fixes. Whether it was the 1985 idea of reinvigorating the economy by combining an investment boom in engineering with renewed reliance on administrative methods, the 1987 idea of introducing market elements in the economy without ensuring financial control, his 1990 idea of a transition to the market without social costs (Yasin 1998), or his (pre-August) 1991 idea of combining a rapid transition to the market with a socialist orientation, or his (post-August) 1991 listening to foreigners who offered him the perspective of turning the USSR into an economic superpower, he never seems to have had much grasp of economic reality, feasible economic policies, and the time lags between policy initiatives and results.

6 Conclusion

The memoir literature enables economic policy in the perestroika period to be seen from the inside. The "budgetary anarchy" of the perestroika period and the investment cycle of 1985–1991 partly resulted from Gorbachev's distrust of the Ministry of Finance and Gosplan resulting from his struggle with them in 1980–1982 when

he was head of a major spending lobby (agriculture). Partly they resulted from the traditional Bolshevik failure to understand the macroeconomic consequences of excessive investment. They also partly resulted from Gorbachev's oft-repeated view that perestroika was a revolution, and that in a revolution "stability" is the slogan of the counterrevolutionary forces. The destabilizing price policy was largely a result of the fear by the leadership of the political consequences of retail price increases. The Western money for which the Soviet leadership repeatedly appealed in 1990–1991 was intended to soften the social costs of economic transformation and hence make it politically feasible. The perestroika leadership saw the West as a model and a source of assistance. The West supported the idea that following the economic policies recommended by the West would generate Western economic assistance and lead to the USSR's becoming an economic superpower. Gorbachev's economic ideas evolved substantially in 1985–1991, but unfortunately only from illusion to illusion, from one failed quick fix to another.

Notes

This chapter is an offshoot of a joint project with Vladimir Kontorovich on the collapse of the Soviet economic system; see Ellman and Kontorovich (1997). The main fruit of this project is a book, Ellman and Kontorovich (1998).

1. The text, both Russian and English, reads "balance of payments" not "balance of money incomes and expenditures of the population," but this seems to be a mistake, in as much as the text makes much more sense if it refers to the "balance of money incomes and expenditures of the population."
2. This "quotation" is a composite one, based on both the Russian and English texts. The last sentence is only in the English text. The contrast between the efforts of the Brezhnev leadership to maintain financial equilibrium and the "budgetary anarchy" under perestroika is a sound one. On the other hand, blaming it on Ryzhkov and Pavlov while ignoring the responsibility of Gorbachev, Yakovlev and Medvedev, is a political distortion.
3. Gorbachev (1995a, p. 234) states that some information about the real state of the budget he only learned shortly before resigning as president (in 1991). He does not say whether this was because officials consciously hid the data from him, or because he was so little aware of the importance of financial control that he did not ask for them. According to Gorbachev (1995b, p. 289) he only obtained a full picture of the budget situation in June–July 1991, during the preparations for the London G7 meeting.
4. Of all his disputes with the Ministry of Finance in 1981–1987, the only one in which Gorbachev seems to have backed an economically rational policy concerns the taxation of individual labor activity (Gorbachev 1995a, p. 345). The Ministry of Finance advocated higher tax rates than Gorbachev and Ryzhkov thought appropriate.
5. "Stagnation" was Gorbachev's term for the Brezhnev period.
6. See his speech to the conference of agrarian-Marxists in December 1929 (Stalin 1949, pp. 143–146).
7. This confirms Shtromas's (1981, p. 66) argument about the political limits on the freedom of action of the Soviet leaders in the economic policy field.
8. According to Gorbachev, the initiative in reconsidering the matter after Brezhnev's death came from Andropov himself, the task was assigned to Gorbachev and Ryzhkov, and Gorbachev was kept in the dark about some important relevant economic variables such as the real situation of the state budget.

Furthermore (unlike Pavlov), Gorbachev does not state that he supported the price increases while Brezhnev was alive.

9. In his 1,200 pages of memoirs, Gorbachev does not mention this at all. He does, however, find room to blame conservative opponents of his policies for "sabotage" (Gorbachev 1995a, pp. 298, 305, 347, 361, 396).

10. Gorbachev seems to have been chairman of the Politburo's Polish commission in 1984–1985. See Pechenev (1991, p. 96). That Gorbachev was the last chairman of the Polish commission (the first was Suslov, who was succeeded by Andropov and then Chernenko) is confirmed by Shakhnazarov (1993, p. 114), although he does not give precise dates. Gorbachev (1995b, pp. 338–339) describes the work of the Politburo's Polish commission but omits to point out that he was its last chairman. This omission may result from his wish to disclaim his joint responsibility for the Politburo's Polish policy in the period before perestroika.

11. The memoirs of Ligachev, Vorotnikov, and Shevardnadze all neglect to discuss this episode. This lacuna is particularly striking in Vorotnikov's memoirs, because he actually describes this Politburo session (Vorotnikov 1995, p. 116) but entirely omits references to its discussion about price policy.

12. The phrase "Prices decide everything" is a reference to the famous Stalinist slogan "Cadres decide everything."

13. In his memoirs Gorbachev (1995b, p. 288) has nothing about Pavlov's personal role, referring instead to "a memorandum of the Cabinet of ministers." Possibly this results from Pavlov's subsequent participation in the August putsch. In retrospect, Gorbachev considers his appointment of Pavlov as head of the government in January 1991 a mistake (Gorbachev 1995a, p. 587). In addition, whereas Cherniaev gives the impression that the Soviet application for IMF membership was approved at this meeting, Gorbachev simply says that the memorandum of the cabinet of ministers was "discussed." Possibly Gorbachev wanted to downplay the positive influence of Pavlov's cabinet prior to the August putsch. Gorbachev's account focuses on his possible participation in the London G7 meeting, although, as subsequent events showed, participation in the IMF by the USSR/FSU was far more important for economic policy than one person's participation in a summit meeting.

References

- Abalkin, L. I. (1992). *K tseli cherez krisis. Spustia god. . .* Moscow: Luch.
- Arbatov, G. A. (1991). *Zatianuvsheesia vyzdorovlenie (1953–1985gg). Svidetel'stvo sovremennika*. Moscow: Mezhdunarodnye otnosheniia.
- Arbatov, G. A. (1992). *The system*. New York: Times Books.
- Baibakov, N. K. (1993). *Sorok let v pravitel'stve*. Moscow: Respublika.
- Bell, J., and J. Rostowski (1995). A note on the confirmation of Podkaminer's hypothesis in post-liberalization Poland, *Europe-Asia Studies* 47(3): 527–530.
- Burlatsky, F. (1996). *Russkie gosudari epokha reformatsii*. Moscow: Shark.
- Chernaiev, A. S. (1993). *Shest' let s Gorbachevym*. Moscow: Kul'tura.
- Chernaiev, A. S. (1997). *1991 god: Dnevnik pomoshchnika Prezidenta*. Moscow: Respublika.
- Cook, E. C. (1992). Agriculture's role in the Soviet economic crisis. In Ellman and Kontorovich (1992), pp. 193–216.
- Ellman, M. (1992). Money in the 1980s. In Ellman and Kontorovich (1992), pp. 106–133.
- Ellman, M., and V. Kontorovich (1992). *The Disintegration of the Soviet Economic System*. London: Routledge.
- Ellman, M., and V. Kontorovich (1997). The collapse of the Soviet system and the memoir literature. *Europe-Asia Studies* 49(2): 259–279.

- Ellman, M., and V. Kontorovich, eds. (1998). *The Destruction of the Soviet Economic System: An Insiders' History*. Armonk, N.Y.: M. E. Sharpe.
- Gorbachev, M. S. (1995a). *Zhizn' i i reformy*, vol. 1. Moscow: Novosti.
- Gorbachev, M. S. (1995b). *Zhizn' i reformy*, vol. 2. Moscow: Novosti.
- Grachev, A. (1994). *Dal'she bez menia . . . ukhod prezidenta*. Moscow: Progress.
- Kornai, J. (1980). *Economics of Shortage*, vol. 2. Amsterdam: North-Holland.
- Kornai, J. (1985). Gomulka on the soft budget constraint: a reply. *Economics of Planning* 19(2): 9–55.
- Kunaev, D. (1994). *Ot Stalina do Gorbacheva (v aspekte istorii Kazakhstana)*. Almaty: Sanat.
- Kurierov, V. (1998). Foreign trade liberalization. In Ellman and Kontorovich (1998), pp. ••.
- Ligachev, Ye. K. (1992). *Zagadka Gorbacheva*. Novosibirsk: Interbuk.
- McKinnon, R. (1993). *The Order of Economic Liberalization: Financial Control in the Transition to a Market Economy*. Second edition. Baltimore: Johns Hopkins University Press.
- Medvedev, V. A. (1994). *V komande Gorbacheva*. Moscow: Bylina.
- Pavlov, V. (1995). *Upushchen li shans?* Moscow: Terra.
- Pechenev, V. (1991). *Gorbachev: k vershinam vlasti*. Moscow: Gospodin narod.
- Peebles, G. (1991). *A Short History of Socialist Money*. Sydney: Allen & Unwin.
- Podkaminer, L. (1982). Estimates of the disequilibria in Poland's consumer markets, 1965–1978. *The Review of Economics and Statistics* 64(3): 423–431.
- Podkaminer, L. (1986). Persistent disequilibrium in Poland's consumer markets: some hypothetical explanations. *Comparative Economic Studies* 28(3): 1–16.
- Podkaminer, L. (1987). On Polish disequilibrium once again, *Soviet Studies* 34(3): 509–512.
- Podkaminer, L. (1988). Disequilibrium on Poland's consumer markets. *Journal of Comparative Economics* 12(1): 43–60.
- Ryzhkov, N. I. (1992). *Perestroika: istoriia predatel'stv*. Moscow: Novosti.
- Ryzhkov, N. I. (1995). *Desiat' let velikikh potriasenii*. Moscow: Assotsiatsiia "Kniga. Prosveshchenie. Miloserdie."
- Shevardnadze, E. A. (1991). *Moi vybor*. Moscow: Novosti.
- Shakhnazarov, G. (1993). *Tsena svobody*. Moscow: Rossika.
- Shtromas, A. (1981). *Political Change and Social Development: The Case of the Soviet Union*. Frankfurt-am-Main: Lang.
- Sitnin, V. K. (1994). Iz opyta finansovo-kreditnykh reform v SSSR (zametki finansista). *Problemy prognozirovaniia* 6: 42–53.
- Stalin, J. V. (1949). Teoriia "ravnovesiia", *Sochineniia* vol. 12, pp. 143–146. Moscow: Politizdat.
- Tsipko, A. (1998). The collapse of Marxism-Leninism. In Ellman and Kontorovich (1998), pp. 169–186.
- Vorotnikov, V. I. (1995). *A bylo eto tak . . .* Moscow: "Sovet veteranov knigoizdaniia" SI-MAR.
- Yakovlev, A. N. (1994). *Gor'kaia chasha*. Yaroslavl: Verkhne-Volzhscoe knizhnoe izdatel'stvo.
- Yasin, Ye. (1998). The parade of market transformation programs. In Ellman and Kontorovich (1998), pp. 228–237.
- Zoteev, G. (1998). The view from Gosplan: growth to the year 2000. In Ellman and Kontorovich (1998), pp. 85–94.

16 **The Internal versus Occupational Labor Market—A Neglected Dimension of Hungary's Postsocialist Transformation**

István R. Gábor

The severe transformational recession of the 1990s in Hungary was eagerly accommodated in the labor market: employment declined in excess of the fall of production, and real wages declined substantially as well. With these socially painful adjustments over, Hungary's labor market is, however, now seemingly reluctant to accommodate the much hoped-for economic upswing: even under the circumstances of positive economic growth and rising investment, employment continued to decline last year (1997); the relatively high rate of unemployment remains, and without effectively mitigating wage inflationary pressure; in spite of the fact that the majority of those squeezed out of employment have withdrawn from the labor force—particularly those with little hope of being rehired later—long-term unemployment is not diminishing. This involves the danger that the ongoing economic recovery will fuel wage inflation to a degree that calls for restrictive economic policies, with the consequence that the recovery may get stuck before growing demand for labor could bring forth any sizable increase in employment. The overtly excluding character of the Hungarian labor market, moreover, as a result of which huge masses of people are compelled to make their living as self-employed, may act to perpetuate the current unfortunate state of affairs in which overpopulated and overfragmented small businesses are a medium of an acute danger of long-term economic dualization.

In our searching into the causes and cures of this unfortunate situation in the rest of this chapter, the legacy of the socialist era, notably the prominent role and systemic characteristics of internal labor markets of large enterprises of the time, will be considered first. Next we will look at how those particular internal labor markets have largely disintegrated and dissolved into occupational labor markets in the 1990s without similarly far-reaching complementary changes in the institutional infrastructure of the labor market having been implemented, which produced a kind of institutional imbalance. We will argue that it is this uneven process of institutional development (combined, perhaps, with factors widely held to be responsible for unemployment persistence in some long-established market economies) that accounts for Hungary's disquieting labor market experience in the current initial stage of economic recovery. The main lesson we will draw from this assessment is that in the years to come a prerequisite of lasting economic growth and small-business modernization in Hungary is the alignment of the institutional infrastructure of the labor market with the logic of operation of occupational labor markets.

The Legacy of Socialism: The Prominent Role and Systemic Characteristics of Internal Labor Markets

Within the context of the indirect mechanism of central economic management of the 1970s and 1980s, with labor power having remained the single typically privately owned resource of production, labor markets were the most “authentic” markets of the Hungarian economy in that they were coordinated dominantly through market (i.e., nonbureaucratic) mechanisms.¹

As a matter of fact, under the conditions of the essentially transactive relationship that existed between buyers and sellers of labor, wage differentials, and variations in working conditions played a prominent role in the allocation of workers among firms and occupations, economic branches and regions, and so on, as well as in the motivation of workers’ on-the-job behavior.² Direct government interventions into the allocation of labor were as a rule campaignlike and/or of limited range of effect; central wage regulation, a permanent component of economic regulation, served as a device to constrain artificially the overall rate of growth of workers’ earnings in order to mitigate inflationary pressure rather than to directly influence earnings differentials; centrally imposed criteria of job classification and remuneration (the so-called wage tariffs) did not in fact substantially diminish employer discretion in the formation of earnings differentials; and forced participation in the labor force was virtually absent.

At the same time, owing to the conditions of a bureaucratically coordinated, resource-constrained economy under which it operated, the Hungarian labor market differed markedly from labor markets in contemporary market economies. Most importantly, its normal state of equilibrium was characterized by an overall excess demand for labor—a disequilibrium that persisted at a high rate of employment³ and in the presence of a sizable and expanding, characteristically small-scale private economy, the so-called second economy, acting as a supplementary source of living rather than an alternative to full-time employment in the large-firm sector.

With forced participation virtually absent, the attainment of this high rate of employment was conditional upon a correspondingly high rate of voluntary participation. Beside social transfers and allowances that reduced the time and money costs of market work and/or discriminated against the unemployed, this high rate of voluntary participation was brought about essentially by two spontaneous mechanisms, both bound to occur under the conditions of full employment, namely: (1) a kind of state-dependence of voluntary participation, augmented by (2) a positive feedback between the rate of employment and the rate of voluntary participation.⁴

As regards the former mechanism, the essence is that expanding employment (possibly through the mobilization of previously nonemployed labor) tends to augment the "normal rate of participation" by giving rise to the proportion of those who are "accustomed to employment" and thus willing to be in the labor force.

As for the latter mechanism, its essence is that the further the rate of employment and, with it, the earner/dependent ratio rise, the weaker the "relative purchasing capacity" of wages becomes (defined as the ratio of the average real wage to the per capita real consumption of the working population), meaning that households can only catch up with rising living standards by adjusting their participation decisions to earlier rises in the earner/dependent ratio, that is, by further expanding their economic activity.

That is, although in the absence of open unemployment labor supply could not even temporarily be curtailed by withdrawals from the labor force of discouraged workers, a specific, long-run version of the so-called added-worker effect was largely responsible for the fast attainment of a high rate of participation. In contrast to its short-run version, this long-run effect is produced not by job losses of some household members (typically primary earners), which, to the extent that their job losses are feared to be permanent, urge some previously nonemployed (economically inactive) members in those same households to seek employment, but by the demonstration effect of the growth of consumption other households achieve by increasing their rates of participation.

Although the rapidly growing rate of employment, via its (by definition negative) impact on the relative purchasing capacity of wages, prompted Hungarian households to increase their labor supply accordingly, it also weakened the motivating potential of the prevailing wage differentials, and so did the growing role of the second economy as a supplementary source of household incomes. To wit, both can be regarded, from the viewpoint of individual households, as processes leading to "more diversified portfolios" of their labor investments, weakening thereby their sensitivity (the subjection of their per capita consumption capacities) to fluctuations in the earnings (returns to the labor investments) of any one of their members.

Furthermore, although this weakened sensitivity would call for a widening of the wage scale as a way to counter the resulted loss in the motivating force of the prevailing wage differentials, the weakened relative purchasing capacity of wages, itself a consequence of the very same conditions (high employment rate coupled with high proportion of parallel participation in the second economy), in fact narrows the feasible range of those differentials for the following reason. Given the average wage, and assuming that minimum and/or reservation wages must be positively related to the costs of living at the prevailing living standards (not a particularly

stringent assumption indeed), which in turn are higher, the higher the earner/dependent ratio and/or the larger the amount of additional labor incomes that households accrue, it follows that the higher that ratio and/or the larger that amount, the closer minimum and/or reservation wages must be to the average wage, meaning that the narrower will be the feasible range of wage differentials. *Q.E.D.*

This syndrome of widening of the necessary and narrowing of the feasible wage scales, with the weakened motivating force of wages it entailed, however, was not only a consequence but, as a factor giving rise to a particular kind of substitution effect in household behavior, also a derived cause of the rising employment rate and growing proportion of second economy incomes, and thus a link in the positive feedback mechanism between the employment rate and the rate of economic activity previously described. To wit, the narrower the existing wage differentials in various dimensions (job effort, qualification, position in the job hierarchy, and so on), and the smaller the effect of a given percentage of variation of one's wage on the per capita consumption capacity of one's household, the less one is motivated to improve one's wage position (through improved performance, increased job search efforts, additional investments to one's human capital, and so on), and the more one is tempted to seek to find additional sources of income instead, that is, to increase the economic activity of one's family and oneself (in the form of a higher earner/dependent ratio, multiple job holding, more extensive exploitation of one's opportunities to earn income in the second economy, and so on). All this led to the paradoxical result of an apparently self-sparing and at the same time self-exploiting population in Hungary, that is, to a kind of aversion among them from jobs that are demanding coupled with overtly work-centered time allocation patterns.

Under the conditions just described, firms responded to the harsh competition for labor in the first economy and the extensive presence of the second economy roughly in three ways.⁵

First, they could try to improve their competitive position in the labor market by raising their wages above the market rates. This course of action was indeed prevalent among those privileged few (mostly giants, relative to the size of the market) who, thanks to their strategically strong bargaining position vis-à-vis the proper authorities, managed to obtain relief from the rules of central wage regulation. As a result of what was just said about the limited motivating potential of wage differentials, however, even they could hardly expect to attain concessions of such a scale that would enable them to achieve permanent improvements of sufficient magnitudes. They could instead expect their competitive position to improve by using their informal bargaining position to make the proper authorities adjust the occu-

pational composition of regular training to their particular needs, and by running apprentice workshops within that training.⁶

Second, in order to get hold of an enlarged pool of job applicants and reduce replacement needs arising from voluntary quits, firms could also experiment with concessions on performance and skill requirements. By doing so, they could rightly expect to appear more attractive as places of employment before those who were satisfied with their secondary earner positions in their households or had a major interest in the second economy as a source of income. Considerable concessions of this type, however, were affordable only for firms with production technologies and work organizations largely immune to the inferior working morale and poor technical skills of a more or less significant part of their labor force.

Finally, firms could try to enhance their labor market attractiveness, motivating capacity and labor retaining power by establishing internal labor markets, that is, offering part of their workers "privileges" by seniority and merits. Considering the predominance of large firms as employers most apt for the establishment of internal labor markets, and the high frequency of production disturbances and bottlenecks (due to short and uncertain supplies of inputs, worn and outdated, often idiosyncratic and ill-matched machinery, and so on) that called for a good amount of firm-specific knowledge at all levels of the job hierarchy as a prerequisite of initiative and effective daily improvisations, no wonder that internal labor markets as a form of adjustment became preponderant in the Hungarian labor market.

In many respects they resemble their market economy counterparts (entry-port jobs, long-term employment relationships with a strong bearing of the length of tenure and on-the-job training on wages and promotion, and so on⁷), whose existence is usually traced back partly to the productivity enhancing effects of specific knowledge and skills acquired while working with the firm, partly to the role a positive association between earnings and tenure as a kind of deferred payment may play in discouraging shirking and quits and attracting job applicants of superior quality, and partly to the role long-term employment relation may play as a device for firms to learn more about the productivity of their workforce and thus improve job matching at reduced screening costs. Due to the earlier described systemic characteristics of the context in which they operated, internal labor markets of Hungarian firms were at the same time distinct in several respects.⁸

First, the rules that governed the allocation and pricing of labor in these internal markets were less transparent. Instead of collective negotiations taking place before the eyes of the work force and according to more or less well-defined procedural rules and principles of representation, they were formed typically in behind-the-

scene informal and selective, occasional bargains. Job assignments and rules of promotion were similarly less transparent and less unambiguous.

Second, the fortune of these internal labor markets was tightly dependent on how far a given firm could influence those bureaucratic decisions that formed the officially recognized catalogs of crafts and occupations, the occupational composition and curricula of regular training, or the allocation of the right to run apprentice workshops.

Finally, under the conditions of the predominance of task wage (as against time wage) as a form of workers' remuneration, the positive association between earnings and tenure, a distinctive feature of any internal labor market, was in its Hungarian version based not so much on one's regular job promotions and pay advancements as on informal practices that provided preferential treatment on the shop floor for the tenured in the daily allocation of tasks of differing wage/effort ratios, and—perhaps even more important—on generous overtime assignments for them. The primary role of informal overtime provisions in the operation of the Hungarian version of internal labor markets (and in the evasion of central wage regulation) as well as the less formalized and less bureaucratic—more informal and marketlike—rules of allocation of labor and pay in these markets in general became particularly evident in the 1980s with the emergence and fast proliferation of intraenterprise work partnerships.⁹

A Paradox of Transformation: Inflexible Occupational Labor Markets

As is well known, the Hungarian economy was hit by a severe transformationary recession in the 1990s.¹⁰ Employment declined by over 25 percent—in excess of the fall of production—while real wages were eroding and registered unemployment rose to a rate of over 10 percent.¹¹ Clearly, a fall of employment in excess of the decline of production (with a corresponding improvement in productivity) would have been impossible had not a good part of the formerly established internal labor markets disintegrated and dissolved into occupational labor markets. To wit, internal labor markets in recessionary periods act as employment buffers in that they reduce the output elasticity of employment and, by so doing, tend to lower (or slow down the growth of) productivity.

That internal labor markets did in fact recede and give ground to occupational labor markets in Hungary in the 1990s should be placed primarily on account of privatization, which brought in its train the predominance of firms too small and shaky to maintain internal labor markets.¹² This is shown particularly blatantly in

the almost explosive spread of self-employment. Obviously, given the natural desire of the overwhelming majority of workers for income and job security, such an abrupt proliferation of self-employment could hardly have occurred had employment opportunities in internal labor markets not diminished drastically. Moreover, self-employment in Hungary today is in a large part only a mask that hides extremely precarious employment relations (representing the polar opposite of the kind of employment relationship characteristic of internal labor markets) whereby separation costs as well as taxes and social security contributions can be saved. The excessive multiplication of this kind of compelled (pseudo) self-employment should be seen as not a manifestation of a working population highly entrepreneurial in spirit but an indication of a temporary detour, or perhaps even worse: a low-equilibrium trap, of small-business development.¹³

Another piece of indirect evidence of the disintegration of formerly established internal labor markets and their dissolution into occupational labor markets is the withdrawal of firms from regular training: the role they would play in regular training by establishing and operating apprentice workshops in their factories (as a device to get access to a larger and selected pool of skilled young workers) is now increasingly assumed partly by workshops run by the training schools themselves, partly by microbusinesses that can thus benefit from tax relief (and also from informal fees that families of apprentice candidates without a place of practical training might be forced to pay).¹⁴ Similar conclusions can be drawn, tentatively at least, from the statistical observation that, other things being equal, those from among the unemployed who prior to unemployment were relatively low paid in their jobs—and can thus be assumed to constitute a subsample in which workers with shorter or no tenure in an internal labor market are overrepresented—if reemployed, will typically register a wage gain while the formerly better-paid reemployed will typically suffer a loss.¹⁵ As for wage employment in microbusinesses, registered employment is rare among them,¹⁶ but where it exists, it does so mostly in the form of casual/seasonal and part-time employment with practically no guarantee of job and/or income security of any kind. Furthermore, small-business owners prefer to rely on references obtainable from fellow entrepreneurs when searching for new employees¹⁷—again a practice that would clearly be foreign to an internal labor market context, where employers had better rely on their workers' recommendations.

I do not want to say that internal labor markets in Hungary have been deprived entirely of their conditions of existence. (It would also be a mistake to ignore, when talking about the predominance of internal labor markets, the role occupational labor markets played in the 1970s through the 1980s as places of employment for

those large segments of workers who were only loosely attached to the labor force and/or eager to accept less remunerative, less demanding jobs or who were simply unskilled and untrainable.) Managers of newly established Hungarian subsidiaries of multinational companies, for instance, make their best to adopt Western-type internal labor markets, as is demonstrated by their efforts to build up their own systems of training as a private source of supply of skilled labor for themselves. In other cases privatization of existing firms, where it was not followed by the introduction of new technologies, left their old-type, informal internal labor markets almost unaltered in spite of management efforts.¹⁸ What I do want to say though is that in Hungary today a much smaller proportion of workers have a chance to be employed in internal labor markets than what they were used to in the 1980s, and a much larger proportion should rest satisfied with employment opportunities occupational labor markets offer (for the most part including self-employment).

Meanwhile, however, the labor market in Hungary has become conspicuously closed and inflexible. Notwithstanding the remarkable gaining ground of occupational labor markets, it responds to economic recovery in ways (not even) a predominance of internal labor markets would imply. As a matter of fact, even under the conditions of positive economic growth and rising investment employment continued to decline last year; the relatively high rate of unemployment remains; in spite of the fact that the majority of those squeezed out of employment in recent years have withdrawn from the labor force—particularly those with little hope of being rehired later—and are now economically inactive (therefore not counted as unemployed), long-term unemployment is not diminishing; wage inflationary pressure has not lessened.¹⁹

This paradoxical behavior of the Hungarian labor market—excessive job-rationing²⁰ and wage-inflationary pressure in the context of occupational labor markets—is intellectually embarrassing indeed. More importantly, it involves the danger that the ongoing economic recovery will fuel wage inflation in a degree that calls for restrictive economic policies, with the consequence that the recovery may get stuck before growing demand for labor could bring forth any sizable increase in employment.²¹ The overtly excluding character of the Hungarian labor market, moreover, as a result of which huge masses of people devoid of entrepreneurial spirit are notoriously deprived of the possibility of wage employment and are compelled to make their living as self-employed, may contribute to the perpetuation of the current unfortunate state of affairs in which overpopulated and overfragmented small businesses are a medium of an acute danger of long-term economic dualization and thus represent a threat to rather than a catalyst of economic transforma-

tion. In order to escape these dangers it is essential to explore what causes lie behind labor-market inflexibility.

Obviously, differences in historical precedents and peculiarities of the transformational recession preclude the possibility of extracting immediately adaptable lessons and direct guidance from treatments of labor-market rigidities that market economies experienced following the oil-price explosions. At the same time, a special account of their findings, one that considers them in the light of the peculiarities of the Hungarian case, may serve as a useful starting point of exploration.

Note, first of all, that similar to the situation the majority of the developed market economies faced following the oil-price explosions, Hungary in the 1990s got acquainted with high unemployment rates coupled with high rates of inflation. This syndrome, commonly known as stagflation, was originally assumed to arise in part from anomalies of adjustment ("friction") accompanying accelerated structural changes. The larger the structural (regional, occupational, and so on) disequilibria of the labor market at a given point of time—the argument goes—the higher the rate of unemployment necessary to discipline inflationary pressure.²²

An important implication of this approach is though that with accelerated structural changes having been accomplished, stagflation is bound to disappear—an implication that was not entirely supported by later experience. True, equilibrium unemployment declined in the 1980s but not as far as its prestagflation level and with considerable variation among individual countries. This directed analysts' attention to the effects of institutional factors such as (1) labor-market corporativity, (2) path-dependency of equilibrium unemployment (usually labelled as hysteresis), or (3) internal versus occupational labor markets may exert on a country's unemployment performance. It is these factors whose responsibility for the inflexibility of the Hungarian labor market we now consider.²³

Initial attempts to find out why the unfavorable effects of the oil-price explosions lasted longer than had been anticipated, and why they varied in degree from country to country, yielded the assertion that there exists a theoretically well-arguable U-shape relationship between labor market corporativity (centralization of wage negotiations) and unemployment performance.²⁴

Centralized (macro-level) collective bargaining is claimed to have the most desirable properties on the ground that unions negotiating in such setting are in a position that both allows and enables them to take full account of the unfortunate effect excessive wage claims (ones that are inconsistent with the anticipated growth of productivity) may exert on the rate of price inflation. They will thus not easily fall victim to wage illusion on one hand. On the other hand, they are aware of the danger an

unanticipated acceleration of price inflation entails, namely, that it normally invokes restrictive economic policies as counteractions, which in turn result in declining employment, imposing involuntary unemployment on a part of their membership and increasing burdens of unemployment benefit contributions on the rest.

Micro-level collective negotiations are claimed to have favorable properties as well, which would explain the relatively good unemployment record of those countries where unions are engaged typically in firm-level bargaining. Here trade unions will feel obliged to show self-restraint, for they know that at a firm, facing highly price-elastic demand for its products, workers face highly wage-elastic employment opportunities. Consequently, unions will fear that higher wages (whether or not increasing production costs will be passed on to buyers in the form of higher prices) may lower the firm's profits (either because of falling revenues or as a consequence of lower profit margins) and induce staff decrease.

Industry-level bargaining, by contrast, is clearly the worst alternative setting as far as equilibrium unemployment implications are concerned, for industry unions negotiating separately will act like noncooperating players in a prisoners'-dilemma game. That is, they will know that none of them bears full responsibility for the impact excessive union wage claims in the aggregate make on the rate of unemployment and inflation, and because each of them will know that everybody else will think this way and act accordingly, they all are likely to end up with showing minimum restraint. Also, with demand for an industry's output being less price elastic than that of a firm (in consequence of the lower substitutability of an industry's output), industry unions have less reason to bother about the industry-level employment impact of their wage claims.

Now, in Hungary today, taking the as yet rudimentary state of industrial relations into consideration,²⁵ one should be skeptical about the merits of an explanation of labor market inflexibility that would draw too heavily on the earlier arguments. Be they differential equilibrium unemployment impacts of different degrees of corporativity as claimed, the most one can extract as lessons relevant to the Hungarian case are "ifs." One can only say that had macro-level tripartite negotiations turned out to be more productive, and/or had the implementation of some kind of Western-type income policy (that would have filled the gap the hasty elimination of central wage regulation in 1990 had left behind) been tried and found effective as a device to moderate wage inflation, and the like, then Hungary might have tided over the economic transformation at lower costs, that is, with less restrictive economic policies and thus fewer job losses.²⁶ Similarly, equilibrium unemployment in the years ahead could in principle be favorably affected by a macro-level collective agreement in which the parties would commit themselves to moderate wage inflation.

Realistically, however, it is exactly the present rudimentary state of industrial relations that makes it hard to believe in the feasibility of such an agreement. Hungary's strongly divided trade unions, keenly competing for a shrinking constituency base, can hardly be expected to enter into such an agreement. Do they still become resigned to doing it, it would be naïve to assume that they will indeed be up to consistently meeting the negotiated terms.

Similar doubts can be cast on the merits of the oldest (but still quite recent) versions of path-dependency arguments²⁷ as far as their relevance to Hungary is concerned. True, they start out from the assumption that trade unions act as partial representatives of insider workers' interests—a kind of partiality that is clearly manifest in union behavior in Hungary as well. This partiality, however, could produce hysteresis only insofar as our unions were massively involved in wage setting.

As against these early explanations, later interpretations of this same phenomenon are based on assumptions that do seem directly applicable to the Hungarian case. Namely, they assume that either employers are (for good reason or not) reluctant to hire from among the long-term unemployed or (if only for this reason) the long-term unemployed lack the confidence necessary to effectively compete for jobs, or both. They argue that because increasing unemployment tends to increase the incidence of long-term unemployment—the proportion of long-term unemployed in total unemployment—it causes an upward shift in equilibrium unemployment, that is, an increase in wage inflation at any given level of total unemployment.²⁸

True, these latter interpretations, too, examine wage setting in a collective bargaining context. Nevertheless, they may serve as a suitable reference in finding out the causes of labor-market rigidity in Hungary, for, in the first place, they allow, implicitly at least, for the possibility that, thanks to the high proportion of long-term unemployed in total unemployment, large segments of insider workers remain defended against the unfavorable impacts of massive unemployment—even in the lack of strong union support. Moreover, their reasoning would even account for the kind of low-employment equilibrium-trap situation where the disciplinary (wage-inflation moderating) effect of an initial increase in total unemployment has been more than offset by the sufficiently large increase of long-term unemployment it called forth. Yet, what would be essential in the adoption of these interpretations to the Hungarian case is to address the issue of what (presumably institutional) factors determine whether a given (increase in the) rate of unemployment will be associated with larger or smaller (growth of the) incidence of long-term unemployment.

Finally, according to a third characteristic explanation, the observed dependency of current rates of equilibrium unemployment on past rates of unemployment has

to do primarily with capacity utilization fluctuating in the course of a business cycle.²⁹ Namely, with capacity utilization having declined in the initial phase of a recession, a fall of investments and an increase in capital scrapping are bound to follow—with the result that capital stock will shrink. Granting this, when recovery comes it soon leads to an overutilization of the undersized capital stock, with the consequence that price and wage inflation will start to accelerate at a relatively low rate of employment (and a correspondingly high unemployment rate).

Now, in view of the enormous magnitude of the reduction of capital stock that accompanied the transformational recession in Hungary, this argument deserves serious consideration. At the same time, it should be kept in mind that hysteresis of such origin, unless the rise of inflation thus induced is misdiagnosed and suppressed by hastily adopted restrictive economic policy measures, is transitory by nature, and that timely anticipation of an imminent and fast recovery may give impetus to business investments early enough to simultaneously accelerate the upswing and prevent an undue transitory rise of inflation.

Although traces of the ideas reviewed and reflected on in the foregoing appear incidentally in the study of the ongoing transformation of the Hungarian labor market, the marked and sudden contraction of the domain of internal labor markets and its bearing on the prospects of economic transformation have somehow remained beyond our range, leaving obvious parallels with the market-economy experience unrecognized and unexploited. The section that follows is meant to amend this negligence.

Labor Market in Institutional Imbalance: A Clue to the Riddle of Inflexible Occupational Labor Markets

Cross-country comparative labor-market studies usually reveal marked correspondence between the historically established patterns of coexistence of internal versus occupational labor markets and certain aspects of the institutional infrastructure that surrounds the labor market (vocational training, statutes of labor, industrial relations, employer and employee attitudes and norms of behavior, and so on).³⁰

The instrumentality of this correspondence becomes obvious in the light of the costs parties to an employment contract—workers and firms—incur if they terminate their contract: the lower these transaction (separation/replacement) costs, the more viable occupational labor markets are.³¹ These separation/replacement costs are in turn lower the closer substitute job offers of different employers are for a worker with a given qualification, and the closer substitute workers with a given

qualification are for a firm, that is, the more standardized the skill requirements of a given occupation and the actually acquired skills of workers with a given qualification are. Now, the more standardized occupational skills and skill requirements are, the more reasonable it is that occupational skills be acquired in specialized training institutions obliged to respect the established occupational standards in their subject matters of instruction, rather than in the form of on-the-job training. Furthermore, under the conditions of low separation/replacement costs that highly standardized occupational skills acquired in specialized training institutions entail, workers can best promote their interests collectively by forming or joining unions that are organized along the lines of crafts and occupations, and by pressing for collective agreements that include well-defined job demarcation lines and formal qualification criteria of admittance that are uniform throughout the economy.

With an internal labor market, by contrast, a two-pillar system of vocational training is compatible whereby basic occupational skills acquired at school are complemented by a large amount of special skills, varying in content from firm to firm, that can only be acquired while working. In view of the long duration of employment contracts and the strong bearing of the length of tenure and on-the-job training on earnings and promotion, workers can best advance their interests collectively by forming or joining unions that are organized at the level of the firm (or group of firms within the same branch of the economy), and by pressing for collective agreements that include well-defined and enforceable rules of hiring, firing, and promotion (possibly together with other kinds of job property-rights provisions).

Whether the configuration of the labor market and its institutional infrastructure in a given country at a given time is more of internal or occupational type is the result of a historic process of institutional evolution. In the course of that process, tensions arising from the perception of unexploited opportunities and/or unfulfilled expectations by workers and firms (their dissatisfaction with one another's attitudes and behavior, with the performance of training institutions, with the prevailing system of collective bargaining, and so on) would call forth institutional adjustments that gradually eliminate inappropriate institutional elements. Presumably it is mainly these gradual adjustments that eventually produce a constellation that is largely self-perpetuating (with compatible and mutually reinforcing opportunities and expectations) even amid extreme swings in economic environment.

The particular equilibrium that such a stable, self-perpetuating institutional configuration represents should be respected by labor-market policies as well (themselves coming within the institutional infrastructure of the labor market): it constrains the choice of the effectively adaptable set of policy instruments. Faced

with a predominantly internal labor market configuration, for instance, active labor market policies should take into account that participation in retraining programs cannot procure tenured jobs for the prime-age unemployed, and to the extent this is known to them, they may be reluctant to undergo retraining. In a configuration where the typically occupational labor markets of small firms dominate, on the other hand, granting wage subsidies to employers hiring from the unemployment pool will not be cost efficient either: nonrenewed, short-term employment relations and high turnover rates will result.

As for the present situation of the Hungarian labor market, it is far from any kind of a self-reproducing institutional equilibrium. As a matter of fact, the remarkable gaining ground of occupational labor markets has not been accommodated by the institutional environment. Namely, parallel with the liquidation of an increasing part of firms' apprentice workshops, the kinds of skills and qualifications acquirable in vocational training have become ever less transparent and standardized; lack of coordination and supervision has led to an enormous inflation of certificates of dubious credence and market value.³² Moreover, falling investments consequent to the recession, by hindering the process of large-scale replacement of obsolete capacities, retarded the devaluation of job-specific skills and thus the diminution of separation costs. Trade union organization along the lines of crafts and occupations has not gathered momentum either. Or, although the expectations of the main army of the unemployed bear the imprints of their earlier internal labor market experience, too many of the newly trained face as job applicants the disheartening experience of distrust toward their qualification certificates on the part of employers,³³ among other difficulties.

Taking established market economies as a reference, the gaining ground of occupational labor markets in Hungary in the course of transformation should thus be assessed as an irregular, uneven process. In the period of transformational recession, the role of internal labor markets as employment buffers has been reduced (perhaps, even beyond what one would regard as normal in a market economy) without the conditions of operation of occupational labor markets (i.e., low separation/replacement costs that could have effectively moderated wage-inflationary pressure by intensifying competition between insider and outsider workers, and so on) and thus a later expansion of employment having been created. Such being the case, active labor market policies (like massive retraining programs or generous wage subsidies) could as a matter of course not be particularly effective either.³⁴

To conclude this section, the main lesson to be drawn from the preceding assessment of the Hungarian situation is that to the extent occupational labor markets retain their prominence in the years to come, a prerequisite of the degree of labor

market flexibility and policy efficiency necessary for sustainable economic growth is the formation of an appropriate institutional infrastructure, that is, one that fits and confirms the logic of operation of occupational labor markets. This could, perhaps, best be promoted here and now by redesigning the system of vocational training in a way that could yield transparency. The work should possibly start with the specification of the skill requirements of occupations and then continue with the elaboration of their standard cores of instruction—in themselves a huge task, which could hardly be carried out successfully without government assistance or involvement of organizations representing workers' and employer's collective interests. The point is that only to the extent the credibility of officially recognized qualification certificates issuable by training institutions is reestablished can the expectations of workers and employers gradually adjust to the principles of operation of occupational labor markets and get reinforced by fulfillment.

Rights and Wrongs of Alternative Institutional Arrangements

In our searching into the causes of the apparent closedness and inflexibility of Hungary's postsocialist labor market, the dimension of internal versus occupational labor market was highlighted. The essence of the diagnosis thus established is that the rapid disintegration of the particular internal labor markets of the 1980s and their dissolution into occupational labor markets in the 1990s have not been matched by a corresponding transformation of their institutional infrastructure, which produced a kind of institutional imbalance. It is fundamentally this institutional imbalance that I blamed for why, after occupational labor markets having become predominant (without which the dramatic decline in waged employment that was experienced in the 1990s could not indeed have taken place), today's Hungarian labor market responds to the initial stage of economic recovery in ways—namely, wage and employment tendencies—characteristic of internal labor markets, thereby endangering lasting economic growth and small-business modernization.

I noted that predominance of internal or occupational labor markets does not mean their exclusiveness. I conceive it in a way similar to how Kornai thinks of the predominance of state versus private ownership as the basis of bureaucratic versus market coordination.³⁵ In the same way that it is not possible to specify exactly what proportion of state ownership is just large enough to exclude the predominance of market coordination—obviously, this critical proportion will depend on, among other things, how far state ownership penetrates into different kinds of economic activities—it cannot be determined specifically in what proportion jobs should be

of this or that type so as to entitle one to speak of the predominance of a given labor market type.³⁶ The most one can say is that in a proportion large enough to bind the infrastructural institutions to conform to the logic of operation of that particular labor market type, and that in this respect it is primarily skilled manual (non-managerial) jobs in the market sector of the economy whose distribution by labor market types matters.³⁷

Before granting that infrastructural institutions (vocational training, collective bargaining, statutes of labour, expectations of employers and workers, active labor market policies, and the like) should therefore conform to the logic of operation of the predominant labor market type, one must pose the question of whether this would not incapacitate the other type and thus result in a degree of institutional homogeneity that, while promoting short-term efficiency, endangers long-term adaptability.³⁸

The answer to this question should differentiate according to how unambiguously workers of different crafts and occupations are separated from each other with respect to the types of labor markets their jobs represent, that is, how segmented labor markets are in this respect.³⁹

Among crafts and occupations where this kind of segmentation is sharp—jobs belonging to them are associated unambiguously with this or that labor market type—the principles of operation of their respective labor market types can be expressed directly in the heterogeneity of the institutional infrastructure.

Yet, such a high degree of segmentation by occupation is an exception rather than a rule—at least among the crafts and occupations we are primarily interested in. Part of those completing regular training and entering the labor force with a given qualification get employed, temporarily or lastingly, in the occupational labor market, others in internal labor markets. Would not an institutional infrastructure adjusted to the logic of the dominant type of labor market of a given occupation lead in that occupation to the incapacitation of the other labor market type and thus eliminate institutional heterogeneity?

True, mutually reinforcing interactions taking place between the dominant type of a labor market and an institutional infrastructure that conforms to it are such that they have an unfavorable impact on the costs of operation of the nondominant type (for instance, they impose excess costs of training, search, collective bargaining, and such on those who would rely on it), inhibiting it from ever taking over predominance. Nevertheless, it can subsist in its nondominant position and thus contribute to the perpetuation of institutional heterogeneity insofar as its comparative advantages are large enough to compensate for these excess costs. (For instance, a firm can benefit from an internal labor market even in an occupational

type institutional environment inasmuch certain individual characteristics of workers are of vital importance but can only be learned after long years of service, or as long as a low turnover rate is worth these excess costs for some other reasons.)

Although it might be redundant, I will conclude by noting that the concept of labor market type was used all along in the foregoing as a descriptive category devoid of any normative connotation. The reason I feel that such an ending is appropriate is twofold, namely: (1) I am afraid that the literature of labor-market segmentation (or its loose reading) may make someone associate the concept of internal labor market with the positive notion of a submissive version of capitalism, or of the upper segment of the labor market, and the concept of occupational labor market with the negative notion of rough (wild) capitalism, or of the lower segment of the labor market, and see on this ground the preponderance of internal labor markets as a desirable goal in its own right.⁴⁰ (2) My insistence on the formation of an institutional infrastructure adjusted to the logic of occupational labor market as a way to the flexibilization of the Hungarian labor market may seem to be motivated by an opposite commitment: a predilection for unbridled competition of *laissez-faire* capitalism.

From the observation that poorly paid, arduous jobs devoid of promotion prospects are located, as is emphatically stressed in the literature of labor market segmentation, primarily outside internal labor markets, it does not follow that it is the logic of operation of occupational labor markets per se that accounts for the accumulation of these unfavorable job characteristics. Suffice it to say that holders of some prestigious degrees of high market value view exactly occupational labor markets as the most suitable institutional setting for them to make a career. Realistically, both types of labor markets have, as I pointed out incidentally in the preceding, right and wrong sides of distinctively their own.

A comparative advantage of occupational labor markets as against internal labor markets is that owing to the more flexible adjustment of employment to production fluctuations, economic recessions are less likely to involve falling productivity and deteriorating international competitiveness. Furthermore, job losses are less likely to result in long-lasting unemployment, and, if only for this reason, an early acceleration of wage inflation is less of a danger in a later recovery. At the same time occupational labor markets have the comparative disadvantage of an inferior working morale, and a stronger resistance to technological innovations that entail large-scale modifications in the skill requirements of an occupation. These are less of a problem in an internal labor market, given the possibilities of workers' internal promotions, reallocations, and retraining. Or, owing to low separation costs, workers employed in an occupational labor market can quit with ease if they are made a better job

offer but can be laid off with ease as well. In an internal labor market, by contrast, workers enjoy higher job security and more calculable career prospects, but at the price of facing substantial losses if it comes to dismissal.

It is more important than going further with the enumeration of the right and wrong sides of the respective types of labor market, from the viewpoint of our topic, to emphasize that, on one hand, their comparative advantages can only unfold in an institutional environment that fits their particular logic of operation; otherwise their advantages remain latent and only their disadvantages become apparent. On the other hand, no such thing can exist like a labor market that would be characterized by the comparative advantages of either of the two types untinged with their disadvantages, that is, one that would combine the respective advantages of both. To be sure, I do not mean to say that their unfortunate effects cannot be restricted. Active labor market policies should serve for exactly this purpose—but again, because they themselves come within the institutional infrastructure of the labor market, they should comply with the requirements of institutional coherence in order to be effective.

Granting that neither of the two types of labor market is superior to the other in terms of their social and/or economic impacts, why do I favor the emergence of an institutional infrastructure that would conform to the logic of occupational labor markets? After all, I could start out from the assumption that, facing an institutional infrastructure seemingly inconsistent with the logic of operation of occupational labor markets, employers will increasingly resort to building up their own internal labor markets, with the result that the present anomalies of the disrupted institutional balance of the Hungarian labor market will gradually be eliminated spontaneously.

My position is dictated by neither a concealed taste for occupational labor markets nor even my rejection of the earlier assumption as being for now unrealistic. Instead, it is based on a confrontation of the above described comparative advantages and disadvantages of the respective types of labor market, and the lessons that were drawn in the equilibrium-unemployment literature after the oil-price shocks, with my diagnosis of the overtly excluding character and inflexibility of the Hungarian labor market. This confrontation suggests to me—and this is what I have tried to demonstrate throughout my chapter—that considering the urgency of a marked reduction of wage inflation as a condition of any sizeable expansion of wage employment (this latter being in turn, as was pointed out, a necessary condition of small businesses embarking on the road to modernization), it is the predominance of occupational labor markets that promises a smoother way to economic development *here and now*.

Notes

1. This straightforward proposition, to which I take the courage from the conception of affinity of ownership forms and modes of coordination put forward in Kornai (1984) (and confirmed in Kornai 1992—cf. his comments on figure 15.1 in chapter 5), will be qualified in the following paragraphs in accord with the relevant arguments of chapters 11 and 16 of Kornai (1980) and of chapters 10 and 23 of his (1992).
2. This interpretation of the role of earnings differentials in Hungary goes as far back as MTA KTI (1971), a hotly debated study of the time written by a collective of reform-minded young economists, which remonstrated against glorification of workers' loyalty to their firms and denunciation of "migratory birds," and pointed out reasonable motives and favorable aspects of the high rate of workers' inter-enterprise mobility that had marked the first two years of the New Economic Mechanism. With full respect to the authors' bold protest against the imposition of artificial constraints on labor mobility and the stigmatization of workers' voluntary quits, I focus attention on those systemic constraints that came to be recognized later as setting objective limits to the potential role of wage differentials in the coordination of labor allocation and in the motivation of workers.
3. Excess demand for labor prevailed even under the conditions of stagnation in the late 1980s, although employment (but not the rate of employment) declined somewhat. For an analysis of the development and fluctuations of labor demand, see Fazekas and Köllő (1990).
4. In the following discussion of these spontaneous mechanisms I draw on Kornai (1980) and (1992) (particularly on the chapters referred to in note 1).
5. The ways and means of adjustment of Hungarian firms to the conditions of labor shortage are a main theme in Fazekas and Köllő (1990), where their findings of more than a decade of labor-market research are presented.
6. For further details consult with Fazekas and Köllő (1990).
7. For the characteristics of internal labor markets in a market economy setting, see Osterman (1984).
8. Their rules of operation, with particular attention to formal and informal mechanisms of wage determination, were brought to light in a large part in the late 1970s and early 1980s by shop-floor case studies carried out at large industrial and service firms by a research group headed by Gábor Révész. (For their findings see the respective chapters by Károly Fazekas and János Köllő in Galasi (1982) and in the working papers collected in Fazekas et al. [1983].) Their findings were confirmed and complemented by further empirical evidence and interpreted as indicating systemic distinctiveness by Stark (1986).
9. See, for example, Neumann (1988) and Stark (1989).
10. The arguments that follow take the views and ideas advanced by Kornai (1993), (1995), and (1997) on the causes of the transformational recession and conditions of recovery as a reference but are confined to aspects strictly related to the labor market.
11. János Timár was among the first who took notice of the fact that compared to the dramatic decline of employment that was brought about by the transformational recession (as a result of which Hungary's employment rate fell to a level even lower than is typical among market economies) the magnitude of the increase of unemployment was disproportionately small in absolute terms. (See Timár [1994].)
12. Smith (1988) finds that only firms large enough (with a staff of 500 at least—or, according to Oi (1983), 1000) establish internal labor markets.
13. This is suggested by the apparent inconsistency of the relatively low rate of failure and retreat of our overpopulated and overfragmented small-business sector: as many as two-thirds of those small and mostly young undertakings that were surveyed in 1993 were still in business in 1996. (See Lengyel [1996].) Moreover, the death rate of those among them that had been started under the pressure of necessity was below that of the total sample. And although the size of the units (measured on their staff) that had survived increased somewhat during that three-year period, this did not hold for the small-business sector as a whole (see figures in note 16). For my view of the current state of health of the Hungarian small-business sector and its developmental prospects see Gábor [1997].

14. Although the number of apprentices enrolled in regular vocational training dropped by 17 percent between 1990–1991 and 1993–1994, those among them who attended training shops run by schools even increased by 58 percent while the number of apprentices trained in groups at firms decreased by 49 percent. At the same time participation in other (nongroup) forms of regular training grew by 12 percent (Benedek [1994]).
15. This observation is reported (and given a hypothetical interpretation that, unlike the one I am proposing, relies on the theory of statistical discrimination) in Köllő and Nagy (1995).
16. According to figures reported in Laky (1996), the average number of employees in unincorporated small undertakings of single owners—typically own-account workers—did not change from 1993 to 1995: it remained as low as 0.3. As for corporations, the ratio of those with a staff of fewer than 20 even increased, from 72.5 to 87 percent, between 1991 and 1995. A particularly striking increase could be observed in the ratio of the smallest corporations: the proportion of those with a staff of 10 or fewer grew from 47.6 to 74.8 percent in the two-year period of 1993 to 1995.
17. See Laki (1997).
18. This was demonstrated by case studies—e.g., Bódis (1996, 1997) and Janky (1996)—of an ongoing research project (entitled “Economic sociological investigation of hierarchical economic organizations”) informed by the ideas of Csontos (1994) and coordinated by him until his early death in 1997.
19. For a solid and carefully documented general account of the labor market impacts of the Hungarian economic transformation see Galasi and Kertesi (1996).
20. The phenomenon of job rationing as a regime in which otherwise apt outside applicants are widely denied jobs even despite their lower wage claims is traced back by efficiency wage theories in part to workers’ unobserved heterogeneity. This being the case, and considering the fact that the existence of internal labor markets are also traced back in part to this heterogeneity, it follows that efficiency wages and internal labor markets are in this respect conditional upon each other. (Cf. Weiss [1990].) Note also that the role internal labor markets may play in maintaining a regime of excessive job rationing is tacitly acknowledged in the so-called insider-outsider approach by Lindbeck and Snower (1988) as well, in which insider workers, understood as “full-fledged employees,” are claimed to be able and willing to make underbidding by the outsiders expensive for the employers to accept and/or disagreeable for the outsiders themselves to engage in.
21. In his discussion of the role the development of wages and consumption played in the transformationary recession, and of the kind of income policy to be pursued in the recovery, Kornai repeatedly calls attention to this danger—see Kornai (1993, 1995, 1997). (For a markedly different position on the same issues see Matolcsy [1997].)
22. That is, the higher the natural rate of unemployment in a competitive economy; and the higher the equilibrium rate of unemployment (NIRU or NAIRU) in imperfect competition—understood as the rate at which workers’ real wage claims and firms’ real profit claims are consistent with each other and, as a result, price-inflation expectations are fulfilled. (Note that the theory of imperfect competition is itself a product of the 1980s; see Blanchard [1986] and Blanchard and Summers [1988].) Because the arguments below refer to the conditions of imperfect competition, the concept of equilibrium unemployment is adopted hereafter.
23. References to the literature on which the relevance of these factors for Hungary is considered are placed in footnotes at the respective sections.
24. See Calmfors and Driffil (1988).
25. For facts substantiating this unfavorable general assessment see Ladó and Tóth (1996).
26. In his assessment of the results of the much discussed corrective measures of the Bokros package, Kornai (1997) himself sees it as a fortunate accident, presumably the consequence of a surprise inflation, that real wages declined considerably in 1995.
27. See Blanchard and Summers (1986, 1991).
28. See Layard and Nickell (1986). Although this line of reasoning appears to differ fundamentally from the previous one in its departure, as Lever (1995) pinpoints: depending on how broadly or restrictively

the notion of “outsiders” is defined—whether it is meant to refer to the long-term (how long?) unemployed only or, more broadly, to the unemployed, or it should embrace nonunion member employees and/or workers in noninternal labor market jobs as well—can one regard the two lines of reasoning as theoretically more or less distinct from each other and attempt to decide (possibly with the help of econometric analysis) which of the two fits reality better.

29. See Soskice and Carlin (1989).

30. See especially Eyraud et al. (1990). Also see Hashimoto and Raisian (1988).

31. These costs include the so-called quasi-fixed costs of employment born by employers (costs of recruitment and selection, direct and indirect costs of training of newly hired workers, severance payments, and so forth) as well as those direct and indirect costs that workers incur in case of a change in their place of employment (search costs, reduction in one's expected lifetime earnings, and the like), which might be substantial in internal labor markets and negligible in occupational markets. (Cf. Oi [1983].) Note that equilibrium unemployment is also seen as a function of institutional factors—namely, of the existence of severance payment, a characteristic of internal labor markets—in the multiple-equilibrium model set up by Saint-Paul (1995) in the spirit of Blanchard and Summers (1988). In that model severance payment has an unfavorable impact on equilibrium unemployment partly by discouraging separations, partly by strengthening the bargaining position of (insider) workers. This gives rise to the conjecture that there is an overlapping between the first two of those explanations of high equilibrium unemployment, which were discussed under the heading of hysteresis unemployment and ones that concentrate on the effects of institutional factors related to the type of the labor market.

32. To refer back to regular vocational training, what lies behind the category of “training not in groups” as a form of training currently in expansion is in a large part training in technologically obsolete and/or narrowly specialized shops in the small-business sector, which are for this reason alone inopportune to train apprentices in the niceties of an occupation.

33. That those who enter the labor market after completing their studies in regular vocational training have nevertheless relatively good chances to get a job can partly be explained by their willingness to accept jobs that were and/or could be filled with unskilled workers. (This explanation is suggested in chapter 5.3 of ILO [1997].) It is indeed reasonable to assume that to the extent school and/or qualification certificates are unsound signals of one's acquired general knowledge and/or occupational skills, employers are induced to hire formally overeducated and/or overskilled workers.

34. Based on a wide range of unemployment literature, Falusné (1991) called attention to the important role active labor market policies would have to play already in the initial stage of economic transformation, and attempted to specify the main “strike directions” that she believed could be appropriate under the constraints of transformation. Yet, her expectations, although moderate compared to the views prevailing that time, turned out to be too ambitious in the light of later experience.

35. See, for example, Kornai (1984).

36. The parallel may apply also to the marketlike versus bureaucratic nature of employment contracts in that employment relations in an internal labor market—although any employment relationship is inherently a dominantly market-type, transactive relationship, in which employees as owners of their labor undertake certain tasks for certain remuneration—are more bureaucratic in nature: employers retain more discretion in managing the labor power of their employees (in exchange for providing them guarantees for job security and promotion) than is the case in an occupational labor market. (For a transaction-costs interpretation of the bureaucratic aspects of internal labor market-type employment relations, see the seminal paper by Williamson et al. [1975]; other interpretations view employment relationship as a particular kind of principal-agent relationship and stress the problem of moral hazard; again others, following the footsteps of Hirschman [1970], point out the importance of “voice” as a political-institutional mechanism of adjustment characteristic of internal labor markets, as against “exit” as a competitive mechanism characteristic of occupational labor markets; see, e.g., Freeman [1980].) Occupational (internal) labor markets may thus be seen as representing a configuration of ownership form and mode of coordination closer to (farther away from) the pole of pure private ownership and pure market coordination.

37. It is by and large this range of jobs to which the comparison of the role internal versus occupational labor markets play in Britain and France is confined in Eyraud, Marsden, and Silvestre (1990).
38. Based on a wide selection of the relevant literature, Grabher and Stark [1997] convincingly demonstrate the importance of organizational diversity in the evolutionary potential of economic institutions in general and under the conditions of postsocialist transformation in particular.
39. For a review of the recent literature on labor-market segmentation, see McNabb and Ryan (1989).
40. Ironically, though, it is exactly internal labor markets where Bartlett (1989) thinks employers have excessive negative (i.e., punitive) event power over their employees. (See his vehement refutation in chapter 7 of the idea of regarding employment relation as a series of bilateral spot contracts with neither of the parties having any real authority over the other—an idea that was put forward in his interpretation by Alchian and Demsetz [1972].)

References

- Alchian, A., and H. Demsetz (1972). Production, information costs, and economic organization. *American Economic Review* 62: 777–795.
- Bartlett, R. (1989). *Economics and Power: An Inquiry into Human Relations and Markets*. Cambridge, Mass.: Harvard University Press.
- Benedek, A., ed. (1994). *Szakképzés Magyarországon* (Vocational Training in Hungary). Budapest: Ministry of Labor.
- Blanchard, O. J. (1986). The wage-price spiral. *Quarterly Journal of Economics* 101: 543–565.
- Blanchard, O. J., and L. Summers (1986). Hysteresis and the European unemployment problem. In S. Fisher, ed., *NBER Macroeconomics Annual* vol. 1, pp. 15–77. Cambridge, Mass.: NBER.
- Blanchard, O. J., and L. Summers (1988). Beyond the natural rate hypothesis. *American Economic Review* 78: 182–187.
- Blanchard, O. J., and L. Summers (1991). Hyteresis in unemployment. In N. G. Mankiw and D. Romer, eds., *New Keynesian Economics*. Cambridge, Mass.: MIT Press.
- Bódis, L. (1996). *Foglyok és dilemmák. Egy egyedi gépgyár munkaszervezetének és vezetésének néhány jellemzője a privatizáció után* (Prisoners and dilemmas: some characteristics of the work organization of a non-series machine-producing factory after privatization). Mimeo.
- Bódis, L. (1997). Privatizáció, munkaszervezet és bérelosztási mechanizmusok egy nagyüzemi varrodában (Privatization, work organization and mechanisms of wage distribution in the sewing workshop of a large firm), Parts I–II. *Közgazdasági Szemle* 44: 698–717 and 799–818.
- Calmfors, L., and J. Driffil (1988). Centralisation of wage bargaining. *Economic Policy* 6: 13–61.
- Csontos, László (1994). Szempontok a privatizáció és a hierarchikus gazdasági szervezetek belső szerkezetének változásai közti összefüggések tanulmányozásához (Considerations for the study of the impacts of privatization on the internal structure of hierarchical business organizations). *Szociológiai Szemle* 4: 83–95.
- Eyraud, F., D. Marsden, and J-J. Silvestre (1990). Occupational and internal labour markets in Britain and France. *International Labour Review* 129: 501–517.
- Falusné, K. Sz. (1991). A foglalkoztatáspolitikai fő irányai (The main directions of employment policy). *Közgazdasági Szemle* 38: 1–17.
- Fazekas, K., I. Kalász, J. Köllő, M. Polgár, G. Révész, and Gy. Sziráczi (1983). *Kereseti és bérvizonyaink* (Earnings and Wage Situation in Hungary). Budapest: MTA Közgazdaságtudományi Intézetének Közleményei. no. 28.
- Fazekas, K., and J. Köllő (1990). *Munkaerőpiac tőkepiac nélkül* (Labor market in the absence of capital market). Budapest: Közgazdasági és Jogi Könyvkiadó.

- Freeman, R. B. (1980). The exit-voice tradeoff in the labor market: unionism, job tenure, quits, and separations. *Quarterly Journal of Economics* 100: 644–673.
- Gábor, I. R. (1997). Too many, too small: small entrepreneurship in Hungary—ailing or prospering? In G. Grabher and D. Stark, eds., *Restructuring Networks in Post-Socialism: Legacies, Linkages, and Localities*. Oxford: Oxford University Press.
- Galasi, P., ed. (1982). *A munkaerőpiac szerkezete és működése Magyarországon* (The structure and functioning of the Hungarian labor market). Budapest: Közgazdasági és Jogi Könyvkiadó.
- Galasi, P., and G. Kertesi, eds. (1996). *Labour Market and Economic Transition: Hungary, 1986–1995*. Budapest: ILO Japan Project.
- Grabher, G., and D. Stark (1997). Organising diversity: evolutionary theory, network analysis and post-socialism. In G. Grabher and D. Stark, eds., *Restructuring Networks in Post-Socialism: Legacies, Linkages, and Localities*. Oxford: Oxford University Press.
- Hashimoto, M., and J. Raisian (1988). The structure and adaptability of short run labor markets in Japan and the United States. In R. A. Hart, ed., *Employment, Unemployment and Labor Utilization*. Boston and London: Unwin and Hyman.
- Hirschman, A. O. (1970). *Exit, Voice, and Loyalty: Responses to Decline in Firms, Organizations, and States*. Cambridge, Mass.: Harvard University Press.
- ILO (1997). *Országfelmérés a foglalkoztatásról és a fenntartható életmódról* (Country report on employment and sustainable ways of living). Budapest: Nemzetközi Munkaügyi Szervezet. (ILO).
- Janky, B. (1996). Övé a gyár—magadnak építed! Munkaszervezet és privatizáció (The factory is theirs—build it for yourself! Work organization and privatization). *Szociológiai Szemle* 3–4: 223–248.
- Kornai, J. (1980). *Economics of Shortage*. Amsterdam: North-Holland, 1980. Also in Hungarian (1980, 1982, 1989), Czech (1981–1982), French (1984), Polish (1985), Chinese (1986), and Russian (1990).
- Kornai, J. (1984). Bureaucratic and market coordination. *Osteuropa Wirtschaft* 29(4): 316–319. Also in Hungarian, Japanese, Chinese, Estonian, and Russian.
- Kornai, J. (1992). *The Socialist System: The Political Economy of Communism*. Princeton, N.J.: Princeton University Press, 1992; Oxford: Oxford University Press, 1992. In Hungarian (1993), German (1995), French (1996), and Bulgarian (1996).
- Kornai, J. (1993). Transformational recession: a general phenomenon examined through the example of Hungary's development. *Economie Appliquée* 46(2): 181–227. Also in Hungarian, German, Czech, Polish, Russian, and French.
- Kornai, J. (1995). The dilemmas of Hungarian economic policy. In Béla K. Király and András Bozóki, eds., *Lawful Revolution in Hungary 1989–94*, pp. 323–349. New York: Columbia University Press.
- Kornai, J. (1997). Adjustment without recession: a case study of Hungarian stabilization. In Salvatore Zecchini, ed., *Lessons from the Economic Transition: Central and Eastern Europe in the 1990s*, pp. 123–152. Dordrecht: Kluwer Academic. Also in Hungarian.
- Köllő, J., and Gy. Nagy (1995). Bérek a munkanélküliség előtt és után (Wages before and after unemployment). *Közgazdasági Szemle* 42: 325–357.
- Ladó, M., and F. Tóth, eds. (1996). *Helyzetkép az érdekegyeztetésről, 1990–1994* (Situation report on interest reconciliation, 1990–1994). Budapest: Érdekegyeztető Tanács Titkársága—PHARE Társadalmi Párbeszéd Project.
- Laki, M. (1997). *Kisvállalkozás a szocializmus után* (Small entrepreneurship in Postsocialism). Budapest: MTA Közgazdaságtudományi Intézet.
- Laky, T. (1996). A magángazdaság kialakulásának hatásai a foglalkoztatottságra 1995-ben (Effects of the formation of private economy on employment in 1995). *Közgazdasági Szemle* 43: 629–652.
- Layard, R., and S. Nickell (1985). The causes of British unemployment. *National Institute Economic Review* (February): 62–85.

- Layard, R., and S. Nickell. (1986). Unemployment in the UK. *Economica* 53: 121–170.
- Lengyel, Gy., ed. (1996). Megszűnt és működő vállalkozások, 1993–1996 (Disfunt and operating business undertakings, 1993–1996). Mimeo.
- Lever, M. H. C. (1995). Insider-outsider effects in wage formation: an empirical survey. *Bulletin of Economic Research* 47: 257–274.
- Lindbeck, A., and D. Snower (1988). *The Insider-Outsider Theory of Employment and Unemployment*. Cambridge, Mass.: MIT Press.
- Matolcsy, Gy (1997). Kiigazítás recesszióval. Kemény költségvetési korlát és puha piaci korlát (Correction with recession: hard budget and soft market constraint). *Közgazdasági Szemle* 44: 782–798.
- McNabb, R., and P. Ryan (1989). Segmented labour market. In Z. Tzannatos and D. Sapsford, eds., *Current Issues in Labour Economics*. London: Macmillan.
- MTA KTI (1971). A munkaerő-vándorlás indítékai és hatásai az új mechanizmusban. A fiatalok munkaerő-vándorlásának sajátosságai (Motives and effects of labor turnover in the new mechanism: particularities of labor turnover among the young). *Közgazdasági Szemle* 18: 805–821.
- Neumann, L. (1988). Piaci viszonyok az üzemi béralkuban? (Market relations in shop-floor wage-bargaining?). *Gazdaság* 22: 64–79.
- Oi, W. (1983). The fixed employment costs of specialized labor. In J. Triplett, ed., *The Measurement of Labor Cost*. Chicago: University of Chicago Press.
- Osterman, P., ed. (1984). *Internal Labor Markets*. Cambridge, Mass.: MIT Press.
- Saint-Paul, G. (1995). The high unemployment trap. *Quarterly Journal of Economics* 90: 527–550.
- Smith, R. (1988). Comparable worth: limited coverage and the exacerbation of inequality. *Industrial and Labor Relations Review* 42: 227–239.
- Soskice, D., and W. Carlin (1989). Medium-run Keynesianism: hysteresis and capital scrapping. In P. Davidson and J. Kregel, eds., *Macroeconomic Problems and Policies of Income Distribution*. Aldershot: Edward Elgar.
- Stark, D. (1986). Rethinking internal labor markets: new insights from a comparative perspective. *American Sociological Review* 51: 492–504.
- Stark, D. (1989). Coexisting organizational forms in Hungary's emerging mixed economy. In V. Nee and D. Stark, eds., *Remaking the Economic Institutions of Socialism: China and Eastern Europe*. Stanford, Calif.: Stanford University Press.
- Timár, J. (1994). A foglalkoztatás és a munkanélküliség sajátosságai a poszt szocialista országokban (Particularities of employment and unemployment in postsocialist countries). *Közgazdasági Szemle* 41: 633–661.
- Weiss, A. (1990). *Efficiency Wages: Models of Unemployment, Layoffs and Wage Dispersion*. Princeton, N.J.: Princeton University Press.
- Williamson, O. E., M. L. Wachter, and J. E. Harris (1975). Understanding the employment relation: the analysis of idiosyncratic exchange. *Bell Journal of Economics* 6: 250–278.

17 Transformation before the Transition: Employment and Wage Setting in Hungarian Firms, 1986–1989

János Köllő

Transitology, primarily engaged with the post-1989 history of Central and East European societies, pays little attention to the decay of the socialist system preceding its sudden political collapse. A “big-bang approach” to transition is certainly appropriate in the case of some countries, such as Romania or Czechoslovakia, where the political regime and the economy remained intact until late 1989, but not in Poland, Hungary, or the Soviet Union where the system underwent remarkable changes between 1985 and 1990.

This chapter analyzes an area of evolutive change in the last years of state socialism using Hungarian data. In particular, it addresses the question why employment began to shrink several years before the collapse of the socialist political system. Why did Hungarian state enterprises cut employment by two-digit percentages in the late 1980s, in a period of modest economic growth (1.4 percent per annum in 1985–1989) and see no dramatic changes in prices, trade relations, or other key economic indicators?

The magnitude of change is illustrated by table 17.1. Total employment (the sum of dependent-status employment and self-employment) fell slightly, by only 2.9 percent in 1985–1990, but the number of employees decreased by more than 8 percent. The decline fell entirely on the enterprise sector, where it exceeded 12 percent and implied a loss of half a million jobs. This figure still underestimates the magnitude of net job destruction in medium-sized and large firms where employment was predictably lower by as much as 17 percent, in 1989 as compared to 1986.¹

One of the reasons why job destruction could remain almost unnoticed was the outstanding role of self-employment (as opposed to open unemployment) in absorbing redundant labor. Many workers left the enterprise sector to start businesses voluntarily but, as Gábor R. (1994), Czakó (1996) and others have emphasized, self-employment was also a resort for workers losing their jobs or at risk of involuntary job loss.

A fall in labor demand (unrelated to external market pressures) was also observed on the micro level in both Hungary and Poland.² In their analysis of Polish state enterprises over the period 1983–1988, Lehmann and Schaffer (1995) found that employment was shrinking in the majority of firms, and they observed roughly the same rate of job destruction in firms paying below and above the estimated marginal product of labour. Kőrösi (1997) found that employment fell by 21 percent in 1986–1989 in Hungarian exporting firms. His repeated cross-section estimates of a

Table 17.1

Employment in the pretransition period in Hungary (in thousands)

	1983	1985	1987	1989	1990
<i>Active population</i>					
Unemployed	—	—	6	14	24
Self-employed ¹	243	n.a.	330	404	501
Employees ²	4,727	(4,641)	4,555	4,418	4,267
<i>Employees</i>					
Budget sector ³	764	784	848	853 ³	876 ³
Enterprises ⁴	3,963	(3,857)	3,707	3,565	3,391

Source: KSH (1991), pp. 19, 38–39, 74.

Notes

1. Includes assisting household members and full-time members in business partnerships without a legal personality.
2. Includes employees on military service and nonretired employees older than the retirement age. Excludes women on child care leave and employed pensioners.
3. Health, education, culture, and public administration. Institutions employing more than fifty workers in 1989 and 1990.
4. Calculated from the table by deducting budget sector employment from total employment. Figures in parentheses are interpolated estimates.

labor demand equation suggested low output elasticities, gradually rising from one-third to three-fourths of the levels observed during the transition. Estrin and Svejnar (1996) also observed net job destruction and loose linkage between output and employment in their Polish and Hungarian firm samples.

These observations motivate me to look at employment decline in detail using enterprise-level information from the period 1986–1989. I start in the first section by discussing employment and wage-setting under state socialist conditions, in a regime where (1) firms were constrained in wage setting; (2) enterprises had an incentive to raise employment beyond the efficient level; and (3) the actual wage-employment combination was achieved in a bargaining process and (therefore) was not optimal for workers.

In the second section I turn to the immediate pretransition period when the enlargement of enterprise autonomy, the hardening of the budget constraint, and the liberalization of firing practices created incentives for downsizing. I also argue that, given the properties of the state socialist heritage (points 1 and 3 above), downsizing could be Pareto-improving. Though the conceivable scenarios of downsizing appear to be different in terms of wage and profit change, they are similar in that they unanimously imply a fall in the level of employment.

I follow the discussion of causes and possible implications with data from 1986–1989, from which I draw conclusions in the final section.

1 The Socialist Firm

Wage Setting

Under central planning, the wage bill was determined by the planners. The wage plan was the single most stringent limit set for enterprises, as argued in detail in chapter 16 of Kornai (1980).

The case of Hungary is more complex because the country abolished central planning in 1968 and has controlled wages by means of taxes since then. Firms were free to set wages within broad limits suggested (rather than prescribed) for occupational categories. Taxes were levied on the wage *increment* with rates depending on the relation of wage change to a reference indicator. The details of the tax system changed frequently, but in the typical setting the tax was a function of the ratio of wage bill growth to value added growth. The tax was punitive with extremely high marginal rates.

It is important to note that the ruling tax regime punished the growth of the wage bill but (in principle) opened an unlimited range for the differentiation of average wages. Enterprises could raise the average wage tax-free by reducing labor hoarding. The tax regime is insufficient for explaining why interenterprise wage differentials remained modest. At least three equally important reasons might be mentioned.

First, though the tax rules explicitly encouraged the trading off of jobs for wages, other rules and incentives strongly discouraged it. Severe political constraints were imposed on “firing without a cause”: a firm dismissing workers for wage gains would have risked serious political sanctions.³

Second, employment cuts resulting in lower output were at odds with other incentives like the rewards attached to growth, market share, capacity utilization, or responsibility for supplies and CMEA-exports.

Third, wage cuts potentially resulting from productivity losses were usually averted by wage subsidies, exemptions from taxation, or value-added growth rates themselves were manipulated by transfers or via prices. (Kornai and Matits 1984). The distribution of subsidies followed relatively simple rules. Bargaining power increased with firm size because the number of established communication channels between the firm and the Party—hence the ability to bargain at all—was a function of size. Keeping firm size in terms of employment constant the likelihood of subsidies increased with capital assets. Cukor and Kertesi (1987) provide empirical evidence on that and argue that the cost of refusing subsidies was higher, other things equal, if low wages reduced the utilization of large assets of fixed capital.

Finally, subsidies were more likely when “equalizing wage differentials” had to be adjusted.⁴

Prohibitive taxes, political constraints, and manager’s incentives on the one hand, and subsidies on the other hand, thus operated as an equalizing mechanism: the risk that some firms set wages high above the industry average, or lag behind substantially, was actually rather small. The assumption of a wage level (or a wage range) quoted by the state seems to be a relevant assumption even in the special case of post-1968 Hungary, where firms had limited freedom in wage-setting.

Employment

To illustrate the difference between socialist and capitalist firms we draw figure 17.1. The curves emanating from the bottom left corner of the wage-employment plane are isoprofit curves: the loci of wage-employment combinations resulting in the same level of profit.⁵

An isoprofit contour is increasing and concave left of the turning point—where marginal product equals the wage—and falls afterward. The curvature of the isoprofit contour beyond the turning point is shaped by decreasing marginal revenue (implying successively increasing wage cuts as we move to higher employment levels) and a size effect (making higher profits available at the cost of successively decreasing wage cuts). As suggested in appendix 17.1, an isoprofit curve is likely concave in the relevant range where marginal product is well above zero.

Under state socialism, firms employed more workers than a capitalist enterprise facing similar wages and prices. The profit-maximizing firm (subject to w^* quoted

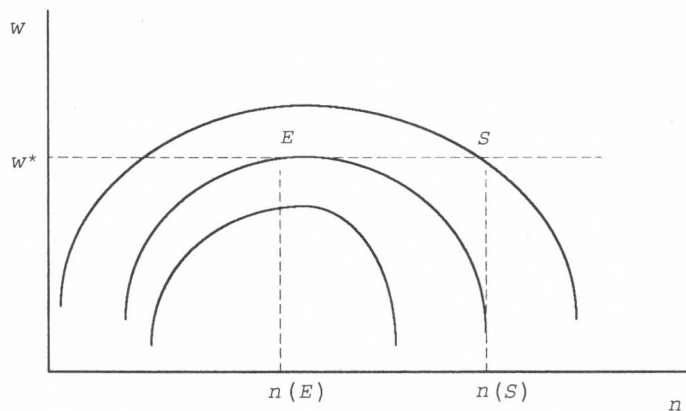


Figure 17.1
Employment and wage

by a union or government) is likely to choose $E(w^*, n_E)$ where the wage equals the marginal product.⁶ The choice of the socialist firm is $S(w^*, n_S)$, east of the efficient point.

E is an equilibrium point in the sense that, other things unchanged, the firm has no motivation to leave it. S is different in that state socialist firms were always ready to expand employment beyond its current level. In the short run firms were prevented from doing so by their budget constraint, but management could bargain for higher prices or fiscal transfers, making higher employment compatible with enlarged postsubsidy revenues. In this sense the assumption of a binding level of profit (setting a maximum for n_S at a given wage) would miss the important point about firm behavior. Kornai's notion of an "unconstrained" demand for labor (1980, chapters 11 and 16) refers exactly to that: irrespective of the level of the wage, in lack of a binding level of profit, n_S has a tendency to move to the right.

Cooperation and Conflict between Workers and Managers

When the budget constraint is soft, workers and managers are interested in cooperating against the state, for they can mutually benefit from subsidies. But it does not follow that they agree on how to distribute the subsidies among themselves.

The optimal choice of workers lies in the tangency of an isoprofit curve, regarded as binding momentarily by the management, and the highest worker indifference curve available at that profit level. For finding the place of this point we need reasonable assumptions about the representative worker's indifference curve (his or her evaluation of wages and employment). Because employment was full under socialism, higher employment in the firm had no extra value for the workers. Lower employment, by contrast, implied a loss of utility for them because of the cost of job search, or the loss of seniority-related wage returns, among others. We may assume therefore that the representative worker's indifference curve was flat right of the actual level of employment, but upward-sloping left of that point.

Suppose that the point S chosen by the firm accidentally provided the highest utility available for workers, at a given profit, as in figure 17.2. Workers and managers could always gain from state subsidies, so they kept on bargaining for state support. We may think of subsidization as a permission to move to a lower pre-tax profit rate, somewhere to the SAC area. Workers obviously preferred A to other points and pressed the management to fight for *wage* subsidies.

The managers of state socialist enterprises benefited from higher wages for a number of reasons (peace within the firm, lower labor turnover) but they were also rewarded for higher output, as well as for higher profit. The success criteria actually applied by government and Party officials under socialism suggested that a "proper"

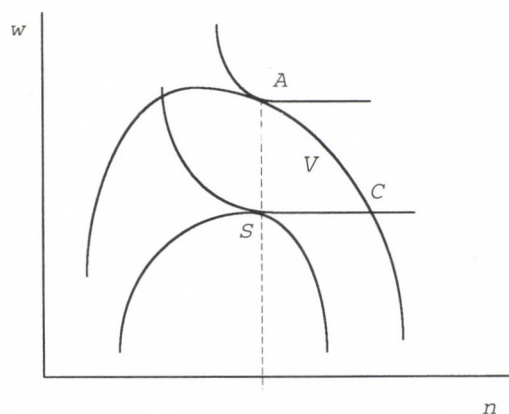


Figure 17.2
Optimum region

mixture of a number of indicators—rather than a single criterion—was what mattered in evaluating managers (Laki 1985). If we assume, accordingly, that managers strived for a combination of relevant success criteria—wage, employment, and profit in our presentation—the point they wanted to achieve was somewhere in the SAC triangle (V). If they had at least some bargaining power, the outcome was different from A —the firm moved away from worker's optimum.⁷

2 The Pretransition Reforms and Their Implications

Unlike in Czechoslovakia, Romania, or the German Democratic Republic, the transition process in Hungary was preceded by a period of evolutive change. The following paragraphs give a brief list of market-oriented reforms relevant for the labor market by grouping them into three clusters:

The delegation of managing rights to firms. The transferring of rights to manage came in both formal and informal ways. The establishment of elected enterprise councils, delegate's assemblies, and general assemblies (instead of subordination to branch ministries) created a legal framework for these changes in 1985. By the end of the period most Hungarian firms could autonomously decide on corporate structure, factor inputs, and the output mix.

Hardening of the budget constraint. A series of market-oriented reforms contributed to the monetization of the economy and the gradual hardening of the

budget constraint. These reforms included cuts of fiscal subsidies, wide-ranging price liberalization, tax reforms (Western-type value-added and personal-income taxation), the establishment of commercial banks, and the foundation of the stock exchange. Though these early reforms did not mean the end of government intervention, fiscal subsidization, and the bailing-out of bankrupt enterprises (huge loss-makers typically remained under direct government control even after 1985), they certainly signaled a departure from a comprehensive system of paternalistic income redistribution. The nature of change is well shown by time series presented in Kornai (1995): the proportion of firms mentioning demand-side factors among the causes of capacity underutilization considerably increased in this period, while the number of complaints on input “shortages” substantially decreased.

Freedom to trade off employment for wages. The pretransition reforms made the employment-wage trade-off a feasible option by removing the political barriers to dismissals and changing the management’s incentives. The government introduced the first elements of an unemployment benefit system, in 1987, signaling the end of political constraints on mass layoffs, demolishing the taboo of full employment, and allowing managers to reduce employment without risking their own jobs. It should be added that wage-bill taxation was maintained until 1991 (in a more sophisticated form in some years). Increasing wages without reducing employment remained a costly option at the eve of transition.

We believe that the reforms changed the motivation of the players and the rules of the game in many ways. The ceasing of stringent state control on one hand, and the lack of outside private owners on the other hand, provided the managers and core workers of state enterprises with a unique degree of autonomy. Equally important, they had stronger motivation to find an optimum within the limits of sales revenues. They also had new options, including downsizing. The targets chosen by firms may have been different—depending on who made the decisions—but the conceivable scenarios were common in that they all implied job destruction.

Some of the possible scenarios are based on worker control. One can assume that workers do not seek higher profit but they do not want to drive the firm out of the market by pushing profit deep below zero. What they likely try to do is maximising utility at close to zero profit. If at the outset the firm’s profit was low (and we may assume it with state socialist firms) they look for points on, or left of, the current isoprofit curve. This is shown in figure 17.3. Apart from the special case when S is a tangency optimum the “union” indifference curve crosses the isoprofit curve at S and C .⁸ The points between the indifference curve and the profit curve are Pareto-superior to S . If workers have their own way they choose B .

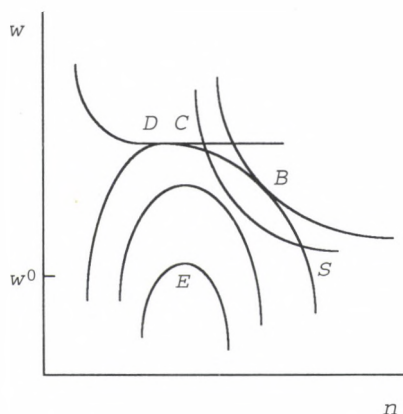


Figure 17.3
Different optima

In a variant of this scenario, decisions are made by, and in the interest of, a small group of core workers. With respect to profit such a small group is likely to behave as the community of all workers, but its utility curve is kinked at the level of group membership, as argued in Carruth and Oswald (1987). If decisions are made by such a group the likely outcome is *D*, the highest possible wage consistent with the prevailing level of profit. The position of *D* with respect to *C* is uncertain, but it is north-west of *S* as well as *B*. The scope for Pareto-improvement is even larger in this case.⁹

In other scenarios control is taken over by managers. A manager in the period under examination had several reasons to seek higher profit: high profit could raise his reputation on the emerging managerial labor market; it could help him in the newly born commercial banking system; or he wanted to become owner of the firm in the coming future. A profit-maximizing manager is generally interested in reaching point *E*, a wage at the reservation level w^0 and an employment level at which marginal product is equal w^0 . Whether he can reach this point is a matter of bargaining power, but he certainly wants to move west or southwest of *S*.

In principle, workers or managers uninterested in profit may have tried to maintain employment and raise wages at the expense of profit, but we do not regard it as a highly relevant option given the proximity of n_s to a level exhausting the firm's intakes at any particular point in time, under socialism. In principle this option would have implied a move to the north of *S*.

A common corollary of the more relevant scenarios of pure worker control or profit-maximizing managerial control—or a mixture of the two—is that the reforms bring about labor shedding without any external market impetus. The degree to

which employment cuts are accompanied by wage increase depends on the nature of decision-making. Similarly, expectations concerning profits are uncertain, but the guess is that even under strong worker influence profits do not change considerably.

The aim of the empirical analysis that follows is to identify typical paths on the wage-employment plane between 1986 and 1989, that is, to assess how output, employment, wages, and profits were changing, and how they related to each other. There is an important detail to be addressed before starting the analysis that is connected with the nature of labor shedding.

If workers are fired on a nonrandom basis the observable outcome of the adjustment process is affected in two ways. On the one hand, productivity-based selection weakens the link between employment change and output change because the dismissal of the least productive workers has a minor impact on output. On the other hand, by shedding the least productive workers the firm's staff changes in favor of the more qualified and more experienced workers. The average wage increases for that reason, but we cannot take it for granted that the remaining workers are better paid. The wage gain from labor shedding can only be measured by the evolution of residual wages, that is, wages controlled for demographic and human-capital variables.

Keeping these warnings in mind, we will analyze the path of firms in three steps in the following sections. First we look at the typical path in terms of output, employment, and wages by estimating a model of simultaneous employment and wage setting. Second, residual wages and their relation to labor shedding will be analyzed in order to check whether the observed link between average wages and labor shedding is real or spurious. Finally, we relate profit change to employment and wage change in order to capture the impact of downsizing on employers.

3 Data

The data come from an enterprise panel covering 2,666 firms observed in both 1986 and 1989 in the framework of the wage survey of the National Labor Centre of Hungary. The firms are randomly selected. The sample covers 51.7 percent of large and medium-sized firms (employing more than 20 workers in 1986). The firms in the sample employed 68.1 percent of all workers in the respective size categories. In addition to enterprise-level information the records comprise individual data of a 10 percent random sample of workers in 1986 as well as in 1989 (but the persons interviewed are not identifiable across waves). The individual data will be used to draw residual wages. The samples are described in detail in Kertesi and Köllő (1995).

Unfortunately, enterprises involved in spin-offs, dismantling, mergers, or any other change affecting their legal personality (and hence their identification code) are automatically excluded from a longitudinal sample. This restriction is severe in an economy where thousands of firms dismantle or merge with others in search of more efficient organizational forms, in order to seek rents, avoid taxes or hide from the eyes of creditors. From the 4,032 firms observed in 1986 we could follow 3,250 over the period 1986–1989, and a further 584 had to be excluded because their data were strongly biased by changes of corporate structure.¹⁰

4 Results

Output, Employment, and Wages

As shown in table 17.2 the output of the median firm did not change between 1986 and 1989, but it cut employment by more than 20 percent. The employment cuts were virtually equal at firms with falling and rising output. At first glance employment seems uncorrelated with sales performance, but wages appear to differ in expanding and contracting firms.

For obtaining more informative results we estimate a model of simultaneous employment and wage setting:

$$\begin{aligned}\Delta \log n &= b_{11} \times \Delta \log q + b_{12} \times \Delta \log(w/p) + b_{13} \times X + e_1 \\ \Delta \log w &= b_{21} \times \Delta \log(q/n) + b_{22} \times \log(p) + b_{23} \times Z + e_2\end{aligned}\tag{17.1}$$

where p is the producer price index, q is output (sales revenues net of material cost, deflated with p), and X and Z are vectors of control variables. Q , p , X , and Z are exogenous. Delta refers to change between 1986 and 1989.

Estimations of (17.1) and related calculations are available for a similar enterprise panel from 1989 to 1992 in Köllő (1996); therefore we have the possibility to

Table 17.2

The median of basic variables in 1989 (1986 = 100)

Median	All firms ($N = 2,666$)	Firms with falling output ($N = 1,334$)	Firms with rising output ($N = 1,332$)
Output	99.0	80.6	125.7
Employment	78.4	76.7	80.1
Real wage*	131.2	124.1	138.2

* Deflated with the branch-level producer price index.

compare the results from the late 1980s with ones coming from the times of transition.

The elasticity of employment to output estimated from (17.1) is 0.19. This figure compared to a similar estimate from 1989 to 1992 (0.35) seems rather low, supportive of the conjecture that the sudden change of employer behavior was a result of institutional change rather than of market pressures.

The rate of job destruction was highest in the medium-size category: large firms (employing more than a thousand workers) had a 12.5 percent higher level of employment than had medium-sized firms, in 1989 relative to 1986, other things being equal. This outcome also seems consistent with the framework developed earlier. Although the state held important positions in the enterprise councils of large firms (and kept some large firms under direct ministry control), the transfer of managing rights to the enterprise was fully completed in the case of medium-sized firms. Therefore more severe employment cuts are expected here than at large firms (or small ones where the management had a higher degree of freedom even before).

The employment equations of table 17.3 suggest a strong trade-off between wages and employment: a 1 percent higher rate of change of the average wage appears to reduce employment by more than 0.5 percent. (The causation can be reversed as the equation was normalized for employment by choice). The parameter (−0.56) is three times as high as in 1989–1992 (−0.17). Though the interpretation of this coefficient can be a matter of debate, in the particular case of the pretransition firm assuming direct linkage between the reduction of employment and fast-growing average wages is supported by at least two arguments. Firms with an accumulated

Table 17.3
Estimation of employment and wage adjustment (2SLS)

Employment equation			Wage equation		
	Coeff.	(<i>t</i>)		Coeff.	(<i>t</i>)
Output	0.190	18.7	Productivity	0.060	6.2
Real wage	−0.564	−11.9	Price index	0.064	1.7
Exporter (Rbl)	−0.006	−0.5	Ratio of manual workers (log)	—	
				0.068	−7.0
Exporter (\$)	0.082	7.2	100% nonmanual	0.086	0.6
Small firm (<100)	0.087	2.8			
Large firm (>1000)	0.125	12.3			
Constant	−0.197	−15.4	Constant	0.790	17.2
aR^2, N	0.365	2,666		0.092	2,666

stock of low-productivity jobs could downsize employment without a major effect on output (and hence with a positive effect on productivity and wages). In addition they could choose to cut both output and employment with the very aim to achieve higher wages, as we have discussed.

The wage equation in table 17.3 suggests a relatively weak connection—an elasticity of 0.06—between wages and productivity (or the branch-specific price index). It might be added that by regressing the producer real wage on productivity in a univariate model we get a far higher elasticity of 0.2. This is consistent with the finding of a highly significant trade-off between employment and wages because the parameter of the univariate estimation combines the effect of a “sharing rule” (higher productivity leads to higher wage) with a feedback effect (that leads from higher wages to lower employment at given sales revenues, hence to higher productivity and back to wages).

The first results suggest that the typical Hungarian firm substantially reduced employment between 1986 and 1989; employment cuts were loosely related to sales performance; and more labor shedding was associated with higher wage increase. However, the wage gains from labor shedding indicated by equation (17.1) may be spurious, calling for the study of residual wages.

Residual Wages

Residual wages are obtained from Mincer-type earnings functions estimated for 1986 and 1989 using the individual data of workers employed in the firms under examination (the equations are presented in appendix 2). We calculate the error terms of the equations, take enterprise means, and study how they relate to other indicators of the firm. In case the finding is that more labor shedding between 1986 and 1989 was associated with a rise of the residual wage, at given sales dynamics, we have more reason to believe that the jobs of fired workers were traded off for higher wages paid to the remaining insiders.¹¹

The calculations are restricted to firms employing at least 500 workers in both 1986 and 1989 and thus providing an adequate sample of fifty or more workers observed in the wage survey.

Table 17.4 presents two wage equations. In the first column the average wage in 1989 (w_{89}) is regressed on a lagged term (w_{86}), change of output, and change of employment. In the second column the average wage figures are replaced for the residual wage figures (r_{89} , r_{86}). All variables are in logs.

In the equations the lagged terms have been used as indicators for the change in the effect of ignored variables—the closer their coefficients to 1, the less these relationships changed. In the wage equation the parameter is close to 1, suggesting that

Table 17.4

The effect of output change and employment change on average wages and average residual wages (OLS estimates)

Dependent	w89		r89	
	Coeff.	(t)	Coeff.	(t)
Lagged term (w86, r86)	0.998	31.1	0.634	18.1
Change of output	0.200	14.2	0.071	5.1
Change of employment	-0.494	32.9	-0.038	3.8
Constant	0.162	0.6	-0.025	4.1

Residual wage equation:

$aR^2 = 0.235$, $F = 120.2$ (0.0000), Nobs = 1,164.

Wage equation:

$aR^2 = 0.5993$, $F = 643.9$ (0.0000), Nobs = 1,164.

the ignored wage effects were rather stable. In the residual wage equation the coefficient is far lower than unity, indicating that some unobserved reasons leading to positive or negative deviations from "expected" earnings may have changed during the period.

The effect of both output and employment is markedly stronger in the average wage equation. The coefficients of the employment terms bear special importance from our point of view: they appear to be negative and significant in both equations, but the parameter in the wage equation is thirteen times larger.

It seems that more labor shedding at given output led to higher wages indeed, even after allowing for changes in the staff's composition, but this effect was very weak. Evaluating the outcome at the mean change of sales and two relevant points of employment change (mean $\pm$ standard deviation) gives the following results. Cutting employment by 2 percent had no effect on the residual wage; cutting jobs by 47 percent induced a 2.4 percent residual wage increase. Undoubtedly, this is a very small gain from shedding half of the firm's staff, raising doubts whether labor shedding was indeed beneficial for the incumbent workers.

We may reverse the question, however, and ask how the actual wage gains emerged. Table 17.5 gives mean figures of output and employment in the quartiles of residual wage change. The data suggest that in the firms where residual wages grew fast employment decreased substantially while output practically remained at its 1986 level.

The conclusions encouraged by the evidence in this section might be summarized in two statements. First, it seems that in many cases the increase of the average wage (associated with massive labor shedding) covered no more than the rising cost of

Table 17.5

Output and employment in four quartiles of residual wage change

	Quartiles of residual wage change		
	Lowest	2nd and 3rd	Highest
<i>Mean 1989/Mean 1986</i>			
Output	0.87	0.91	0.98
Employment	0.79	0.83	0.81
<i>Mean errors in the quartiles:</i>			
Mean error 1986	0.038	-0.003	-0.031
Mean error 1989	-0.121	-0.009	0.130

Note: Firms employing more than 500 workers (Nobs = 1,164)

labor due to its improving quality. Second, we nevertheless observe that residual wage gains were generated in firms where a relatively large proportion of the employees were dismissed, at constant output.

It should be added that by restricting the analysis to a longitudinal sample of large firms we certainly underestimate the importance of downsizing—as a source of wage gains—in the economy. The dissolution of large state firms often took the form of spin-offs, cases when smaller groups of managers and core workers shifted assets from the master enterprise to new firms and set higher-than-average wages (financed from higher *labor* productivity).

Profits

For a brief overview of how profits were affected by downsizing, we grouped the enterprises into quintiles by change of employment and change of the real product wage and looked at profit change in these groups. The change of the profit rate in a cell is affected by the relation of productivity change and wage change. We are particularly interested in the question whether the productivity gains associated with large-scale labor shedding could cover the rapid growth of wages at downsizing firms.

To reduce the disturbances caused by industry and price effects, the deviation of the observed profit change from its predicted value was used as a success indicator of employers. The prediction was calculated from a model regressing the percentage change of the profit rate on the producer price index, output change, and dummies for exporters and one-digit industries (appendix 3). The residual profit change obtained in this way is expected to respond to firm-level decisions, concerning employment and wages, more than does the raw variable. Table 17.6 shows

Table 17.6

Percentage change of the residual profit/sales ratio: Weighted group means by change of employment and change of the real product wage (number of firms in parentheses)

Quintiles by change of wage	Quintiles by change of employment					Wage 1989/86
	1	2	3	4	5	
1	-1.02 (246)	-2.97 (96)	-5.18 (69)	-3.18 (57)	-1.67 (66)	1.70
2	-0.41 (101)	-1.68 (117)	-2.78 (105)	-3.71 (109)	-1.50 (101)	1.40
3	1.90 (72)	-0.74 (109)	-1.36 (132)	-2.16 (99)	-1.69 (121)	1.30
4	2.70 (59)	2.01 (103)	-0.94 (118)	-2.24 (136)	-1.14 (117)	1.19
5	4.33 (56)	1.81 (108)	0.20 (109)	-1.61 (132)	0.14 (128)	1.02
Employment 1989/86	0.53	0.71	0.80	0.88	1.00	<i>N</i> = 2,666

the mean change of the residual profit and the number of firms in each cell. (While we calculated the mean, firms were weighted with employment in the base year.)

The cells where profits grew or decreased by less than 1 percent are separated from the others by a winding line in table 17.7. The impression is that large-scale downsizing was associated with relatively favorable profit records (cells left of the line). It seems that the major source of profit squeeze was the *lack* of downsizing, especially when it was accompanied by above-the-average wage growth.

The firms in the upper left cell are particularly interesting from the perspective of the discussion. These enterprises cut the number of jobs by almost 50 percent and increased the average wage by 70 percent between 1986 and 1989. The number of firms in the group is large, 246 as opposed to an expected frequency of 107. Despite the fast wage growth characteristic of the group, profits decreased by only 1.02 percentage points, which appears to be a minor fall compared to the average.

The test applied here is undoubtedly elementary but, perhaps, allows the conclusion that contracting firms had relatively favorable profit records. Although it is true that large-scale labor shedding tended to induce a rise in the average cost of labor, and sometimes was associated with growing residual wages, it seems that the rising average product of labor could generally accommodate the wage increment.

5 Conclusions

The immediate pretransition years do not give the impression of a motionless world waiting for the "big bang." Actually, the firm-level and individual data suggest substantial changes of employment and wages. The transformation of the labor market

definitely started several years before the political collapse of state socialism in Hungary.

I have attributed massive labor shedding to institutional reforms rather than market-related shocks, arguing that the elimination of labor hoarding on the one hand, and a move to the west from the starting wage-employment combination on the other, was in the interest of workers, managers, or both parties.

The nature of downsizing (in terms of profit change and wage change) is a question to be addressed in empirical research. A short panel of Hungarian enterprises from 1989 to 1996 suggested massive labor shedding that was rather loosely correlated with output. Average wages grew faster in firms shedding a larger proportion of their staff but we found that a large part of the resulting productivity gain was actually used to finance the growing average cost of a smaller (and better) labor force. Worker's gains from labor shedding were relatively modest (albeit positive). Huge residual wage gains were nevertheless achieved in firms where severe employment cuts occurred at constant output. We observed minor impact on profits: it seemed that the returns from downsizing could generally finance the growth of wages.

These observations seem to support the conclusion that the elimination of labor hoarding had been started already before 1990 and the bias toward high employment and low wages, characteristic of the state socialist regime, was to some extent corrected in the pretransition period. Workers remaining in employment gained rather than lost from labor shedding, whereas firms did not lose in terms of profitability (or they lost for reasons other than employment and wage adjustment). The loss of workers pushed out of employment belongs to the full picture and waits for deeper research in the future.

Appendix 17.1

The second derivative of the isoprofit curve yields a condition of concavity

$$R'' < 2(R' - R/n)/n$$

that is difficult to evaluate. We argue that at levels of n where the marginal product is significantly higher than zero, the isoprofit curve is concave. We show it for the case of the zero-profit curve.

Let n be an employment level above the efficient point. Along the zero-profit curve the wage is set so that $w = R(n)/n$. Concavity implies that:

$$w(n+1) - w(n) < w(n) - w(n-1) \quad (17A.1)$$

that is:

$$R(n+1)/(n+1) + R(n-1)/(n-1) < 2R(n)/n \quad (17A.2)$$

We can rewrite (17A.2) as:

$$[n/(n+1)]R(n+1) + [n/(n-1)]R(n-1) < 2R(n). \quad (17A.3)$$

The concavity of R implies that:

$$[n/(n+1)]R(n+1) + [1 - n/(n+1)]R(n-1) < R(n). \quad (17A.4)$$

If it is true that:

$$\{n/(n-1) - [1 - n/(n+1)]\}R(n-1) < R(n) \quad (17A.5)$$

then (17A.3) will hold. By rewriting (17A.5) we get (17A.6) as a condition of concavity:

$$(n^2 + 1)/(n^2 - 1) < R(n)/R(n-1). \quad (17A.6)$$

Consider firms with 10, 100, ..., 1000 employees. Adding one worker to the staff implies a percentage change of 10 percent, 1 percent, and 0.1 percent respectively. The critical values $(n^2 + 1)/(n^2 - 1)$ are 1.02, 1.0002, ..., 1.000002. It follows that unless an 1 percent change of employment implies less than 0.2 percent, 0.002 percent, ..., 0.0002 percent change in output, (17A.6) holds and we are at the concave region of the isoprofit curve.

Appendix 17.2

Earnings functions for 1986 and 1989 (sample: workers in the 2,666 firms in the panel; dependent: log monthly earnings 1986, 1989)

	1986		1989	
	Coeff.	<i>t</i> -value	Coeff.	<i>t</i> -value
<i>Gender = male</i>	0.298	138.3	0.311	121.2
<i>Age</i>	0.038	66.6	0.040	58.9
<i>Age squared (/00)</i>	-0.037	-52.4	-0.040	-46.1
<i>Education*</i>				
Less than completed primary	-0.084	-52.4	-0.059	-13.9
Vocational manual	0.083	33.4	0.079	27.1
Secondary	0.090	30.2	0.110	30.7
Higher	0.274	50.2	0.381	60.1

(continued)

	1986		1989	
<i>Job category*</i>				
Nonmanual	0.092	30.2	0.178	50.2
Manager	0.544	53.0	0.831	72.3
<i>Region dummies (16 regions)</i>	yes		yes	
<i>Industry dummies (36 branches)</i>	yes		yes	
<i>Constant</i>	7.684		8.022	
aR ² , Nobs:	0.4377	88,089	0.4392	87,747

* Reference categories are completed primary school; manual.

Appendix 17.3

Estimation of profit change in 1986–89 (OLS)

Dependent: Percentage change of profit/sales	Coeff.	(t)
Producer price index	9.909	3.3
Output 1989/86	1.199	8.8
Ruble exporter (1986)	-1.271	2.9
Dollar exporter (1986)	-0.623	1.5
One-digit industry dummies	yes	27.2
Constant	-18.959	5.1

aR²: 0.1114, F = 34.4 (0.0000), Root MSE: 7.66. Joint significance test for industry dummies.

Notes

I started research on employment and wage adjustment in 1993–1994 when I was a visiting fellow in János Kornai's research group at Collegium Budapest—Institute for Advanced Study. I am grateful to him for his encouragement and criticism. I also thank György Lázár for support with data and Olivier Blanchard, Simon Commander, and Gábor Kőrösi for comments on earlier versions. Errors are mine.

1. This figure comes from the firm sample introduced later in the chapter, covers enterprises employing more than 20 workers, and compares yearly average figures.
2. Though a part of the outflow from the enterprise sector was clearly supply driven, it still remains a question why enterprises, for the first time in socialist history, did not try to fill their vacancies opened by voluntary quits.
3. The only case when hundreds of workers were dismissed collectively under János Kádár's reign (Rába, city of Győr, 805 workers) was put on the agenda of a Politburo session on May 29, 1979.
4. As, for instance, the price of hard manual work increased—reflecting a general improvement of educational levels, the ceasing of migration from villages to industrial centers, and a consequent fall in the supply of unskilled labor—the wage levels generated by the “value added-wage growth rule” would have been insufficient to maintain staff levels at firms with predominantly unqualified jobs. Subsidies were often explicitly targeted to mitigate “labor shortages” of this kind.

5. Let $R(n)$ be a revenue function with the usual properties $R'(n) > 0$ and $R''(n) < 0$. If (w^*, n^*) is a wage-employment combination for which $R'(n^*) = w^*$ and $R(n^*) - w^*n^* = P^*$ then $R(n) - wn = P^*$ defines an isoprofit curve in the wage-employment plane the slope of which is given by $[R'(n) - w]/n$ and thus its turning point is at (w^*, n^*) . The isoprofit curves serve as indifference contours with lower curves representing higher profit. The framework used here follows the introductory sections of McDonald and Solow (1982).
6. The optimum is different if both employment and wages are set in a bargain, as shown by McDonald and Solow.
7. Note that after a shift to higher employment levels (say from S to V) the union indifference map should be rescaled. This is so because in this presentation worker's utility is a function of the change, not the level, of employment. Their preferences are transitive on the $(w, \Delta n)$ plane but not on the (w, n) plane.
8. Whether the worker's indifference curve is kinked or well-behaved bears no importance in this case.
9. The boundaries of the area Pareto-superior to S are the actual isoprofit curve, the demand curve, and a horizontal line at height w_s .
10. The methods of data cleaning are presented in Köllő (1996).
11. Deviations from predicted wages may reflect a series of unobserved factors like work intensity, polluted workplace, or the risk of corruption, but they can also hint at rents paid above (or below) the market level of wages. The equalizing differentials probably do not change considerably over a relatively short period of time, but the rents are certainly more variable and are expected to respond to changes of the average product.

References

- Basu, S., S. Estrin, and J. Svejnar (1994). Employment and Wage Behavior of Enterprises in Transition. Paper presented at the workshop on Enterprise Adjustment in Eastern Europe. Washington, D.C.: The World Bank.
- Carruth, A., and A. Oswald (1987). On union preferences and labor market models: insiders and outsiders. *The Economic Journal* (June): 431–445.
- Cukor, E., and G. Kertesi (1987). Inter-enterprise wage differentials in Hungary: causes and implications. *Acta Oeconomica* 38(1–2): 79–115.
- Czakó, Á. (1996). Unemployment in Hungary—The First Five Years. Mimeo, Social Cost of Transition (SOCO) Project, IWM, Vienna.
- Estrin, S., and J. Svejnar (1996). Employment and Wage Determination in the Early Years of Transition. Paper presented at the workshop on Unemployment, Restructuring, and the Labor Market in Eastern Europe and Russia. Washington, D.C.: The World Bank.
- Gábor, R. I. (1994). Small entrepreneurship in Hungary: ailing or prospering? *Acta Oeconomica* 46(3–4): 333–346.
- Kertesi, G., and J. Köllő (1995). Wages and Unemployment in Hungary 1986–94. Mimeo, ILO-Japan Project, Budapest.
- Köllő, J. (1996). Employment and Wage-Setting in Three Stages of Hungary's Labor Market Transition. Paper presented at the workshop on Unemployment, Restructuring, and the Labor Market in Eastern Europe and Russia. Washington, D.C.: The World Bank.
- Kornai, J. (1980). *The Economics of Shortage*. Amsterdam: North Holland.
- Kornai, J. (1995). Eliminating the shortage economy: a general analysis and examination of the developments in Hungary. *Economics of Transition* 3(1): 13–37; 3(2): 149–168. Also in Hungarian, German, and Russian.
- Kornai, J., and Á. Matits (1984). Softness of the budget constraint—an analysis relying on data of firms. (With Agnes Matits.) *Acta Oeconomica* 32(3/4): 223–249. Also in Hungarian, Japanese, and Chinese.

Lehmann, H., and M. E. Schaffer (1995). Productivity, employment and labor demand in Polish industry in the 1980s: some preliminary results from enterprise-level data. *Economics of Planning* 28(1): 1–27.

KSH (1991). A nemzetgazdaság munkaerőhelyzete a 80-as években. Budapest: Központi Statisztikai Hivatal.

Kőrösi, G. (1997). Labor Demand during Transition in Hungary (Econometric Analysis of Hungarian Firms 1986–1995). Working Paper, Institute of Economics, Hungarian Academy of Sciences.

Laki, M. (1985). Economic policy and the practice of outstanding firms—report on a workshop held in Győr, 28–29 May 1985. (In Hungarian.) *Közgazdasági Szemle* (1985): 12.

McDonald, I., and R. Solow (1982). Wage bargaining and employment. *American Economic Review* 71: 896–908.

18 How Does the Transitional State Behave? Evidence from Retained State Ownership in Mongolian Privatization

Georges Korsun and Peter Murrell

1 Introduction

János Kornai's contributions to economics are rightly renowned for their deep conceptual advances, which have stimulated much work by other researchers, promoted extensive debate, and become part of the lexicon of the economics discipline. All scholars of comparative economic systems, for example, are versed in the notion of overcentralization (Kornai 1959), in the concept of a soft budget constraint and the concomitant economics of shortage (Kornai 1980), and in the idea of the affinity between ownership forms and coordination mechanisms (Kornai 1986).

A less noted aspect of Kornai's work is its methodological style, perhaps best exemplified by the arrangement of the titles of the following two works: *Overcentralization in Economic Administration: A Critical Analysis Based on Experience in Hungarian Light Industry* and *The Affinity Between Ownership Forms and Coordination Mechanisms: The Common Experience of Reform in Socialist Countries*. The conceptual insight embodied in the titles derives from deep study of those real-world microeconomic processes identified in the subtitles. This is the hallmark of Kornai's work: the attention to empirical detail that precedes the inductive reasoning, which in turn results in the theoretical insight.

If we are to use this empirically based, inductive approach for the burgeoning field of transition studies, then we are now in the phase where grand conceptual theorizing has to take a back seat to collection of information and to detailed empirical study of the ongoing processes in the many countries undergoing economic revolution. Kornai's observations on the evolution of financial discipline, the causes of recession during transition, and the role of the state are models in this respect (Kornai 1992, 1993, 1994).

The present chapter contributes to knowledge of transition processes by offering information on the behavior of the state in transition. In debates on the strategy and tactics of reform that took place in the beginning of transition, one of the prime areas of contention was how the state would behave in this new market environment. Sweeping visions of economic reform were based on strong assumptions about how the state would behave. But we now have enough years behind us, and enough empirical information to digest, that we are able to examine these assumptions.

Mongolia is a small country at the opposite end of the spectrum from Kornai's Hungary in terms of standard of living, degree of pretransition reform, and

geography. But it offers interesting evidence for the same reason as Hungary—there has been a distinctive style to its policies, with Mongolia being one of the most advanced reformers in the set of very poor transition countries.

In Mongolia, as in many countries, the state held residual shares in enterprises that went through the privatization process. The questions addressed in the following are what determined the size of these residual shares and what does this tell us about the way in which a reformist transitional government behaves. In answering this question, we follow Kornai's methodology of inductive reasoning, examining what the data show before proceeding to broader interpretations of the results.

In the next section, we briefly describe the Mongolian privatization program. The third section presents evidence on the size and the nature of residual state ownership. The fourth section contains a regression analysis, with residual state share as the dependent variable, and interprets the results in terms of the basic characteristics of the postsocialist environment. The fifth section provides further interpretation using theories of privatization that have been developed recently. Concluding comments appear in the last section.

2 Mongolian Large-Enterprise Privatization

For seventy years, Mongolia was a *de facto* republic of the Soviet Union, although it was formally independent. Therefore, its circumstances were highly comparable to those of the new nations spawned by the demise of the USSR. But the fact of *de jure* independence meant that Mongolia began its reforms before the ex-Soviet republics, proceeding very fast on some aspects, such as privatization. Here, we provide basic facts on Mongolian privatization to set the context of the results that follow.¹ A detailed overview appears in Korsun and Murrell (1995a).

As the 1990s began, Mongolia was still under the hegemony of the Soviet Union. Only a small amount of decentralizing reform had occurred during the perestroika era. In early 1990, a political revolution led to the country's first free elections. The old Mongolian People's Revolutionary Party (MPRP) won, but it formed a coalition government with the new parties. Significantly, the economics portfolio was given to a member of the National Progress Party (NPP), which had been formed by a group of young economists who were convinced that the country needed truly radical reforms to move ahead. The implementation of the central element of these reforms, privatization, kept remarkably close to a very ambitious timetable.² Here we focus on "large privatization," the main component of which was the privatization of large-scale enterprises in the industrial, trade, and service sectors.

The economists in the NPP designed a far-reaching and innovative privatization program, which was accepted swiftly by the government. In January 1991 the government established the Privatization Commission, whose staff had considerable decision-making power, and the Stock Exchange, which performed a key role in the privatization of large enterprises. The basic components of the original program made their way into the law passed in May 1991. The first privatization of a large enterprise took place in March 1992, to be followed by more than 300 more in the next sixteen months.³

On the supply side, large privatization was highly centralized. The Privatization Commission staff ordered enterprises to prepare plans to be reviewed by the commission's experts. There was little discretion in the preparation of plans and no discretion in the method of privatization—every large enterprise would be sold for vouchers. However, the number of shares placed on sale was often less than the total, because the state retained part ownership in a large number of enterprises. We focus on this element of the privatization process: In which enterprises did the state keep a residual ownership share?⁴

Decisions on the residual state share resulted from the interactions of a tangle of actors. The Privatization Commission, which had the power to make the detailed decisions, was led by a politician advocating the most extensive reforms possible and was staffed by like-minded individuals. The Privatization Commission was an arm of a government that was much more conservative, but that nevertheless seemed to pay only cursory interest to the details of privatization. Guidelines that defined the enterprises that were to remain under some degree of state ownership could be loosely interpreted by the Privatization Commission staff. Enterprise insiders who had no formal power whatever could nevertheless use obfuscation, delay, and the mobilization of work forces. Informal bargaining rather than open debate led to the decisions on state share. Extensive interviews with the major participants have not served to untangle the objectives and constraints that drove the elements of this bargaining process.⁵ Hence, we now turn to a statistical summary of the results of that process to ask whether there are any regularities in the data that might provide some enlightenment.

3 The Size of the Residual State Share

The Mongolian privatization program was as comprehensive as any in the ex-socialist countries. Outside the utility and mining sectors, virtually every enterprise was slated for privatization. But, as in many countries (Perotti 1995; Pistor and

Turkewitz 1994), privatization fell far short of 100 percent private ownership. The state retained stakes ranging from 15 percent to 80 percent in 41 percent of the first 303 enterprises to be privatized.⁶

Table 18.1 presents basic information on the dimensions of the residual state share in those large enterprises that were privatized within sixteen months of Mongolia's first large enterprise privatization.⁷ The official decision on residual state ownership in any enterprise was cast in terms of the percentage of shares that would be retained by the state instead of being either sold directly to employees or auctioned on the stock exchange. In the first row of the first column, table 18.1 presents the mean of those percentages for all privatized large Mongolian enterprises. On average, then, the Mongolian government retained an ownership share of more than 20 percent.

In an initial examination of variations in state share, the most interesting disaggregation is along a political fissure, capturing the political environment under which firms were privatized. The Government (or national) Privatization Commission administered the privatization of nearly half of the enterprises, with local privatization commissions administering the remainder. Outside the capital, Ulaanbaatar, and the two other large cities, Darhan and Erdenet, these local privatization commissions were much more likely to be under the sway of old-style politicians than were their national or urban counterparts. The last two rows of table 18.1

Table 18.1

The residual state share in Mongolian enterprises (data for 303 enterprises)

	(a)	(b)	(c)
	Mean percentage of enterprises shares	Percentage of the total net worth of privatized enterprises, book value	Percentage of total net worth of privatized enterprises, stock market valuations
All Mongolian enterprises	20.77	21.12	26.04
Urban $\cup$ national enterprises	16.73	18.37	24.48
Rural $\cap$ local enterprises	27.30	31.20	37.64

Urban $\cup$ national grouping: all enterprises located in the three biggest cities and all enterprises privatized by the Central Privatization Commission.

Rural $\cap$ local grouping: all enterprises located outside the three big cities that were privatized under local privatization commissions.

Sources: From data provided by the Mongolian Government Privatization Commission and the Mongolian Stock Exchange.

therefore disaggregate the data between those enterprises in rural areas privatized by local commissions (the “rural $\cap$ local” group) and those privatized by the more reformist national or urban commissions (the “urban $\cup$ national” group).

Not surprisingly, this partition shows that the state share is higher in the rural $\cap$ local group than in the urban $\cup$ national one, indeed more than 60 percent higher. At least at this beginning level of analysis, these results suggest an effect of the political orientation of officials on privatization decisions.

The data in the first column, the mean residual state share across enterprises, might not accurately reflect the aggregate economic importance of the state share. One requires means weighted by the size of the enterprises. The remaining two columns of table 18.1 provide the pertinent information. The second column gives the share of the aggregate capital stock of privatized enterprises retained by the state, when that capital stock is measured using book values. Book valuations were the ones that were available to officials when they were making their decisions on state shares. Using these book values to measure the capital stock retained by the state, one obtains essentially the same picture as that from data on mean enterprise shares, as is clear in comparing the first column to the second column.

The figures in the second column reflect backward-looking valuations, whereas the third column employs forward-looking valuations. Each privatized enterprise was sold for vouchers on the stock exchange in a quasi-auction procedure, which resulted in relative valuations being established on the basis of investors' demands. The data in the third column show the share of the aggregate capital stock of privatized enterprises retained by the state when that capital stock is measured on the basis of these investor valuations. This forward-looking measure presents a rather different picture, suggesting a rather larger residual presence of the state in the privatized sectors of the economy.

There are two very different causal mechanisms that could account for the large differences between the data based on book values and the data based on market valuations. First, it is clear that privatization officials had inside knowledge of which enterprises had the best prospects under the new system, for example, on which capital stocks would be less valued in a market economy than in a socialist economy. Given this information, these officials might have decided to retain a larger state share in the enterprises with better prospects, leading to a higher retained share when measured using investors' valuations than when measured using book values.

Alternatively, investors might have decided that enterprises with larger state shares represented a better prospect when these investors used their vouchers in the stock market auctions. If this were the case, then stock-market valuations would

be positively correlated with the size of the residual state share, resulting in the different pictures conveyed by the second and third columns of table 18.1.

Obviously the data in table 18.1 cannot discriminate between the two causal mechanisms described in the two previous paragraphs. However, additional information is presented in the next section, which suggests that the first mechanism—the state retaining better prospects—can be ruled out.

4 Interpreting Patterns in the Residual State Share

The following empirical exercise describes variations in the size of the residual state share across Mongolian enterprises. Received theory is so indirectly related to the data at hand that such a theory could be suggestive only in guiding the formulation of a regression equation. Thus, in obtaining the regression results presented in the following, inchoate ideas on the forces determining the residual state were used to suggest variables that were likely candidates for study. The final version of the estimated equation materialized from a process in which these ideas were modified as a consequence of the learning involved in examining the data.

Given the poor base of official data on which to build empirical work, the following statistical exercise relies on both official statistics and information collected from a survey of enterprises. In 1993, a survey questionnaire was administered to all those enterprises in Ulaanbaatar that had already passed through the large privatization program. The results appearing below reflect the circumstances of these enterprises: the sample covers the one-third of privatized enterprises that are in the capital city, rather than the population of more than three hundred privatized firms included in table 18.1.

The results appear in table 18.2. The dependent variable is the percentage post-privatization state ownership in an enterprise. Because this variable is zero for many enterprises in our sample, a tobit regression is the natural tool.⁸ With several actors making interdependent decisions in the privatization process, the regression with residual state shares as the dependent variable should be regarded as a reduced form, summarizing a politico-economic equilibrium. The independent variables included in the regression are listed in the first column of table 18.2.

Because a fairly conservative government was in power, one might expect the regression results to reflect the residue of the socialist idea. That is, the state would be expected to retain activities that had primacy and cachet under socialism, particularly large enterprises in heavy industry. But this idea secures only minimal support. In the regression, heavy industry is the omitted sector dummy. For the six

Table 18.2
Patterns in the residual state share in Mongolian enterprises

(a)	(b)	(c)	Tobit regression: percentage state share is the dependent variable	
			(d)	(e)
	Mean values	Standard deviation	Tobit coefficient	Marginal effect
Number of employees	344.609	474.500	-0.009 (-0.67)	-0.001
Total capital stock, in millions of tugriks	44.509	75.432	0.329 (3.90)	0.031
Total debt, in millions of tugriks	19.609	41.504	-0.364 (-1.75)	-0.035
Profits per share, thousands of tugriks	0.048	0.117	-493.119 (-2.98)	-46.846
International Contacts: dummy = 1 if firm had contacts with foreign firm in 1990	0.517	0.503	-4.821 (-0.63)	-0.458
State orders as a percentage of output in 1991	48.057	44.554	0.250 (3.11)	0.024
Divestiture: dummy = 1 if firm was a spin-off from a larger firm	0.310	0.465	12.298 (1.22)	1.168
Influence of workers on enterprise decisions in 1990, percentage scale	15.023	1.705	-6.123 (-2.51)	-0.582
Influence of management on enterprise decisions in 1990, percentage	43.839	6.050	-1.052 (-1.73)	-0.100
Date of approval of privatization plan, days after 1/1/91	420.920	98.316	-0.171 (-2.91)	-0.016
Delay between plan approval and sale of shares, in days	121.897	76.002	-0.099 (-1.71)	-0.009
Sector dummy: light industry	0.276	0.450	11.702 (0.70)	1.112
Sector dummy: agricultural processing	0.069	0.255	56.724 (3.03)	5.389
Sector dummy: construction	0.264	0.444	39.812 (3.17)	3.782
Sector dummy: transportation	0.092	0.291	58.844 (3.68)	5.590
Sector dummy: service	0.092	0.291	-143.531 (0.00)	-13.635
Sector dummy: wholesale trade	0.034	0.184	75.996 (3.56)	7.220
Intercept			187.773 (3.16)	
Dependent variable: percentage of shares owned by the state	16.13	21.108		

Notes: There are eighty-seven observations. T-statistics are in parentheses.

sector dummies included in the regression, five coefficients are positive and the only negative coefficient has the lowest *t*-statistic.⁹ Apparently, heavy industry was less attractive to the state than most other sectors, including such poor cousins of socialism as agricultural processing and wholesale trade.

Of course, the size of an enterprise was also a positive feature under socialism, and it would not be surprising if larger enterprises were more likely to be retained by the state. This notion receives mixed support. When size is judged by the number of employees, there is a negative sign on the pertinent coefficient—larger enterprises have a smaller state share. If size is measured by capital stock, the pertinent coefficient is positive.

There is another interpretation of the capital stock variable, however, one that hangs on the particular circumstance of Mongolia, a very poor country that was the recipient of huge amounts of Soviet aid directed at the large enterprise sector. Enterprises with a large capital stock and low debt would have been most favored in the past by such aid, having been built in the context of the Soviet-inspired program of industrialization. Therefore, the signs of the coefficients of the capital stock and debt variables suggest that the government retained holdings in the enterprises that were most favored under the old system. In turn, these enterprises are the ones that are most likely to be vulnerable in the new era, after the demise of the Soviet system and of the CMEA.

Is there evidence that the state maintains ownership in those enterprises likely to be most vulnerable during reforms?¹⁰ There are several variables in the regression that capture the likely vulnerability of enterprises in the new market environment. First, the volume of state purchase orders in 1990 indicates the degree to which enterprises formerly relied on government purchases.¹¹ Second, pre-reform international contacts are evidence of preparation for the new world of an open economy. Third, the profits-per-share variable measures the profitability of an enterprise on the eve of the privatization process, a measure that certainly reflects some of the difficulties to be faced in the new environment. Fourth, a dummy variable for enterprise divestiture captures the vulnerability of an enterprise that has had existing ties broken in the restructuring process.¹² The signs of the coefficients for all these variables are consistent with the hypothesis that the state retains a larger share of ownership in those enterprises most likely to be vulnerable in the new environment.

In our interpretation of the results so far, we have concentrated on the motives of the government officials making privatization decisions. This is natural because any description of the formal aspects of the Mongolian privatization process would suggest that enterprise insiders could not affect such decisions. The administrative

aspects of the program were very centralized. The virtual absence of concessions to workers and managers suggests that the political power of insiders was minimal. Nevertheless, enterprise management could decide whether to cooperate with the privatization process, a power that they might have been able to use to press their concerns.

Two sets of variables examine whether there are any discernible effects of enterprise bargaining power in the patterns of residual state shares. First, our enterprise survey provides measures of how influential an enterprise's workers and management were within the pre-privatization enterprise, relative to government.¹³ Second, administrative delays (in approval of the privatization plan and between the approval of the plan and the sale of shares) reflect the power of enterprises to resist central dictates.¹⁴

The results suggest that the measured power of workers and managers within enterprises is negatively related to the size of state ownership after privatization. Also, the two variables proxying the ability or the willingness to resist privatization—the approval time and the delay variables—have signs consistent with an inverse relationship between state share and enterprise power. These results suggest that enterprise insiders wanted smaller state ownership in their enterprises and that they were able to use their bargaining power during implementation stages, even when these insiders were not able to achieve their goals in the formal political process.

5 Alternative Interpretations

The interpretations of the results that we have provided so far have been based on rather general understandings of the goals of actors involved in the privatization process. Now we turn to more precise recent theories that might be used indirectly to aid in the interpretation of the results. In particular, we examine predictions arising from the theories of Perotti (1995) and of Shleifer and Vishny (1994) and Boycko, Shleifer, and Vishny (1996).

Perotti (1995) reflects on why governments of nontransition countries might retain part ownership in privatized enterprises, when enterprises are being sold for cash. His model rests on asymmetric information. Private investors are uncertain about whether the government intends to intervene in the future in a manner that will reduce the profits of privatized enterprises. That is, investors are uncertain whether the government is a committed one (to policies of nonintervention) or a populist one. A committed government can credibly convey this information by

retaining part ownership for some time. Temporary residual state ownership maximizes the value of privatization sales made by the committed government, because a higher price is extracted when the retained shares are sold in a later period, after the government has been able to present credible evidence of its commitment through the implementation of policy that affects the return on the shares. In contrast, a populist government maximizes its receipts from privatization by selling all shares immediately, because the value of any retained shares would fall once the government necessarily showed its hand.

Given that Mongolian privatization sales were not for cash, but rather for vouchers, there remains the question of whether Perotti's analysis is pertinent here. The question of pertinence hangs on whether the Mongolian government had a strong positive incentive to collect vouchers speedily (by selling all shares for vouchers rather than retaining some ownership). If the government had such an incentive, then it faced a trade-off analogous to that in Perotti's model—whether to maximize the immediate collection of vouchers or to retain some shares in order to sell them later for cash.

In the Mongolian case, it is clear that the government did have a very strong incentive to collect vouchers immediately. The success of the privatization program was politically very important, and an indicator of that success would be the number of citizens who had used their vouchers. Moreover, the fewer vouchers left in the hands of the populace after the initial phases of the privatization program, the less political pressure there would be at a later date to place additional assets on sale for vouchers. The voucher sale of such assets would be costly to the government, given that these assets would also have a cash value. Therefore, we can conclude that the Mongolian privatization did have characteristics that conform approximately to those of Perotti's model.¹⁵

Perotti's (1995) predictions do help in offering alternative interpretations of some of the results in table 18.2. For example, if there is a sequence of privatizations, he predicts that complete sales will come later in the privatization process, after a committed government's intentions are clear. This conjecture is obviously consistent with the negative coefficient on the variable measuring the amount of time that elapses between the beginning of reforms and the approval of an enterprise's privatization plan (the date at which the residual state share is finally determined.)

Perotti (1995, p. 856) also predicts that there will be higher residual state shares for enterprises that are earning rents from protected markets. Such enterprises offer larger potential opportunities for intervention, therefore tempting a populist government and offering greater gains on the residual state share for a committed government that is able to establish its credentials. This prediction is not consistent with

the results of table 18.2. If profits are any indication of rents, then our results indicate that enterprises with higher rents have a smaller state share, a result that is inconsistent with Perotti's conjectures.

We now turn to an examination of the theories of Shleifer and Vishny (1994) and Boycko, Shleifer, and Vishny (1996), who explore the interaction between politicians and enterprise managers, predicting when this interaction will be most damaging to general economic efficiency. The larger state ownership is, the more likely that politicians and the enterprise manager will make an agreement that results in an inefficient outcome.¹⁶ If a committed reforming government had this analysis in mind, then it would undertake privatization with a view to binding any future government that might be less reform-minded. Privatization will make it more costly for future politicians to strike efficiency-reducing agreements.

This theory would provide predictions pertinent to the present analysis, if the reforming officials faced limitations on the total amount of privatization that they could accomplish but had flexibility in allocating the residual state share across enterprises. The limitations might arise from political factors, for example. Then, a government committed to reform would apportion the residual state shares so that the privatization process as a whole had the maximum binding effect on future governments. Hence, there would be complete privatization for those enterprises most likely to strike efficiency-reducing political deals with future noncommitted governments and a positive residual state share in those enterprises least likely to offer scope for such efficiency-reducing bargains. For example, Shleifer and Vishny (1994, p. 1018) argue that more profitable enterprises should be privatized first, in as much as larger profits expand the scope for efficiency-reducing bargains.

The prediction ending the previous paragraph is essentially the opposite of the second of the predictions of Perotti (1995), noted above. In Perotti's case a committed government retains higher stakes in enterprises that offer the most scope for political deals in the future, because such enterprises give a committed government most chance to show its pedigree. In our use of Shleifer and Vishny (1994) and Boycko, Shleifer, and Vishny (1996), a reform government tries to restrain a future, less committed government, therefore trying to minimize the scope for future political deals. Hence, there is no essential inconsistency between the two theories. Rather, the two theories embody different assumptions about the expectations of the privatizing government concerning the reform orientation of future governments.¹⁷

In fact, most of the variables in the regression of table 18.2 can be given an interpretation in terms of the model of Boycko, Shleifer, and Vishny (1996). Two essential elements of that model are used here. First, there is the assumption that explicit

subsidies from the state treasury to finance inefficiencies are more costly to politicians than a loss in state-enterprise profits due to an efficiency-reducing bargain. Hence, efficiency-reducing agreements are most likely to occur between politicians and managers of profitable enterprises. Second, the benefits to a politician of aiding an enterprise are positively related to the political power of the workers in that enterprise, implying that politicians are more likely to strike efficiency-reducing bargains with an enterprise with greater political power.

We use this logic in table 18.3 to make predictions concerning the signs of the coefficients on the variables that appear in table 18.2. The logic in each row of table 18.3 follows in a straightforward manner from the model of Boycko, Shleifer, and Vishny (1996).¹⁸ For each enterprise, there is a critical value of the residual state share below which political intervention ceases. This critical value will be lowest in profitable enterprises with workers who are more politically powerful. Given this general logic, the second column of table 18.3 then presents the essential elements of the reasoning for each variable included in the regression of table 18.2. The third column provides the prediction on whether increases in the variable will lead to a greater or a lesser likelihood of efficiency-reducing agreements. Enterprises that have high values for variables that have a "positive" in the third column are then exactly the ones that a committed government would be most eager to privatize, thus leading to a "negative" prediction in the fourth column. This prediction is then matched against the signs of the coefficients in table 18.2, which for convenience are repeated in the fifth column of table 18.3.

As is immediately obvious from table 18.3, there is a high degree of conformity between the predictions loosely derived from the analysis of Boycko, Shleifer, and Vishny (1996) and the results in table 18.2. Certainly, there is a strong case to be made that the vision of the reformers was consistent with that of a reformist government following the implications for privatization of the bargaining model of Boycko, Shleifer, and Vishny (1996).

6 Conclusions

This chapter began by noting the empirically based inductive methodology that has been central to the work of János Kornai, a methodology that has been used in the essay. In interpreting our results, we have also used themes closely connected to ideas that have been central to Kornai's economic vision. In noting that more vulnerable enterprises had larger residual state shares, we have suggested that the reformist government might have been seeking to protect such enterprises from the

Table 18.3

Applying the logic of Boycko, Shleifer, and Vishny (1996) to the interpretation of the variables describing patterns in the residual state share in Mongolian enterprises

(a)	(b)	(c)	(d)	(e)
Variable included in the regression in table 18.2	Essential element of the logic relating this variable to the bargaining between enterprise managers and future noncommitted governments.	Effect of an increase in this variable on the likelihood that future governments would strike efficiency-reducing bargains after incomplete privatization	Predicted sign of the coefficient in the regression in table 18.2, if a reforming government wants to constrain a future government	Actual sign of the coefficient in the regression in table 18.2
Number of employees	The political power of an enterprise increases more than proportionately with the number of the employees and therefore an efficiency-reducing bargain with a larger enterprise is more attractive to the politician.	positive	negative	negative
Total capital stock	Enterprises with the largest capital stocks were built with Soviet aid and were more closely tied to the old Soviet system. These enterprises would suffer most during transition and be least profitable, reducing the scope for efficiency-reducing bargains.	negative	positive	positive
Total debt	The existence of enterprise debt implies an additional instrument (debt forgiveness) that could be used furtively to aid the enterprise. Therefore enterprises with high debt offer more scope for efficiency-reducing bargains.	positive	negative	negative

Table 18.3 (continued)

(a)	(b)	(c)	(d)	(e)
Profits per share	The higher are profits the more scope there is for an efficiency reducing bargain to occur (Shleifer and Vishny, 1994, p. 1018).	positive	negative	negative
International contacts in 1990	The more international contacts, the more likely that the enterprise will be profitable. The logic in the preceding row of the table then holds.	positive	negative	negative
State orders as a percentage of output in 1991	Enterprises that relied on state orders in the past are more likely to be unprofitable in the future, leading to the same logic as in the previous two rows of the table.	negative	positive	positive
The enterprise was a spin-off	If an enterprise was split up, then both its political power and its profitability would be diminished, reducing the attractiveness of efficiency-reducing bargains to the politician.	negative	positive	positive
(1) Worker influence (2) Management influence (3) Delay in approval of privatization plan (4) Delay in share sale	These four variables represent, or are proxies for, the political power of the enterprise. An enterprise that is politically strong provides greater benefits for the politician when efficiency-reducing bargains are struck.	positive	negative	negative
Sector effects, relative to heavy industry	The heavy industrial sector faces the most problems of adjustment and therefore is least likely to be profitable, leading to less scope for efficiency-reducing bargains in the future.	positive	negative	positive

storms that were inevitably coming.¹⁹ Thus, the state was motivated by paternalism, an objective that is much emphasized in Kornai's vision of political economy (Kornai 1980).

We have also shown that our results can be given an alternative interpretation, one that uses major themes of Kornai's work—the bargaining between politicians and managers that results in soft budget constraints. Whether we assume that reformist politicians want to protect vulnerable enterprises or to bind future governments, that assumption is embedded in a implicit framework in which politicians and enterprise managers bargain over state intervention. Such bargaining, of course, is the essence of the process that leads to the soft budget constraints, which Kornai's work has helped to place at the center of both academic and practical economic deliberations.

The implications for the effect of privatization on the economy are therefore unclear. If giving aid to dying enterprises offers large political benefits to politicians, then one could argue that the Mongolians conducted privatization in a manner that will lead to much state intervention in the future. If, following Boycko, Shleifer, and Vishny (1996), the absence of profits makes intervention very costly to politicians, then one could argue that the Mongolians have conducted privatization in a manner that will lead to little state intervention in the future.

Notes

Georges Korsun is affiliated with Chemonics International, Washington, D.C., and Peter Murrell is a member of the Department of Economics and the IRIS Center, University of Maryland, College Park. We thank Jim Anderson, Ts. Batbold, Peter Boone, Minjin Hashbat, Mike Reynolds, Randi Ryterman, and Yanjema for their generous help, and Josef Brada, Bruce Kogut, John McMillan, and Ronald McKinnon for their comments on an earlier version of this paper. We are deeply indebted to many Mongolian officials, enterprise directors, and academics without whose assistance the research for this paper could not have been accomplished. This paper was made possible through support provided by the U.S. Agency for International Development under Cooperative Agreement No. DHR-0015-A-00-0031-00 to the Center on Institutional Reform and the Informal Sector (IRIS).

1. The factual information presented in this paper is based on interviews, numbering in the hundreds, conducted by the authors. Although there are no citations to such sources, any interested reader can contact the authors as to the source of the facts.
2. On other reforms, progress was slower. See Boone (1994) for a discussion of stabilization and Murrell, Dunn, and Korsun (1996) for the development of price liberalization.
3. These are the privatizations that are the subject of the present study. Since the data were collected, privatization has proceeded, albeit at a slower pace. The population of large privatized firms has now reached 500.
4. For details of the subsequent voucher-auction process and its consequences, the reader is referred to Korsun (1995) and Korsun and Murrell (1995a).
5. Indeed, many important participants were not aware of the overall dimensions of residual state share.

6. Data on ownership patterns across important groups are surprisingly hard to obtain. Earle and Estrin (1994) present a picture of worker ownership for a sample of the leading major reforming countries, while Pistor and Turkewitz (1994) do the same for residual state ownership. Blasi (1994) presents data based on a sample of Russian enterprises.
7. The first privatization was completed in March 1992. A new government was elected in July 1992, but the privatizations included in our data were prepared before the new government changed any significant elements of privatization policy. Therefore, our data reflect the decisions of the first reform government, which is briefly described in the second section of the paper.
8. In the regression presented, thirty-six of eighty-seven observations are non-limit observations.
9. Given the process of generating the final form of the regression, the use of the *t*-statistics in the reported results is for descriptive purposes, rather than for conducting tests. To judge the magnitude of the effect of the variable, information is given in the fifth column on the marginal effect (Maddala 1983, p. 60).
10. We phrase the causal mechanism here as the state retaining ownership of more vulnerable enterprises, because this seems to be the most natural way to frame the motives of actors, based on our conversations with officials. There is an alternative phrasing that could be adopted, based on the work of Shleifer and Vishny (1994) and Boycko, Shleifer, and Vishny (1996), which we discuss in the next section.
11. Note that this is not a measure of *present* ties to the state. By 1993, state orders were at much smaller levels than in 1990. This is a measure of *previous* ties to the state.
12. The results of Lizal, Singer, and Svejnar (1994) for the Czech Republic confirm that performance depreciates when enterprises are broken up. Moreover, a number of these Mongolian spun-off enterprises have reemerged in subsequent years in the face of economic pressures.
13. See Korsun and Murrell (1995b) for the details of the construction of this variable. In the survey, enterprise directors were asked to give retrospective information, from which this variable was constructed. The retrospective aspect of the variable introduces the possibility that memories are affected by events ongoing at the time of the survey. We feel that this danger of bias is less than that from omitting the variable.
14. This follows information gained in interviews with privatization officials.
15. Perotti (1995, p. 856) suggests caution in applying his model to mass privatization in view of the endogeneity of policy preferences during a systemic transformation.
16. We use the element of the analysis in which it is assumed that managers exercise control rights over the enterprises, an assumption appropriate for Mongolia in the era of privatization.
17. Conversations with government officials lead us to conclude that the Boycko, Shleifer, and Vishny (1996) assumption was the more appropriate one for Mongolia.
18. Convenient extra assumptions are that a 100 percent state share would lead to efficiency-reducing bargains in all enterprises, and that a 100 percent private share would end all such bargains. With this extra assumption, it is easy to see that the allocation of residual state share across enterprises is a constrained optimization problem with the solution allocating higher residual state shares to those enterprises that are least likely to strike efficiency-reducing bargains with politicians.
19. The logic that higher state shares serve to protect vulnerable enterprises follows immediately from the Boycko, Shleifer, and Vishny (1996) framework if we adopt Kornai's assumption of paternalism and assume that rescuing a failing enterprise has much greater political benefits than aiding any other enterprise.

References

- Blasi, Joseph (1994). Ownership, governance, and restructuring. In Ira W. Lieberman and John Nellis, eds., *Russia: Creating Private Enterprises and Efficient Markets*. Washington, D.C.: The World Bank, pp. 125–140.

- Boone, Peter (1994). Grassroots macroeconomic reform in Mongolia. *Journal of Comparative Economics* 18(3): 314–328.
- Boycko, Maxim, Andrei Shleifer, and Robert W. Vishny (1996). A theory of privatization. *Economic Journal* 106: 309–319.
- Earle, John, and Saul Estrin (1996). Employee ownership in transition. In Roman Frydman, Cheryl W. Gray, and Andrzej Rapaczynski, eds., *Corporate Governance in Central Europe and Russia: Insiders and the State*, volume 2, 1–61. Budapest: Central European University Press.
- Kornai, János (1959). *Overcentralization in Economic Administration: A Critical Analysis Based on Experience in Hungarian Light Industry*. Oxford: Oxford University Press.
- Kornai, János (1980). *Economics of Shortage*. Amsterdam: North-Holland.
- Kornai, János (1990). The affinity between ownership forms and coordination mechanisms: the common experience of reform in socialist countries. *Journal of Economic Perspectives* 4(3): 131–147.
- Kornai, János (1992). The postsocialist transition and the state: reflections in the light of Hungarian fiscal problems. *American Economic Review* 82(2): 1–21.
- Kornai, János (1993). The evolution of financial discipline under the postsocialist system. *Kyklos* 46(3): 315–336.
- Kornai, János (1994). Transformational recession: the main causes. *Journal of Comparative Economics* 19(1): 39–63.
- Korsun, Georges (1995). Inside Ownership, Internal Influence, and Enterprise Behavior: Evidence from the Mongolian Large Privatization. Ph.D. dissertation, University of Maryland, College Park.
- Korsun, Georges, and Peter Murrell (1995a). The politics and economics of Mongolia's privatization program. *Asian Survey* 35(5): 472–486.
- Korsun, Georges, and Peter Murrell (1995b). History versus Policy: How Much Does Enterprise Governance Change After Mass Privatization? IRIS Working Paper, University of Maryland.
- Lizal, L., M. Singer, and J. Svejnar (1994). Manager Interests, Breakups and Performance of State Enterprises in Transition. Paper presented at the Workshop on Enterprise Adjustment in Eastern Europe. Washington, D.C., The World Bank.
- Maddala, G. F. (1983). *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge: Cambridge University Press.
- Murrell, Peter, Karen Turner Dunn, and Georges Korsun (1996). The culture of policy making in the transition from socialism: price policy in Mongolia. *Economic Development and Cultural Change* 45(1): 175–194.
- Perotti, Enrico (1995). Credible privatization. *American Economic Review* 85(4): 847–859.
- Pistor, Katharina, and Joel Turkewitz (1996). Coping with hydra-state ownership after privatization: a comparative study of the Czech Republic, Hungary, and Russia. In Roman Frydman, Cheryl W. Gray, and Andrzej Rapaczynski, eds., *Corporate Governance in Central Europe and Russia: Insiders and the State*, volume 2, 192–244. Budapest: Central European University Press.
- Shleifer, Andrei, and Robert W. Vishny (1994). Politicians and firms. *Quarterly Journal of Economics* 109(4): 995–1025.

19 Do Power Consumption Data Tell the Story? Electricity Intensity and Hidden Economy in Postsocialist Countries

Mária Lackó

The hidden economy is universal. It is present in any economic system, in developed market economies, in socialist and postsocialist countries, and in developing economies. In spite of such a universality, in the different systems the direct causes of the hidden economies are different. In *developed market economies* the main factors that give rise to the hidden economy are high tax rates, the burden of regulation by the state, varying possibilities of employment for labor in the official economy, and low tax morality. In *developing countries* the causes are the same, though their weights are different: the major determinant is the lack of opportunities for the agents to participate in the official business and the official labor market.

The causes of emergence and perpetuation of the hidden economy in *socialist economies* were entirely different from those in market economies. In fact, the habitual factors of market economies could not be significant in the socialist environment because of the extreme narrow scope of formal taxation and the full employment of the labor force. Major factors triggering the expansion of the hidden economy in socialist countries were the pressure of permanent shortages, the negligible size of the legal private economy, the low level of services provided by the state, and the lack of competition (see Kornai 1992a).

In the *postsocialist system*, a gradual restructuring of the factors that trigger the hidden activities takes place: the factors that had been characteristic of the socialist system lose importance, while those characterizing the market system gain ground. Major new developments such as the disappearance of shortages, the emergence of competition, a gradually expanding market of services, and the growing role of the legal private economy (Kornai 1995) all lead to the disappearance of the earlier causes of hidden economy and give way to the traditional causes present in market economies, such as gains from tax evasion and from the black-market labor of unemployed and other inactive people.

In the course of their gradual restructuring, postsocialist countries also establish financial discipline and tax discipline, although very slowly. "In the socialist system discipline was enforced by the bureaucracy itself, many times with arbitrary and brutal means," Kornai writes (1993, p. 326). Distrust and indifference toward the state characterized the behavior of the citizens. In the early 1990s, in the course of the transition process, the means of enforcement have gradually changed to those of the constitutional state. However, to establish trust in the state takes much longer than changing the regulatory system (see Kornai 1992b, 1993). In recent years, as a consequence of economic liberalization and continuing distrust in the state, the scale of hidden activities has grown fast in these economies.

In the literature of transition economics the size of the hidden economy has a prominent place, and for several reasons. First, the size of the hidden economy has far-reaching consequences for the actual, as opposed to the registered, rate of output decline. Second, the pervasiveness of hidden activities greatly determines the size of missing revenues from the government budget, which is of particular importance in times when stabilization endeavor seeks all possible sources of budgetary revenues.

Once as a student, and later as a colleague and friend, of János Kornai I got acquainted with the characteristic features of socialist and postsocialist economies under his guidance. One of the general principles that I learned from him during these twenty-eight years was that for the description, analysis, and understanding of economic phenomena *measurement* is indispensable, even in the case of arcane concepts such as the hidden economy.

The measurement of the size of the hidden economy has not been solved in any economic system, let alone for postsocialist economies. Many estimation methods have been established and applied that have produced results that support and complement, but many times contradict, each other. A consensus about the reliable and acceptable approaches is far from being established.

In this chapter I deal with such methods of estimation of the hidden economy, which rely to a great extent on aggregate electricity consumption data. The application of these methods has become widespread, and its results are frequently cited in the transition literature. Following the presentation of the method devised by Dobozi and Pohl (1995) and Kaufmann (1995*a*, 1995*b*), I show why these methods are inappropriate for measuring the hidden economy in postsocialist economies. At the end of the chapter, as an alternative, I briefly present my own model, which is based (partly) on the development of residential electricity consumption, instead of aggregate electricity consumption.

1 The Method of Dobozi and Pohl to Measure the Change in the Size of the Hidden Economy

In 1995, two supplementary papers were published in the newsletter *Transition* with talkative titles: “Real Output Decline in Transition Economies—Forget GDP, Try Power Consumption Data!” (Dobozi and Pohl 1995), and “Electricity Consumption and Output Decline—An Update” (Dobozi 1995). In these papers the authors, István Dobozi and Gerhard Pohl, claim that in the postsocialist countries, especially in the case of countries of former Soviet Union (FSU) the statistically registered

drop in GDP after 1989 exaggerates the actual decline, possibly by a very large margin. They state that in market economies, aggregate economic activity and electric power consumption usually move in lockstep (with an electricity-GDP elasticity close to one). Data for postsocialist countries do not show the same pattern, and the huge difference between the growth/decline of electric power consumption and GDP cannot be explained rationally across postsocialist countries without assuming a rapid growth of the hidden economy. The papers of Dobozi and Pohl have not remained without effects: they have become recurrent motifs of serious written analyses (EBRD 1995, for example) and of noted conferences dealing with postsocialist transition.

In the period 1989–1994 in Eastern European transition economies (with the possible exception of the Czech Republic), the cumulative decline in power consumption closely matches the drop in GDP, yielding an electricity-GDP elasticity of about 1 (i.e., that 1 percent GDP fall was associated with a 1 percent drop in electricity consumption). As Dobozi and Pohl write, “Even in those East-European countries where the economic structure and product lines changed drastically, as in Poland, the correlation between power use and economic activity remained fairly close” (Dobozi and Pohl 1995, p. 10).

In most FSU countries however, the reported output declines are completely inconsistent with the power consumption trends; thus, according to Dobozi and Pohl, the reliability of official statistics has to be questioned seriously. The gap between increasing electricity consumption and falling GDP can be explained only by a gross underreporting of the GDP. In the period 1989 and 1994, output downturn in Russia and Ukraine may have been inflated by official statistics more than twofold, in Azerbaijan as much as threefold, and in Georgia, Kazakhstan, Latvia, and Moldova by 50 to 90 percent.

Several factors distort official statistics, such as:

- a widespread underreporting of output in order to avoid high taxes;
- overrepresentation of large state-owned industrial enterprises that are undergoing major retrenchment; and
- shortcomings of data collection in capturing ever increasing private activities (Dobozi 1995, p. 19).

2 The Kaufmann Method

Based on the above considerations, Daniel Kaufmann (1995*b*) already reported the results of some concrete calculations concerning how, with what speed, and to what

size the hidden economy grew in Ukraine between 1989 and 1994. His concept is summarized the following way:

To measure overall economic activity in an economy, electric power consumption is usually the single-best physical indicator of economic activity. Overall economic activity and electricity consumption have been empirically observed throughout the world to move in lock-step—with an electricity/GDP elasticity close to one (Dobozi and Pohl 1995). For the specific case of economies in transition a careful consideration of all special features (very low electricity prices, with gradual adjustment upward; untapped efficiency improvement potential; and restructuring towards less electricity intensive activities) suggests that on balance electricity efficiency per unit of overall GDP may increase somewhat overtime (that is, there may be less than unitary elasticity during the transition). While still being a good proxy, this means that the changes in electricity consumption may to some extent underestimate the changes in overall GDP.

On the basis of the subtraction of the two variables measuring overall and official GDP, respectively, we can arrive an estimate of the unofficial economy. If the estimate of changes of overall GDP, approximated by the rate of change of electric power consumption, is somewhat underestimated, it means that the unofficial economy will also be somewhat underestimated. Consequently, this estimate of the unofficial economy (derived by subtracting the proxy for the overall economy from official GDP, if biased, is likely to be a conservative estimate. (Kaufmann 1995*b*, p. 1)

As a starting point of estimating the size of the unofficial economy in Ukraine, Kaufmann uses the results of the well-known Berkeley-Duke research on the second economy of the USSR (Alexeev et al. 1987) conducted during the late 1980s. That study had put the size of the second economy roughly in the range between 8 percent and 16 percent of the total economic activity. Thus, for purposes of his calculation, Kaufmann uses the midpoint estimate of 12 percent, assuming that this share of all the economic activities had been unofficial in 1989.

The next step in his calculation is to derive the figures for the overall GDP proxy, based on overall electricity consumption. He also computes the growth indices of official GDP. The figures in table 19.1 are Kaufmann's proxy variables for the calculations on the overall and official GDP, respectively. They incorporate the base-line 1989 estimate for the unofficial economy of 12 percent. Table 19.2 presents Kaufmann's calculations in index numbers on the evolution of the Ukrainian unofficial economy in the period 1989–1994, starting with 12.0 in 1989 (the base index is 100.0 for the overall economy). The first and third rows come from the calculations in table 19.1, while the second row (unofficial economy) is the difference between the third and the first rows.

From table 19.2 it can be seen that

Table 19.1

Changes in electricity consumption (as a proxy of overall GDP) and official GDP in Ukraine, 1989–1994

	1989	1990	1991	1992	1993	1994
Growth rate in electricity consumption		0	–2.2	–6.2	–7.8	–11.7
Electricity consumption index (1989 = 100)	100	100	97.8	91.7	84.6	74.7
Growth rate in official GDP		–3.8	–13.4	–17.5	–14.9	–24.5
Official GDP index (index 1989 = 88)	88	84.5	73.2	60.4	51.4	38.8

Source: Kaufmann (1995*b*).**Table 19.2**

Evolution of the official and unofficial economy in Ukraine, 1989–1994 (percent)

	1989	1990	1991	1992	1993	1994
Official economy index	88	84.5	73.2	60.4	51.4	38.8
Unofficial economy index	12	15.5	24.6	31.3	33.2	35.9
Overall economy index	100	100	97.8	91.7	84.6	74.7

Source: Kaufmann (1995*b*).

- by 1994 the unofficial economy tripled, while the official economy contracted to less than half its size in 1989;
- the decline in the output of overall economy since 1989 was one-fourth, which, while significant, is still less than half the decline derived from official statistics. The decline in the official economy was mitigated by the rapid growth in the unofficial economy during the period.

On the basis of table 19.2 Kaufmann calculates the relative shares of the official and unofficial economies. According to table 19.3, by 1994 the estimated share of the unofficial economy in the overall Ukrainian economy was 48.1 percent. Kaufmann adds that this result is consistent with microsurvey estimates.

The aforementioned estimation method of the unofficial economy is very simple and appealing. However, as Dobozi confesses himself, there are people who are skeptical about it: “Although our article was generally welcomed as being on the right track to obtain more reliable—and certainly low-cost—estimates of the extent of output retrenchment during the systemic transition, some skeptics argued that while power consumption and economic activity tend to move in tandem in market economies, it may not be relevant for transition economies that are experiencing

Table 19.3

Relative shares of official and unofficial GDP in Ukraine in 1989–1994 (percent)

	1989	1990	1991	1992	1993	1994
Official GDP	88.0	84.5	74.9	65.9	60.8	51.7
Unofficial GDP	12.0	15.5	25.1	34.1	39.2	48.1
Overall GDP	100.0	100.0	100.0	100.0	100.0	100.0

Source: Kaufmann (1995b).

rapid and massive structural changes. Many argue that the increase in electricity consumption may reflect structural movement toward higher electricity intensity in GDP" (Dobozi 1995, p. 19).

Under the impact of criticism, in his recent calculations with his coauthors, Kaufmann eases the strict assumption of unitary elasticity (Johnson, Kaufmann, and Shleifer 1997, cited by EBRD 1997). According to his new concept he assumes an elasticity of 0.9 for Eastern European countries, 1.0 for the Baltic countries, and 1.15 for CIS countries. With this modification, he and his coauthors use his method for estimating the hidden economy for seventeen countries. His results are presented in table 19.4.

I belong to the skeptics, despite the fact that I agree with the approach that uses indicators of electricity consumption for the estimation of the share of hidden economy in GDP. In fact, in another essay I also worked out an estimation method for this purpose, a method that is suitable for international comparison and that extensively makes use of the indicators of electricity consumption.

My skepticism derives from several factors. One cannot avoid asking: how is it possible that, according to the calculations of Kaufmann and his coauthors, the hidden economy did not grow in Romania and took a very low position in Uzbekistan during the years of transition, whereas in other countries the size of hidden activities was large and grew rapidly? It is also surprising that in 1994 and 1995 the ratio of the hidden economy to GDP is much smaller in Slovakia than in the Czech Republic. Likewise, this ratio also seems to be too low in Poland if compared to other reform countries and to anecdotal evidence.

The second kind of doubt derives from the examination of Finland's case. In the early 1990s Finland, just like the East European economies, and to some extent due to similar factors, experienced a significant fall of the GDP. Between 1990 and 1993 in Finland GDP decreased by 13.6 percent, while electricity consumption, far from decreasing, increased by 5.5 percent.

Table 19.4

Share of the unofficial economy in GDP, 1989–1995, selected transition economies (in percent)

	1989	1990	1991	1992	1993	1994	1995
Azerbaijan	12.0	21.9	22.7	39.2	51.2	58.0	60.6
Belarus	12.0	15.4	16.6	13.2	11.0	18.9	19.3
Bulgaria	22.8	25.1	23.9	25.0	29.9	29.1	36.2
Czech Republic	6.0	6.7	12.9	16.9	16.9	17.6	11.3
Estonia	12.0	19.9	26.2	25.4	24.1	25.1	11.8
Georgia	12.0	24.9	36.0	52.3	61.0	63.5	62.6
Hungary	27.0	28.0	32.9	30.6	28.5	27.7	29.0
Kazakhstan	12.0	17.0	19.7	24.9	27.2	34.1	34.3
Latvia	12.0	12.8	19.0	34.3	31.0	34.2	35.3
Lithuania	12.0	11.3	21.8	39.2	31.7	28.7	21.6
Moldova	12.0	18.1	27.1	37.3	34.0	39.7	35.7
Poland	15.7	19.6	23.5	19.7	18.5	15.2	12.6
Romania	22.3	13.7	15.7	18.0	16.4	17.4	19.1
Russia	12.0	14.7	23.5	32.8	36.7	40.3	41.6
Slovak Republic	6.0	7.7	15.1	17.6	16.2	14.6	5.8
Ukraine	12.0	16.3	25.6	33.6	38.0	45.7	48.9
Uzbekistan	12.0	11.4	7.8	11.7	10.1	9.5	6.5

Source: Johnson, Kaufmann, and Shleifer (1997), cited by EBRD (1997).

The growth of electricity intensity in Finland in 1990–1993 (22%), was not much less than the average of the eighteen postsocialist countries in 1989–1994 (32%). Within the postsocialist group there were differences, with East European countries showing the lowest growth (0.5%) and CIS countries the highest (55%). The Baltic countries, however, were very close to their Finnish neighbor (23.7%). Do the similarities suggest that, just like in the transition economies, Finland's growth of electricity intensity can also be explained by the sudden jump in the size of the hidden economy?

To see whether this can be true, let us carry out the same calculation for Finland that Kaufmann did for Ukraine. The initial share of the unofficial economy, which is relevant for 1990 (13%), was taken from my own estimation (Lackó 1998).

If we follow the method of calculation suggested by Kaufmann, the share of the hidden economy in Finland would increase from 13 percent to 30 percent in three years, that is, its relative size would more than double. We get a similarly surprising and unrealistic result if the starting share is 6.5 percent instead of 13 percent. In that case the share of the hidden economy in 1993 would be 24.7 percent, which would

mean an even more radical change, in that the result would be almost four times that of the starting size.

These results are astonishing and, given all the general knowledge about Finland, simply cannot reflect reality. The calculation was presented in order to show that the method by which the results were derived is rather debatable. In fact, the growth of electricity intensity in Finland is connected with the normal effects of recessions, rather than the expansion of the unofficial economy: in recessions electricity consumption decreases not as much as the GDP, because the fixed (overhead) electricity use does not contract in proportion to the drop in general capacity utilization. This effect is also mentioned by Dobozi and Pohl when they discuss electricity intensity of countries in transition; however, they add, "It is plausible to assume that the consumption-increasing effect of this factor was largely offset by the combined impact of higher electricity tariffs and shifts in the output mix away from heavy industry" (Dobozi and Pohl 1995, p. 18).

In the following two sections I will show that in the eighteen postsocialist countries under investigation it is mainly the differences in the size of structural changes of the economy that determine the differences in changes of electricity intensity, rather than the expansion of the hidden economy.

3 Analysis of Changes in Electricity Intensity in Eighteen Postsocialist Countries in 1989–1994

It is obvious that at times of recession, due to smaller than usual capacity utilization, electricity intensity grows, as we saw in Finland's case. In postsocialist countries, however, additional factors are also present. According to OECD experts, since 1990, in individual countries of Eastern Europe, but particularly in FSU countries, the industrial sector has largely maintained its level of energy consumption, despite the fact that industrial production, in value-added terms, has fallen substantially (OECD 1994). Reasons for this included the following:

1. The ratio of industrial output contributed by energy intensive basic industries, including energy production, has actually increased.
2. Inefficient plants have not been closed, but instead have operated at lower part-load efficiencies. For instance, in Russia, according to Roxburgh and Shapiro (1996), the so-called excess wages tax actually operated as an employment subsidy, and thereby unemployment has been made internal to the enterprise: this tax has given a strong incentive for the firm to retain excess workers on low wages, rather than making them redundant. This kind of tax influences not only state-owned firms, but

also privatized ones; moreover, it was found that privatized firms made fewer workers redundant than state-owned companies (Standing 1994). Internal unemployment most probably leads to higher electricity intensity in the industry than external unemployment would do.

3. Electricity price rises have been limited in real terms, and nominal price adjustments have often been accompanied by consumers' refusals to pay.
4. Little or no investments have been made in new plants to improve electricity efficiency.
5. The decision-makers' ability to evaluate the energy-use implications of their investment choices is lacking. Under the centralized economic system, decisions on investments in fixed assets were based primarily on the available production capacity and plan targets for physical output, with little emphasis on productivity or efficiency. This legacy has produced industrial enterprise managers largely unfamiliar with cost-accounting procedures.

Based on this list of specific factors that determine electricity use in postsocialist economies we may assume that it is the rate at which structural changes take place in the individual postsocialist countries that will mostly explain differences in changes in electricity intensity across these economies. In the following I investigate this problem with the help of a cross-sectional examination of eighteen postsocialist countries. At this point I would like to emphasize that I do not exclude the possibility that the growth of hidden economy also influences the growth of electricity intensity. I only claim that the method initiated by Dobozi, Pohl, and Kaufmann is not the right one to apply for the periods and the economies that these authors wanted to use it for.

In my analysis three different indicators of structural change are defined:

1. The change in the share of industry in the production of GDP between 1989 and 1994: *ind* (measured in percentage points)
2. The difference between the decrease of electricity consumption in industry and the decrease of total electricity consumption in 1989–1993: *d* (in percentage points)
3. The maximal rate of unemployment between 1989 and 1994: *u* (percent).

We assume that the larger structural changes take place in a particular economy, that is, the faster the dismantling of industries of socialist type is carried out, the less growth in electricity intensity occurs. More precisely:

- the more the value added by the industrial branches decreases compared to the official GDP, the less growth in electricity intensity occurs,

- the more the electricity consumption in industrial branches decreases in comparison to total electricity consumption, the more pronounced the shift occurs from highly electricity intensive branches to branches with less electricity intensity, and therefore the less the electricity intensity of the economy grows;
- the higher the rate of unemployment (which is to a high degree structural unemployment), the larger structural changes occurred in the form of dismantling industries of socialist type, and therefore the slower the rise of electricity intensity in the economy.

Table 19.5 shows the average size of the individual structural indicators for the different groups of countries. In the last column the average change in electricity intensity is presented. The table shows that for the Central and Eastern European countries each indicator of structural change is 1.6–2 times as large as the average indicator of the full sample of eighteen countries. The indicators of the Baltic Countries are similar to the total average, whereas the indicators of the countries of the Commonwealth of Independent States are half as large as the average indicators. In the last column of the table it can also be noticed that, in line with our assumption, the bigger the structural change, the less is the growth in electricity intensity. The regularity suggested by the interrelations between country group averages, that is that the development of electricity consumption is significantly influenced by the scale of structural changes, is subjected to a more exact econometric analysis later.

Let us look first at the correlation between the indicators of the individual countries. From table 19.6 we can see that there is a close positive relationship between the change of electricity consumption and the change of the official GDP. We find also an interesting relation between two indicators of structural change: the larger the decrease of the share of industrial branches within the official GDP, the larger the rate of unemployment.

The table also shows an important correlation between the change of the official GDP and the structural indicator related to the electricity consumption: the less the

Table 19.5

Mean values of the indicators of structural change and electricity intensity by group of countries

	<i>ind</i>	<i>d</i>	<i>u</i>	<i>dint</i>
Central and Eastern Europe	-15.9	-18.1	12.3	1.005
Baltic countries	-8.3	-9.3	4.7	1.237
Commonwealth of Independent States	-3.5	-7.0	2.6	1.550
Total	-8.4	-11.0	6.2	1.320

Table 19.6

Correlations among the indicators of the countries

	de_i	$dgdpi$	$dinti$	$indi$	u_i	d_i
de_i	1					
$dgdpi$	0.81	1				
$dinti$	-0.5	-0.84	1			
$indi$	0.23	0.13	0.15	1		
u_i	0.1	0.45	-0.38	-0.75	1	
d_i	-0.47	-0.72	0.66	0.32	-0.39	1

Notes:

 de_i —the change in electricity consumption between 1989 and 1994. $dgdpi$ —the change of the official GDP between 1989 and 1994. $dinti$ —the change of electricity intensity between 1989 and 1994. $indi$ —the change of the rate of industry within the GDP. u_i —the maximal rate of unemployment between 1989 and 1994. d_i —the difference between the change of electricity consumption in the industrial branches and the change in the total electricity consumption (1989–1993). i —country.

electricity consumption in industrial branches decreases in comparison with total electricity consumption, in other words: the smaller the structural change in this field, the larger the fall is experienced in the official GDP. We have to be careful with the interpretation of this result. If electricity consumption in the industrial branches comprises a larger share of unregistered elements compared to that in total electricity consumption, then this would give sufficient explanation for a larger decline in the official GDP. In this case the growth in total electricity intensity would indeed be caused by the growth of the hidden economy. However, this assumption is most probably not well founded: if the electricity consumption due to nonregistered production is large and growing in the industrial branches, then at least the same would apply to the total electricity consumption, in as much as trade and service activities are much easier to keep unregistered.

From table 19.6 it is evident that the growth of electricity intensity is correlated with the measure of structural change related to electricity consumption in industry: the less the decrease in electricity consumption in the industrial branches compared to total electricity consumption, the more electricity intensity increases.

4 Econometric Analysis

In the following we formulate our hypothesis in an equation and attempt at verifying it with a cross-section econometric estimation.

Equation (19.1) expresses our already sketched hypothesis—that in addition to the changes in official GDP, it is the scale of the three kinds of structural changes that determine total electricity consumption. With this, of course, we do not state that the growth of unofficial economy does not influence electricity consumption or the growth of total GDP. We state only that the difference between the shifts in electricity consumption and the change of the official GDP is not that mystical and hidden, and it shows a strong relation to the scale of structural changes.

The equation is as follows:

$$de_i = \delta_1 dgdp_i + \delta_2 ind_i + \delta_3 d_i + \delta_4 u_i + \delta_5 \quad (19.1)$$

$$\delta_1 > 0 \quad \delta_2 > 0 \quad \delta_3 > 0 \quad \delta_4 < 0$$

where

de_i : the change in electricity consumption between 1989 and 1994

$dgdp_i$: the change of the official GDP between 1989 and 1994

ind_i : the change of the share of industry in the production of GDP between 1989 and 1994

u_i : the maximal rate of unemployment between 1989 and 1994

d_i : the difference between the change of electricity consumption in the industrial branches and the change in total electricity consumption, 1898–1993

i : country indicator. The countries are: Bulgaria, the Czech Republic, Hungary, Poland, Romania, Slovakia, Estonia, Latvia, Lithuania, Armenia, Azerbaijan, Belarus, Georgia, Kazakhstan, Moldova, Russia, Ukraine, and Uzbekistan.

In table 19.7 the results of a number of econometric estimations are shown. The estimations were carried out with the ordinary least squares method, based on the cross-sectional data of eighteen countries. The signs of the parameters are mostly as expected, and the indicators of fitting are acceptable. In the course of the estimations, three dummy variables were used for Romania, Georgia, and Uzbekistan.

Besides the structural shifts across sectors that are taken into account in equation (19.1), Romania's special feature was the structural change that took place *within* the industry. In the mid-1980s Romania had by far the largest petrochemical sector in Eastern Europe. Between 1989 and 1992 this industrial branch virtually collapsed: its output decreased by 50 percent. Due to the high electricity use in

Table 19.7Results of estimation of equation (1) (dependent variable: de_i)

<i>Independent variables</i>	(A1)	(B1)	(C1)	(D1)	(E1)	(F1)
$dgd p_i$	0.531 (5.43)	0.6535 (6.06)	0.7052 (4.22)	0.8542 (10.23)	1.002 (11.56)	1.0209 (11.04)
ind_i		0.0047 (2.46)				-0.001 (0.70)
d_i			0.0043 (1.31)		0.004 (2.72)	0.0043 (2.75)
$\ln u_i$				-0.077 (5.65)	-0.077 (6.92)	-0.084 (5.42)
<i>Dummy variables</i>						
Romania		-0.141 (1.91)	-0.193 (2.31)	-0.148 (3.18)	-0.164 (4.32)	-0.17 (4.27)
Georgia		0.044 (0.48)	-0.034 (0.35)	0.127 (2.04)	0.1556 (3.03)	0.1529 (2.90)
Uzbekistan		-0.099 (1.23)	-0.052 (0.59)	-0.297 (4.37)	-0.322 (5.78)	-0.334 (5.61)
constant	0.4055 (6.56)	0.3825 (7.84)	0.3643 (7.54)	0.3287 (7.84)	0.2839 (7.54)	0.2787 (7.11)
aR^2	0.627	0.76	0.69	0.902	0.936	0.933
F	29.57	11.84	8.43	32.28	42.5	34.83
Standard error of regression	0.0845	0.067	0.078	0.0434	0.035	0.036
Mean of dependent variable	0.7236	0.7236	0.7236	0.7236	0.7236	0.7236

petrochemical industries, this single specific development has contributed to the decline in Romania's total electricity intensity to a large extent (see OECD 1993).

In the case of Uzbekistan it is also the heritage of the past that explains the country's outlier position. During the Soviet era, Uzbekistan became the cotton-growing center of the former Soviet Union. The development of basic industries was neglected and the republic was among the least industrialized ones in the USSR (Ebel 1997). Cotton accounts for 70 percent of the country's exports and 50 percent of its GDP. Exactly because of this reliance on a single commodity, Uzbekistan was barely struck by the transitional recession, and the country's electricity intensity did not grow either. The latter was also connected with the fact that Uzbekistan, alone of all the republics of the former Soviet Union, managed to increase its oil and gas production in the first years of transition. Therefore, currently it has control over more energy sources than before transition, thereby capable of substituting those

for electricity. Georgia's special situation can be explained by the conflicts and wars in South Ossetia and Abkhazia, which lasted until 1994 and paralyzed the economy.

In estimation A1 we have not yet taken into account the differences in the scale of structural changes across the countries. According to the results obtained here a 1 percent decrease of GDP was accompanied on average by a 0.5 percent decrease in electricity consumption in the eighteen postsocialist countries. In fact, this estimation does not help the analysis; it only describes the investigated development.

In estimations B1, C1, and D1 we included the three different structural indicators individually. In all three cases the sign of the structural parameter showed up as expected; however, the significance of the parameter in estimation C1 is not appropriate. The reason for this is the multicollinearity caused by the close negative correlation between the variables of gdp_i and d_i . The fitting of the equation improved in all three cases in comparison to estimation A1, with the most spectacular improvement in the case of the inclusion of the rate of unemployment (estimation D1).

The parameter of the rate of unemployment is significantly negative. The negative sign is as expected, because we assumed that larger unemployment indicates a larger scale of structural change, which, *ceteris paribus*, reduces electricity consumption.

The assumption of a negative sign here is not, however, evident at all: we could have assumed and obtained a positive sign likewise due to the widely known close relationship between the size of the hidden economy and the level of unemployment. It is a tendency in developed market economies that the larger the rate of unemployment, the larger the size of the hidden economy. This process, in which the pool of the unemployed strengthens the activities in the hidden economy, will most probably gradually become more pronounced in the Eastern European countries. Inactive members of the families start to participate intensively in the hidden economy, and through this they contribute to the increase of electricity consumption and intensity. This growth of electricity intensity is blurred by the effect of the structural changes (the reduction of the usual role of socialist heavy industry) and the subsequent cut in electricity intensity. Our calculations testify that in the period 1989–1994, from the two effects structural changes turned out to be more powerful: this is shown by the significantly negative parameter in equations D1, E1, and F1.

The results of estimation E1 shows a further interesting result. If there was no difference between the eighteen postsocialist countries' speed of structural change, then GDP and electricity consumption would really closely move together (the value of parameter belonging to the variable $dgdp_i$ is 1.00).

In estimation F1 all three structural indicators are included. Here, due to the close negative relation between ind_i and u_i , multicollinearity arose. Because of the multicollinearity between $dgdpi$ and d_i , equation (19.1) had to be reformulated and reestimated. In the new function, equation (19.2), the dependent variable is the change of electricity intensity which is the function of the indicators of structural change.

$$dint_i = \theta_1 ind_i + \theta_2 d_i + \theta_3 u_i + \theta_4 \quad (19.2)$$

$$\theta_1 > 0 \quad \theta_2 > 0 \quad \theta_3 < 0$$

where

$dint_i$: the change of electricity intensity between 1989 and 1994,

We did a couple of estimations for equation (19.2), the results of which are shown in table 19.8. In these estimations the parameters proved to be significant and showed the expected signs (except for that of the variable ind_i in estimation F2

Table 19.8
Results of estimation of equation (2) (dependent variable: $dint_i$)

Independent variables	(A2)	(B2)	(C2)	(D2)	(E2)	(F2)
ind_i	0.0163 (2.98)				0.0091 (2.08)	-0.027 (1.04)
d_i			0.022 (4.58)	0.0129 (5.59)	0.0043 (2.75)	0.0133 (5.71)
$ln u_i$		-0.21 (6.98)		-0.156 (5.65)		-0.174 (6.78)
<i>Dummy variables</i>						
Romania	-0.254 (1.10)	-0.22 (1.61)	-0.383 (2.15)	-0.242 (3.22)	-0.289 (1.75)	-0.255 (3.36)
Georgia	1.4102 (6.04)	1.3572 (10.05)	1.003 (5.52)	1.2067 (15.30)	1.1657 (6.48)	1.1801 (14.31)
Uzbekistan	-0.462 (2.02)	-0.824 (5.37)	-0.232 (1.31)	-0.651 (7.27)	-0.339 (2.03)	-0.666 (7.37)
constant	1.418 (20.52)	1.5706 (30.17)	1.5442 (21.90)	1.6442 (52.30)	1.5628 (24.68)	1.6498 (51.96)
aR^2	0.71	0.9	0.81	0.97	0.85	0.97
F	11.4	37.32	19.53	105.75	20.52	88.96
Standard error of regression	0.213	0.129	0.171	0.07	0.15	0.07
Mean of dependent variable	1.3189	1.3189	1.3189	1.3189	1.3189	1.3189

because of multicollinearity). The most satisfactory result came out of estimation D2, where no multicollinearity distorted the estimated parameters.

From the results of the estimations and the data in table 19.5 we can draw the conclusion that if there had not been a radical structural change in the official economy in Eastern Europe, then the growth of electricity intensity would have been larger, because the economic structure inherited from the past used a large amount of electricity even in the case when production was falling. We can formulate this also the opposite way: had there been more powerful structural changes in the official economy of the CIS countries, the growth of electricity intensity would have been far less in that region than experienced recently. Even within Eastern Europe differences can be seen with respect to structural changes in the official economy. For instance, in Slovakia structural changes, such as cuts in the production of metallurgy and other heavy industries that were triggered by the shifts in demand, combined with the high unemployment were the cause of the fast decline in the country's electricity intensity in 1993–1994. This was the reason for the larger decrease of electricity intensity in Slovakia than in the Czech Republic, and not a smaller and more rapidly decreasing hidden economy, as implied by the estimation of Kaufmann and his coauthors (see table 19.4).

At the end of our econometric analysis we return to the case of Finland to check the applicability of our results to this particular country. Our question is the following: How big is the difference between estimation and reality, if we estimate the change in Finland's electricity intensity with the help of the parameters of equation (19.2) (i.e., with the parameters that were deducted from the analysis of eighteen postsocialist economies)?

During the recession of 1990–1993 in Finland the rate of unemployment increased by 14.3 percentage points and electricity consumption in the industrial branches increased by 3.6 percent, while total electricity consumption increased by 5.7 percent; accordingly, the value of the variable d_i is -2.1 . If we put these numbers into equation (19.2), and use the parameters produced by estimation D2, we get an estimation of 20.2 percent increase in Finland's electricity intensity. The actual increase in electricity intensity was 22.3 percent; the difference between actual and estimated change is a lot smaller than the margin of error of the estimated function. In our equation the growth of electricity intensity in Finland was explained by the various structural changes, rather than by the explosive growth of the hidden economy.

Based on the results of the statistical analysis and the econometric estimations of the third and fourth sections, we believe that our hypothesis has been proved: structural changes take place in the countries with different speed and these changes are

decisive in the development of electricity intensity. We do not exclude that a fast growth of the hidden economy could have influenced the growth of electricity intensity: we just state that total electricity intensity can be explained sufficiently by aggregate data on structural change and these data can not be excluded from such explanation.

5 Residential Electricity and Hidden Economy—Description of a New Method

Based on the results summarized above, I decided to investigate the size of the hidden economy on a level that is less dominated by structural changes. In several studies (Lackó 1995a, 1995b, 1996, 1998) I analyzed the size of hidden activities present in households in a cross-section of countries and used these results for the estimation of the volume of hidden activities on the national level. That model utilized the data of residential electricity consumption, a part of aggregate electricity consumption. In the following I summarize this model. The method helps to establish the *share* of the hidden economy in different countries in a given period: I made calculations for twenty OECD economies and four postsocialist countries. Because of the unavailability of the necessary data I could not attempt to calculate the *growth* of the hidden economy, and in this respect my calculations are not comparable with the Dobozi-Pohl-Kaufmann calculations.

An advantage of the model based on residential electricity consumption is that it relies in meso-level analysis, that is, the investigation of the behavior of households. This way it can ignore the substantial differences in the macrostructure of the different economies, and the impact of shifts in this macrostructure on electricity consumption, and through this, on calculation of the size of the hidden economy.

The use of households for the analysis of the hidden economy is beneficial for another reason. One characteristic feature of the economic transition in Eastern Europe is the mushrooming of small private business, which is set up practically in the family-household framework. In this milieu this fast-growing economic activity can easily operate hidden from state registration and thereby host a substantial part of the hidden economy. Thus my investigation also focuses on one of the basic and classical territories of hidden activities.

The method is based on an econometrically tested model that was developed to measure the size of hidden activities in developed market economies. Subsequently, the estimations of the size of the hidden economy in the socialist-postsocialist countries required some modifications of the model according to the specificities of these systems.

The first premise for the core model is that in each country a part of the household consumption of electricity is used in the hidden economy. We asserted that the electricity consumption of households in a country was determined not only by such visible factors as the size of the population, the level of development, the country's geographical location (climate and weather), the relative price of electricity, and access to other energy sources, but also by the extent of the hidden economy.

In the model the hidden economy is represented by three proxy variables: the tax/GDP ratio, the inactive/active labor ratio, and the ratio of public social welfare expenditures to GDP. The first two proxies represent well-known relationships: the higher these ratios the higher the share of the hidden economy. The third indicator is related to the enforcement of taxes: the higher the third ratio, the stronger efforts are made by the state to collect outstanding taxes.

The parameters of the model were estimated by a cross-section of countries in different variants: (1) for nineteen OECD countries in 1990, (2) for nineteen OECD countries in 1989, and (3) through a panel data base made up of the data for 1989–1990. The estimated parameters were significant, and the signs coincided with the expected ones. Thus the results supported our assumptions about the determinants of household electricity consumption, including the impact of the hidden economy.

After the estimation of the parameters of the model, residential electricity consumption could be decomposed to two parts with the help of which indicators were created for each country showing the per capita household consumption of electricity related to the hidden economy as a share of total per capita household electricity consumption. This calculation was carried out not only for nineteen OECD countries but, following a necessary modification, also for four Eastern European countries (Hungary, Poland, the Czech Republic, and Slovakia). The parameters obtained for the developed market economies were applied for these four post-socialist economies to establish that part of household electricity consumption that was independent of the hidden economy. After subtracting this part of the consumption from the actual amount of electricity consumption we got the share of household electricity consumption used in the hidden economy. As this sequence of calculations shows, in the case of Eastern European economies the proxies characterizing the determinants of the hidden economy in developed market economies were not made use of.

This exercise aimed at eventually determining the contribution of the hidden economy to GDP in the individual countries. However, the share of the hidden economy in the GDP cannot be calculated without the knowledge of how much

GDP is produced by one unit of electricity in the hidden economy in each country. Because these data are unavailable, an indirect conversion method had to be used. This method was rather rudimentary: the results of one of the alternative estimations for the size of the hidden economy known from the literature were taken (a calculation carried out for a single country for the early 1990s), and the other countries' data (their index of hidden economy expressed in terms of residential electricity consumption) were proportioned to this base country. Table 19.9 summarizes the results of our estimations.

According to the results of our investigations, in the early phase of transition the size of the hidden economy in two postsocialist countries, Hungary and Poland, was two times larger than in the average of developed market economies, and much larger than in those developed economies that had the largest hidden part (Spain, Greece, Ireland, Belgium, and Italy). The share of hidden economy in another two

Table 19.9

The shares of the hidden economy in percent of GDP

	1990	1991	1992	1993	1994
Poland	30.8	29.7	33.0	33.6	32.8
Hungary	26.7	32.4	34.8	32.8	31.0
Czech Republic	n.a.	15.2	19.9	14.4	15.4
Slovakia	n.a.	11.2	14.7	15.0	22.3
Spain	22.9				
Greece	21.8				
Ireland	20.6				
Belgium	19.8				
Italy	19.6				
Denmark	16.9				
Austria	15.5				
Australia	15.1				
Germany	14.6				
Portugal	13.8				
Netherlands	13.4				
Finland	13.3				
Japan	13.2				
United Kingdom	13.1				
France	12.3				
Canada	11.7				
Sweden	11.0				
United States	10.5				
Switzerland	10.2				
Norway	9.3				

Source: Lackó (1998).

countries, the Czech Republic and Slovakia, was smaller, but in Slovakia it has increased steadily and fast since transition started.

6 Summary

I have disputed the frequently presented and quoted statement that in postsocialist economies data on total power consumption are better indicators for aggregate output changes than official GDP. The development of electricity consumption allegedly reflects the combined growth of official and hidden economies, and it tells the true story about the decline of output in transition countries between 1989 and 1994.

Dobozi, Pohl, and Kaufmann claim that the statistically registered drop in GDP after 1989 in postsocialist countries, but especially in the successor states of the Soviet Union, exaggerates the actual decline, possibly by a very large margin. They take it for granted that in market economies aggregate economic activity and electric power consumption usually move in lockstep (with an electricity-GDP elasticity close to one). Because the postsocialist countries do not show the same pattern, they claim that this feature can be explained rationally only by a rapid growth of the hidden (unofficial) economy.

In the analysis of the validity of this approach, first the Kaufmann method was applied for Finland, a developed market economy that suffered an output decline in the early 1990s comparable to that in the postsocialist economies. The result was astonishing: it showed a completely unrealistic implicit growth of the hidden economy in Finland and inevitably led to questioning the underlying assumptions of the Dobozi-Pohl-Kaufmann approach.

It was then shown that the variation of electricity intensities in postsocialist countries does not necessarily reflect the growth of the hidden part of the economy. Statistical and econometric analysis of data for eighteen postsocialist economies testify that the measured and registered structural changes are sufficient to explain the differences in the changes of electricity intensity in the region.

In the course of these investigations I have become convinced that using aggregate electricity consumption data and the assumption of constant electricity intensity is not the proper way to calculate the size or growth of the hidden economy either in mature market economies or in economies in transition. One has to use other indices (such as residential electricity consumption) and other assumptions to get more satisfactory estimations of the size of the hidden economy in developed market economies and postsocialist countries.

Notes

The research was carried out during the author's Phare-ACE Fellowship No. P95-2776-F at the International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria.

References

- Alexeev, M., G. Grossman, N. Malyshev, A. Sayer, and V. Trembl (1987). Studies on the Soviet second economy. *Berkeley-Duke Occasional Papers on the Second Economy in the USSR* 11.
- Dobozi, I. (1995). Electricity consumption and output decline—an update. *Transition* 6(9–10): 19–20.
- Dobozi, I., and G. Pohl (1995). Real output declines in transition economies—forget GDP, try power consumption data! *Transition* 6(1–2).
- Ebel, R. E. (1997). *Energy Choices in the Near Abroad: The Have and Have-nots Face the Future*. Washington, D.C.: Center for Strategic and International Studies.
- EBRD (1995). *Transition Report 1995—Investment and Enterprise Development*. London: European Bank for Reconstruction and Development.
- EBRD (1997). *Transition Report 1997—Enterprise Performance and Growth*. London: European Bank for Reconstruction and Development.
- Freeman, C., and L. Soete (1994). *Work for All or Mass Unemployment*. London: Pinter.
- IEA (1994). *Yearbook*. Paris: International Energy Agency.
- Kaufmann, D. (1995a). On Ukraine's Economy: A glimpse at the Past, pausing at the present, peering into the Future. Mimeo, Workshop on Evaluation of Ukraine's GATT/WTO Membership and Alternatives, IIASA, Laxenburg, Austria.
- Kaufmann, D. (1995b). How Large is the Unofficial Economy in Ukraine: An Electrical Approach to Measurement. Mimeo, World Bank.
- Kaufmann, D., and A. Kaliberda (1996). Integrating the unofficial economy into dynamics of post-socialist economies: A framework for analysis and evidence. *Development Discussion Papers* 558, pp. 81–120. Harvard Institute for International Development.
- Kornai, J. (1992a). *The Socialist System: The Political Economy of Communism*. Princeton, N.J.: Princeton University Press.
- Kornai, J. (1992b). The postsocialist transition and the state: reflections in the light of Hungarian fiscal problems. *American Economic Review, Papers and Proceedings*, May, 82(2): 1–20.
- Kornai, J. (1993). The evolution of financial discipline under the postsocialist system. *Kyklos* 46(3): 315–336.
- Kornai, J. (1995). Eliminating the shortage economy: a general analysis and examination of the developments in Hungary. *Economics of Transition* 3(1): 13–37.
- Lackó, M. (1992). The extent of the illegal economy in Hungary between 1970 and 1989—a monetary model. *Acta Oeconomica* 44(1–2): 161–190.
- Lackó, M. (1995a). Rejtett gazdaság nemzetközi összehasonlításban (hidden economy in international comparison). *Közgazdasági Szemle* 52(5): 486–510.
- Lackó, M. (1995b). Hungarian Hidden Economy in International Comparison. Paper presented at the conference Hungary: Toward a Market Economy. Budapest, 20–21 October.
- Lackó, M. (1996). Hidden Economy in East-European Countries in International Comparison, Mimeo, IIASA.

Lackó, M. (1998). The hidden economies of Visegrád countries in international comparison: a household electricity approach. In L. Halpern and Ch. Wyplosz, eds., *Hungary: Towards a Market Economy*. Cambridge: Cambridge University Press.

OECD (1993). *Electricity Policies of Romania*. Paris: Organization for Economic Co-operation and Development.

OECD (1994). *Electricity in European Economies in Transition*. Paris: Organization for Economic Co-operation and Development.

OECD (1995a). *Energy Policies of the Russian Federation—1995 Survey*. Paris: Organization for Economic Co-operation and Development.

OECD (1995b). *Energy Policies of Hungary—1995 Survey*. Paris: Organization for Economic Co-operation and Development.

Roxburgh, I., and J. Shapiro (1996). Russian unemployment and the excess wages tax. *Communist Economies and Economic Transformation* 8(1): 5–28.

Standing, E. G. (1994). Labor market dynamics in Russian industry in 1993. *ILO-CEET Reports, 1994*. Budapest.

UN (1994). *Annual Bulletin for Electric Energy Statistics*. New York: United Nations.

UN (1995). *Annual Bulletin for Electric Energy Statistics*. New York: United Nations.

UN ECE (1993). *Economic Bulletin for Europe*. Geneva: United Nations.

World Bank (1993). *Ukraine Energy Sector Review*. Washington, D.C.: World Bank.

20 Transition to the Market in Asia: The Case of Vietnam

Marie Lavigne

János Kornai opens his analysis of the “Points of Departure,” in his *Socialist System*, with the following lines:

Let me begin describing the book’s subject matter by taking an example: present-day China. . . . Here I shall mention just a few of its attributes:

1. The Communist Party has been in power in China for more than four decades. This has left its mark on all spheres of society, politics, and the economy.
2. China is part of the “Third World.” It is among the “developing” countries, which are poor and backward by comparison with the industrially developed countries.
3. Geographically, China is part of Asia. Numerous typically Asian attributes are therefore displayed in its history, its cultural heritage, its religious and philosophical traditions, and its people’s way of life and relations with one another.
4. Whereas the previous three points concern similarities that China bears to three different groups of countries (other countries under control of Communist parties, other developing countries, and other Asian countries) there are many things in which China is unique and cannot be compared with any other countries.

This book does not take the approach of rejecting generalization and “common regularities” of every kind, or agree that individual features alone exist in every country. It recognizes general influences that apply in similar ways in countries that differ greatly in other respects. On the other hand, the book eschews single-factor explanations of any kind, deeming it necessary to make a multicausal, multifactorial analysis of society. (Kornai 1992, pp. 3–4)

The purpose of this chapter is to apply this methodological framework to the analysis of the transition process in Vietnam. Vietnam shares with China all the features mentioned by Kornai: the Communist Party has been in power there for a longer period than in China; it is a Third World country, poorer than China; it is an Asian country; it has unique features of its own. All these features have influenced Vietnam’s transformation.

The fateful year 1989 saw Vietnam entering the first year of its market-oriented reforms (if one considers the beginning of *doi moi* in 1986 as a failed attempt). At the same time, it was the first year when Vietnam really recovered its unity, its independence, and peace.

Unity was achieved on April 30, 1975, when the U.S. army left Saigon, and became official with the elections of July 1976. But deep political divisions remained between the North and the South; the forced socialization of the South between 1975 and 1978 triggered a huge emigration. More than 1 million people (ethnic

Chinese and also Vietnamese) fled the country, most of those as boat people who landed in the West or in camps in Thailand and elsewhere.

Vietnam had long been a Chinese colony (from the second century B.C. to the tenth century A.D.), then occasionally often lost its independence to Chinese invaders, and in 1859 became a French colony known as Indochina. The country was occupied by Japan at the end of World War II; France fought to regain its colony after 1945. The war with France ended in July 1954, with as a consequence the division of the country into two separate states belonging to two different political and economic systems. The second Vietnam War started soon after between North and South Vietnam, that is, the Democratic Republic of Vietnam (under Communist rule) and the Republic of Vietnam. It intensified after 1964 when the United States decided to support the Republic of Vietnam and ended in 1975. However, the independence of the country remained fragile as China exerted constant pressures at the northern border and over the archipelagos of the Sea of China, and ultimately supported Cambodia against Vietnam.

Peace, achieved in 1975, was again jeopardized when Vietnam occupied Cambodia in 1979. Following lengthy negotiations involving the ASEAN countries, the United States, the Soviet Union and China, as well as France and the United Kingdom, an agreement was reached in 1989 under the aegis of the United Nations, and Vietnamese troops left the country in September. Peace was then consolidated.

At the same time, in 1988–1989, Vietnam launched the second stage of its reform program, of which the first stage was initiated in 1986 under the name of *doi moi* (renovation). Although the first reform experiment clearly belonged to a “market socialism” type, with many inconsistencies reminding of the Soviet *perestroika* implemented at the same time, the second one was later hailed by the international financial organizations as a true transition to the market, albeit in the framework of a Communist political regime. What in the Vietnamese transformation may be attributed to this political context, and what has been determined by the fact that Vietnam is a Third World country, situated in Asia and hence using Asian ways of development? We would look at the Asian factor as really determining, but its impact is also linked with the two other features.

1 Vietnam: A Third World Country in Transition

The development of a country may be measured by indicators of the *level* of development, by data relating to the *structure* of the economy, and by the *policies* conducted to achieve and sustain growth.

1. Vietnam was in 1994 the poorest of all the countries in transition, with a *GNP per capita* of US\$200 (as compared, for instance, with Mongolia at \$300, or Tajikistan, the poorest country among the NIS at \$360, or Albania, the poorest among the Central and Eastern European countries at \$380). In the same year the *GNP per capita* of China was \$560 (World Bank 1996). It ranks as the twelfth country in the world for the *GNP/capita* indicator (ascending order), between Nepal and Bangladesh, while China is the 39th (World Bank 1996). However, its *GDP per capita*, based on purchasing power parity (PPP), was in 1994 \$1208, six times as high as the *GNP* based on the official rate of exchange (UNDP, *Human Development Report 1997*). Some 55 percent of the population lived below the poverty line (based on a daily diet of 2,100 calories plus nonfood essentials) in 1993, down from 75 percent in 1984 (World Bank 1994, p. 68); 90 percent of the poor lived in rural areas.

However, Vietnam fares much better as regards the Human Development Index (HDI) estimated by the UNDP. It ranked 121th (in decreasing order) out of 175 countries in 1994, but according to the PPP-based *GDP per capita* index used by the UNDP it ranked 147th, which means that for the indicators included in the HDI (representing health and education levels) Vietnam is more advanced than its level of *GDP* might suggest (UNDP, *Human Development Report 1997*).

2. Vietnam displays the *structural features* of a developing country, first of all by the share of agriculture in the *GDP* and in employment (table 20.1). Hence the measures taken in 1988–1989 for the liberalization of this sector had immediately a dramatic impact on the overall performance of the economy: not only did the agricultural output increase, but Vietnam also became in a few years the third rice exporter in the world¹ (table 20.2).

3. It follows that development policies focused on agriculture are critical to the future of the country. Here there may be a conflict between policies aimed at increasing growth and those aimed at reducing poverty. Improving the conditions of production in the agricultural sector and at the same time trying to achieve a more balanced regional growth, in particular through irrigation and the development of infrastructures—roads, access to water and to electricity networks—will have a significant impact on poverty (van de Walle 1996); an industrial policy focusing on fast-growing manufacturing branches, and on export-oriented industries might quicken the rate of growth but with a smaller impact on poverty, and a negative impact on the distribution of incomes. Has the Vietnam government clearly expressed its policy choices? It seems caught in uneasy efforts at combining a regional policy focusing on the development of poorer rural areas, a strategy of industrialization and import-substitution belonging to the socialist legacy, and an export-led growth strategy modeled on the Southeast Asian example.

Table 20.1

GDP and employment by economic sector and form of ownership

A. GDP by economic sector and form of ownership (percent)

	1990	1991	1992	1993	1994	1995	1996
GDP	100	100	100	100	100	100	100
share of the state in total GDP		32.5	33.3	36.2	39.2	40.2	42.2
<i>Agriculture, forestry, fisheries</i>	37.5	40.5	33.9	29.9	28.7	27.5	32.5
share of the state	1.0	1.1	0.9	0.7	0.7	0.8	
share of state ownership in the overall output of agriculture	2.9	2.9	2.8	2.8	2.9	2.8	
<i>Industry and construction</i>	23.9	23.8	27.3	28.9	29.6	30.1	28.9
share of the state	14.4	15.2	17.7	19.1	19.4	19.9	
share of state ownership in the overall industrial output	62.8	63.5	64.7	65.7	65.2	66.1	
<i>Services</i>	38.6	35.7	38.8	41.2	41.6	42.3	38.6
share of the state	17.1	17.0	17.6	19.3	20.0	21.5	
share of state ownership in the output of services	44.2	47.5	45.5	46.8	48.1	51.0	

B. Employment by sector and form of ownership (percent)

	1990	1991	1992	1993	1994	1995	1996
Total employment	100	100	100	100	100	100	100
share of the state	11.3	10.2	9.4	9.1	8.7	8.6	8.9
<i>Agriculture, forestry, fisheries</i>	72.3	73.2	72.9	73.0	73.0	72.6	
share of the state	1.3	1.2	1.1	1.1	1.0	0.9	
share of the state sector in agricultural employment	1.9	1.7	1.5	1.4	1.3	1.2	
<i>Industry and construction</i>	13.9	13.6	13.4	13.4	13.3	13.5	
share of the state	3.9	3.2	3.0	3.1	3.0	3.0	
share of the state sector in overall industrial employment	28.3	23.9	22.9	22.8	22.7	22.1	
<i>Services</i>	13.8	13.2	13.7	13.6	13.7	13.8	
share of the state	6.1	5.8	5.3	4.9	4.7	4.8	
share of the state sector in overall employment in services	43.1	41.0	37.3	36.7	35.3	34.8	

Source: General Statistical Office (*Statistical Yearbook*, Hanoi, 1995 and 1996). For 1996, World Bank (1997).

Table 20.2

Rice (paddy) output, export and domestic availability

Rice (in 1,000t.)	1988	1989	1990	1991	1992	1993	1994	1995
Output	17,000	18,996	19,225	19,622	21,590	22,836	23,528	24,964
Rate of growth		11.70%	1.2%	2.06%	10.02%	5.77%	3.03%	6.1%
Official export	91	1,420	1,624	1,033	1,946	1,722	1,950	2,150
Informal export (estimates)	200	200	200	200	200	200	500	600
Domestic availability	16,709	17,375	17,401	18,389	19,444	20,914	21,078	22,214
Availability in kg/cap	255	260	263	271	280	294	291	300

Source: *Statistical Yearbook, 1995*; the estimates of informal exports are taken from press reports.

2 Vietnam: A Socialist Country in Transition to the Market

According to the criteria retained by János Kornai to define a socialist system, the political one comes first. It is the political and ideological monopoly of the Communist party. "The key to an understanding of the socialist system is the *structure of power*, which receives little or no attention in many comparative studies of economic systems. In my opinion, the characteristics of the power structure are precisely the source from which the chief regularities of the system can be deduced" (Kornai 1992, p. 33). In Vietnam the Communist Party solidly keeps the reins of power, as the 8th Congress, held in June 1996, has shown. Though the party members made up 2.7 percent of the population in 1986 (a rather low figure compared with Eastern Europe and the Soviet Union, where the percentage ranged from 6.6 in the case of the USSR to 15.6 for Romania, and even with China, where the percentage was 4.2), the influence of the party is very important, especially as many cadres still belong to the war generation that fought for independence and reunification, despite the rejuvenation of the leaders in September 1997.

The party leaders were able to promote a successful transition to the market—not without benefiting from it² for corruption and rent-seeking are widely spread among the party officials. Is really Vietnam a case for the compatibility of capitalism and socialism? János Kornai has argued, in the case of China, that capitalism and socialism were incompatible in the long run but that a market economy could coexist for some time with a Communist political system.³ In China as well as in Vietnam, the party leadership has decided to go on with a market-oriented reform combined with a socialist ideology. The outcome is a mixed economy, but certainly not market socialism if one defines this concept according to the Lange-Lerner theoretical approach of the 1930s.

Was Vietnam more “radical” than China in its approach to the reforms? Both the World Bank (World Bank 1996, p. 21) and the IMF (IMF 1996a, p. 4) have praised Vietnam for having more quickly introduced a more radical reform than China. “Vietnam’s transition is not complete. Industrial production remains concentrated in state enterprises, and administrative controls remain pervasive. But its liberalization and stabilization measures were closer to those of Eastern Europe than they were to those of China. *Not all East Asian reforms have been phased or gradual*” (my emphasis) (World Bank 1996, p. 21). However, many authors speak of gradualism when discussing Vietnam’s approach (as expressed in the titles of the papers by Drabek 1990 and Truong and Gates 1996). The issue is not clear-cut. It is discussed at length by Marcus Diehl (1994), who notes that the stabilization-cum-liberalization was not totally radical in as much as fiscal discipline was not immediately enforced. However, Vietnam did not use “heterodox” stabilization measures such as price freezes and a fixed exchange rate (the last point is strongly stressed by the IMF as well). Ultimately the program worked because the government policy seemed to have a high credibility (Diehl 1994, pp. 17–18). The same ambiguity is to be found in Adams and Krkoska (1996): “The reforms in China and Vietnam may be referred to as partial or sequential, but they occurred relatively rapidly” (p. 135). However, since 1997 international organizations and experts rejoin in a much harsher assessment of Vietnam’s reforms: the state sector and the financial system are increasingly inefficient, and trade liberalization has yet to be completed. The younger members of the new political leadership are indeed prone to accelerate the reforms while the old guard is afraid that such reforms would increase inequality and spread social unrest.⁴

What is still “socialist” in the Vietnamese economy? We will again use Kornai’s framework of analysis, beginning with the property relations and then examining the coordination mechanisms. Speaking in the language of transition economics, we will first deal with the structural transformation, and then with liberalization cum stabilization.*

Property Relations

As in standard socialist economies, the state sector was dominant in industry and services, and collective ownership was dominant in agriculture, where cooperatives accounted for over 80 percent of output in 1988; overall the “socialist” sector accounted in 1988 for about 75 percent of production. In addition, there was a significant parallel or shadow sector, which had developed for various reasons: it went together with the war economy during the Vietnam war; it expressed the resistance of the South to socialization after 1975; it was tolerated by the authorities as a way

of alleviating the shortages when the first attempts to reform the economy were made in the early 1980s. Since the beginning of *doi moi* in 1986, the aim was to develop a "multisector commodity-producing economy."

The reforms addressed the issue of the property rights from the outset, beginning with *agriculture*. This was prompted by urgency, for the growth in agricultural production had been decelerating since 1985, with an absolute decrease in 1987 due to adverse climatic circumstances. The main change in property rights was introduced by Decree 10 of April 1988, which legalized household farming. The Land Law of 1993 retains the constitutional principle that land belongs to the people under the state's control, but provides for extended land-use rights that are allocated to families on a long-term basis (ranging from twenty to fifty years) and are transferable by inheritance. This reform, along with the liberalizing of prices, brought about an immediate supply response as early as 1989 (see table 20.2 for the output of rice). By comparison with the European economies in transition, two major differences have to be stressed. First, the reform was successful despite the fact that formal ownership of land was not granted to peasants; in this sense the land reform is closer to the Chinese pattern of household responsibility. Cooperatives were dismantled as compulsory production units; they evolved into organizations providing supplies to the peasants, performing various construction works, organizing new forms of production, and creating small village enterprises. New cooperatives emerged on a voluntary base. Second, there does not seem to have been as strong a rural exodus as was witnessed in Eastern Europe, though in 1988 the limitations to migration were abolished and many inhabitants from rural areas tried to find work in the cities.

The changes in the *state sector in industry and services* were much more paradoxical. First, the state sector was not dismantled. Its share increased slightly instead (see table 20.1). There was no change in the property rights. What was achieved was more like a "perestroika-type" reform of the kind that is deemed to remain unsuccessful, according to Kornai. Beginning in 1989 state enterprises were granted *more autonomy* in decision-making; they could fix prices, choose their suppliers and customers, determine their output and their investments, and use their profits for self-financing.

Moreover, they could reduce their labor force, and contrary to what has always happened in Eastern Europe in the framework of the pre-1989 reforms they made use of this opportunity. What made them behave in this way? The reasons were twofold. First, the Vietnamese workers benefited from much less social protection than their East European counterparts. Second, the dismissed workers, amounting to slightly less than 1 million people between 1988 and 1993—that is, about 30 percent of the labor force in the state-owned industrial sector—seem to have found

jobs in the new private sector, which expanded by more than 5 million during the same period (Diehl 1996, p. 8). Labor statistics do not allow for precise estimates. One might say, as far as the labor market is concerned, that Vietnam is a specific combination of a socialist-like attachment to full employment and of the labor flexibility characteristic of the East Asian developing economies. Officially unemployment is small (slightly under 6 percent in 1995 of the total labor force, according to the 1995 *Statistical Yearbook*), but in Vietnamese literature one can find estimates of underemployment expressing the gap between the theoretically fully worked time and the actual time of work; with such estimates real unemployment would reach up to 30 percent (Nguyen van Than et al. 1995). Diehl (1996), quoting a Vietnamese source, estimates the share of unemployed at 17 percent in 1993 including the underemployed. However, people adjust to such a situation. They combine different short-time hired jobs (which is expressed in a lively labor market for people coming to the cities from rural areas for temporary jobs). Conversely, unemployed people in the cities go back to their rural bases to help on farms or start businesses. Despite an outflow of workers from rural areas to the cities, the labor force in agriculture and forestry increased by more than 2.5 million persons between 1988 and 1993 (about 10 percent of the labor force of 1993).

Adjustments in the state sector thus express the *hardening of the budget constraint* imposed on the state enterprises. First, they were deprived of subsidies and became on the contrary main providers of budget revenues (table 20.3). However there remain loss-making enterprises (Andersen 1995, p. 16, puts them at 11 percent of the state enterprises) that are bailed out by the state. The number of these enterprises increases over time. The Finance Ministry admitted in 1998 that half of the six thousand state enterprises were unprofitable.⁵ In addition, the contributions of the state enterprises to the budget have been declining in share since 1995. Second, the state enterprises faced high real interest rates (table 20.4) and restrictions in the availability of credit, the banks having been allowed to extend credit only to profitable enterprises. Actually it seems that the banks have in fact extended bad loans (Grant 1996), and the number of bank scandals and defaults increased dramatically in 1997 (World Bank 1997, p. 51). Third, there has been a substantial amount of enterprises shutdowns, however not in the form of bankruptcies. The number of state-owned enterprises dropped from twelve thousand in 1988 to fewer than six thousand in 1997. Most that disappeared were local enterprises ("out of 4,584 non-profitable enterprises in 1989, 4,083 were local," Ton 1995, p. 257). In fact most of them were absorbed by larger central enterprises. The latter were also committed to take up the debts, which explains why the interenterprise arrears declined after 1992.

Table 20.3

The state budget (in percent of GDP)

	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	rev. 1996
<i>Current expenditure</i>											
Wages	13.2	12.7	14.2	18.3	17.7	14.4	16.9	21.5	20.9	19.7	18.3
Subsidies	0.9	1.0	1.7	4.6	4.2	2.6	...	6.5	8.0	6.5	...
Interests (scheduled)	2.9	4.9	5.3	0.0	—	—	—	0.1	0.0	0.1	—
Other	0.2	0.1	0.2	2.9	3.0	3.0	2.9	2.7	2.6	1.9	1.6
Capital expenditure	9.2	6.8	7.0	10.8	10.5	8.8	...	12.2	10.1	11.2	...
<i>Revenues</i>	5.9	3.9	4.4	5.8	5.1	2.6	5.8	7.0	6.9	5.4	6.0
Taxes and duties other than on state enterprises	3.2	12.2	11.3	13.8	14.8	13.5	18.3	22.5	24.7	23.9	23.6
Transfers from state enterprises	2.9	2.2	2.9	3.9	4.1	3.7	5.0	8.3	9.9	10.9	10.7
Other revenues	9.5	9.2	7.2	8.0	8.7	8.1	10.8	11.2	12.1	9.8	9.7
Deficit (on accrual basis)	0.8	0.9	1.2	2.0	2.0	1.4	2.5	2.2	2.1	2.9	2.5
	5.9	4.4	7.2	10.3	8.0	3.7	4.4	6.0	3.1	1.2	0.7
<i>Financing</i>											
Foreign grants and loans	2.2	1.4	2.4	1.5	3.0	1.0	2.4	2.7	0.1	-0.7	0.0
Domestic state bank loans	3.6	2.9	2.9	6.8	2.0	0.9	1.5	1.7	0.5	0.0	0.0
State securities	0.0	0.1	0.1	0.9	1.3	1.8	0.5	1.8	0.9	1.2	0.2

Source: World Bank (1994); IMF, *Vietnam*, Statistical Tables, Washington, D.C.: 1995; World Bank (1997).

Table 20.4

Inflation, interest rates, exchange rate (1987-1995)

	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996
Annual inflation rate (percent)*	301	311	96.0	67.1	61.5	17.5	5.2	14.4	12.7	4.5
Exchange rate** (dong for 1\$)	225	900	4,300	5,000	12,500	10,530	10,510	11,000	11,000	11,175
Monthly interest rate (percent)***		3.3	11.0	3.5	3.0	2.7	2.2	2.1	2.1	1.4
Monthly inflation rate (percent)		8.9	9.1	4.4	4.4	1.4	0.4	1.1	0.9	0.3

Source: *Statistical Yearbook*; IMF (1996a); World Bank (1997); press reports.

* End-year basis.

** The exchange rate regime of Vietnam is a managed float. In the beginning of 1997, the government widened the band of fluctuation of the dong around the central rate to +/- five percent. In October 1997, the band was widened to +/- 10 percent. A new central rate of 11,800 dong for \$1 was introduced in February 1998 and the dong immediately slid to its all-time lowest value of 12,980 dong for \$1. However the dong remains overvalued at its real effective rate, especially against the currencies of the region.

*** Lending interest rate for industrial enterprises.

What is going to be the next step? In principle a consistent transition would imply *privatization*. However, this is not really on the agenda. Moreover, although the World Bank is pressing for privatization (World Bank 1996, pp. 21, 50, 143), the IMF report on Vietnam praises the country for its prudent approach to privatization. First, it is stated, a “crucial factor in hardening budget constraints was that the Vietnamese authorities have also enforced their ‘ownership’ tights. In some countries a ‘lack of ownership’ or accountability emerged during the transition to a market-based system, leading to widespread asset stripping and to unsustainable bonus payments and other abuses. In Vietnam, few such doubts about ownership have arisen” (IMF 1996a, p. 15). Second, the IMF report asks the question, should one press for privatization when state-owned enterprises provide most of the budget revenue? “A crucial point that needs to be factored into the decision about how much and how fast to privatize is that government revenue continues to depend heavily on the performance of SOEs. . . . Care needs to be taken that tax collection is not harmed by reform efforts” (World Bank 1996a, p. 15).

Meanwhile, the state-owned sector is indeed evolving. A first trend is *concentration* of SOEs, in sharp contrast with the demonopolization policies advocated in other countries in transition. The three hundred biggest state enterprises account for three-quarters of state-owned assets and of the state sector’s contribution to the state budget (Truong and Gates 1996, p. 86). According to Government Decree 91 of 1994 and the Law on State Enterprises of April 1995, the existing state-owned enterprises in strategic sectors should be regrouped into a dozen “general corporations” on a sectoral basis (steel, oil, coal, electricity, cement, construction chemicals, tobacco, paper, rubber, coffee, textile and garment industries, post and telecommunications, and civil aviation). In 1997 there were eighteen such corporations; once the process completed there should be no more than ninety to one hundred such groups. Some of them are monopolies (such as Petro Vietnam or Vietnam Airlines). These consortia may integrate vertically their suppliers and their research and development or marketing activities, and they may create their own banks or financial agencies. Each such corporation has to integrate at least five enterprises and have a capital of at least 1 trillion dong (a little less than \$90 million). Is it, as Gates (1996, p. 39) contends, a model inspired by the Korean *chaebols* or the Japanese *keiretsu*? Freeman (1996, p. 392) quotes an observer likening these entities to the big state corporations in Italy, which seems more relevant as the Vietnamese corporations are vertically integrated in a given industry instead of being a conglomerate type. In any case, neither the World Bank (Freeman 1996, p. 392) nor the IMF (IMF 1996a, p. 15: “recent action to form conglomerates in different sectors may not be consistent with increasing efficiency and competitiveness”) approve of this trend.

Although privatization is not yet on the agenda on a large scale, the *equitization* of state enterprises is to be implemented. The five-year plan of 1996–2000 explicitly states that “equitization is not synonymous with privatization” on the grounds that money earned from the sale of state equities “should be reinvested in expanding production and business, ensuring a constant growth of state-owned assets” (Orientations 1996, p. 78). Up to 1998, on the basis of an experimental program launched in 1992, twenty-one enterprises have been equitized. A government decree of May 7, 1996, provides for an expansion of the program on a larger scale. The process is twofold. First, the selected enterprises are to be turned into shareholding companies. Second, they will be able to sell part of the shares, mainly to their own wage-earners. The enterprises to be equitized are mostly small and medium-sized, and it may thus affect about two-thirds of the existing state enterprises. New prospects, however, emerged following the political appointments of 1997. From 150 to 200 companies should have been equitized in 1998, along with changes in legislation that would quicken the process.⁶

A stock exchange might be created by the end of 1998.⁷ The selected companies allowed on the trading floor, picked among those already equitized, might then sell shares (either directly, or in the form of bonds convertible into shares) or issue corporate bonds.

All these trends have turned Vietnam to a kind of state capitalism rather than to market socialism. The production sector was initially able to respond to market signals more quickly and efficiently than in Eastern Europe. But the lack of completion of structural reforms, which was of less concern at the time of fast growth, is now taking its toll as the economy is heading for a sharp deceleration in the rate of growth.

Coordination Mechanisms

The success of stabilization in Vietnam can be summarized in two points. First, it achieved macroeconomic stability in a few years. Inflation was curbed from a three-digit rate (400 percent) in 1988 to less than 5 percent in 1996 (table 20.4). The budget deficit fell from 10.3 percent of the GDP in 1989 to slightly over 1 percent in 1996, with government revenues rising from 14.7 percent of GDP in 1990 to around 24 percent in 1995 (see table 20.3). The rate of exchange of the dong was stabilized following a series of devaluations and remained stable during the years 1992–1997, on the basis of a floating rate regime (table 20.4). All this was obtained through a mix of standard stabilization-cum-liberalization policies:

- liberalizing most prices in 1989 (short of utilities, rents, and medicines);
- liberalizing domestic and foreign trade;
- introducing positive real interest rates in 1989, and lowering the nominal interest rates as soon as the rate of inflation began to decline;
- using a market-oriented exchange rate regime by first unifying the official rate with the rate on the parallel market and then letting the exchange rate float freely;
- balancing the budget with a cut in subsidies to the state enterprises, while the latter increasingly contributed to the budget.

However, Vietnam did not suffer from what János Kornai calls a “transformational recession” that, according to him, hit all the postcommunist countries. His explanation of this specific type of recession is that there has been a brutal shift from a resource-constrained sellers’ economy to a demand-constrained buyers’ economy, and that the economic agents have not been able to shift quickly from the first to the second and consequently have not been able to respond to market signals. Nothing of the sort happened in Vietnam. Agricultural supply responded first to liberalization measures, turning Vietnam into an exporter in a few years. Then the services and industrial sectors resumed growth. Demand remained sustained despite the austerity measures. The perverse effects of increased real rates did not occur, in particular there was no dramatic buildup of interenterprise arrears. As a result, Vietnam displayed a sustained growth, with an annual rate reaching over 9 percent in 1995 and 1996, and still about 8 percent in 1997, from a 5.1 percent increase in 1990. World Bank estimates forecast a rate of growth of only 5 percent in 1998, as a result of the Asian crisis but also of the lags in reforms.

What are the explanations of the Vietnam “miracle” in the early 1990s? Several have been offered:

- Growth was sustained by the fact that oil production from fields developed with the Soviet aid just came onstream (IMF 1996a, pp. 3 and 6).
- The collapse of the CMEA boosted the production of goods until then imported from CMEA countries, such as steel and fertilizers (IMF 1996a, p. 7); Vietnam was less integrated within the CMEA than were Eastern European countries (IMF 1996b, p. 1).
- The labor market was flexible and the private sector absorbed the redundant workers of the public sector (IMF 1996a, p. 9).
- New export markets could easily be found among Vietnam’s growing neighbors, while Eastern Europe had to face the Western European recession in 1991–1993 (IMF 1996a, p. 9);

- State investment had been low in the years before the transition and could easily be boosted by the inflow of foreign capital (IMF 1996a, p. 9); to this one could object that the increase in capital flows was gradual and began soaring in 1992–1993. Vietnam was also left without foreign assistance in the first years of the transition, and began receiving assistance on a significant scale only after 1993.

Most of these reasons do not hold true any longer. Export markets have been dwindling because of the recession in Asia, but also because of a lack of competitiveness of the state-owned sector and because of an overvalued currency (though the dong was devalued in the wake of the Asian crisis; see table 20.4). Foreign investment is declining; the total value of the contracts fell by 40 percent in 1997, because of the Asian crisis—for 70 percent of foreign investment comes from Asia—and because of more long-lasting reasons such as fuzzy legislation and the extent of corruption.⁸

3 Vietnam: An East Asian Country

There has been a large debate about Asian culture and Asian values as factors of development. It has been said that the Asian traditions of priority to the group over the individual, of personal control over one's desires, of discipline, self-restraint, and dedication to work, added to the values of Confucianism (Chowdury and Islam 1993), could explain many features of the East Asian development, including the high productivity and intensity of labor and the high level of savings, as well as the specific kind of social protection, that was ensured in the framework of the local community or the family-at-large rather than by welfare institutions. Other authors (for example, Funabashi 1993) have argued that "a new breed of cross-fertilized civilization" combines the cultures of China, Indochina, Japan, and the United States. A sharp controversy was fueled by an article by Paul Krugman (1994) on the East Asian miracle, which he dubbed a "myth". According to Krugman the new industrializing East Asian economies just duplicated the Soviet model of mobilization of resources "that would have made Stalin proud" (Krugman 1994, p. 70), through increase in investment, improvement of human capital, and involvement of most of the population able to work into employment. An earlier report by the World Bank (1993) insists on the role of the market, adding that the clue to East Asian development is not, as what the report calls the "revisionist view" states, a "state-led" model, but rather a "market-friendly" policy where the governments basically let the market operate but made selective interventions so as "to ensure adequate investment in people, provide a competitive climate for private enterprise, keep the

economy open to international trade, and maintain a stable macro economy" (World Bank 1993, p. 10).

How does all this apply to Vietnam? Since the beginning of the transition the country displays some of the characteristics of the South-East Asian developing countries: a commitment to the market, development policies conducted by the State, and an international openness that was acknowledged by its admission in 1995 to the ASEAN. However, there are features proper to Vietnam, resulting from history and geopolitics. Could one say that the first feature—its Communist past—substantiates Krugman's analogy between Stalinist and East Asian development? Not quite. Vietnam had not applied the forced industrialization model and has always privileged the role of agriculture, which during the war had not only an economic but also a political and strategic dimension as the resistance to the enemy could be successful only when based on a close link between the party and the peasants. Soviet assistance was essentially military aid during the war. Following the end of the war and the reunification that aid was indeed directed toward the development of heavy industry (mainly the power industry and oil exploration), but it also indirectly strengthened the orientation toward agriculture and light industry because the Soviet Union wanted the debt to be repaid in food products and manufactured goods (footwear, textiles, garments). Also, Vietnam never achieved the rates of investment and domestic savings (relatively to the GDP) that have been typical both for the Soviet-type economies and the East Asian ones. Again, war conditions and the low levels of incomes after the war may explain this feature. One may also add that following the reunification, the South—which was richer and could have contributed to the investment mobilization—resisted what was perceived as sheer confiscation (through emigration and concealment of resources) until the beginning of the transition.

The combination between the market orientation and the role of the state in defining selective fields of intervention is also something specific to Vietnam in its implementation. We have seen earlier that the market orientation is associated with property structures still privileging the role of state-owned enterprises. However, the liberalization goes on. A significant move was the announcement in January 1997 that the private companies would be allowed to sell rice, a relaxation of the state monopoly on rice exports, which was the most significant export monopoly left. As for state intervention, Vietnam applies the same methods as other Asian countries did at the beginning of their industrialization. It applies an industrial policy (both sectoral and regional); the main instrument is the control on foreign investment, which is channeled toward export-oriented, technologically advanced industries, and also toward labor-intensive industries using local resources. Foreign investors are

also encouraged—with less success—to invest in the central and northern parts of the country, so as to create a better balance between the south and the north. Vietnam uses a commercial policy to protect its infant industries, though it is committed to a lowering of import duties beginning in 2003 in the framework of AFTA (Asian Free Trade Association). It has a five-year plan that can be likened more to the strategic development planning practiced in other Southeast Asian countries than to Soviet-type mandatory plans. But the role and the perception of the state remain different. The Southeast Asian countries have authoritarian (not to say dictatorial) governments in a capitalist environment characterized by wide income inequalities and a collusion between the leading companies and the government. In Vietnam, the state is still very much subordinated to the party. In the new “capitalist” world the political leadership has an influence built on family links between party bureaucrats and managers. At the same time, the deep intermingling of the party cadres and the rural basis was in the past the only way to win the war. The strongholds of the conservative fraction of the party in the rural areas still build on this past.

Finally, one has to mention the Chinese factor, which is very complex. Vietnam has been a Chinese colony at different moments of its history and for many centuries. The cultural and economic influence of China, both from within—through the ethnic Chinese, or *Hoa*—and from outside, is very important, notwithstanding the numerous conflicts, all of which are not yet settled, between the two countries, such as the impact of the Sino-Soviet dispute, or of the occupation of Cambodia by the Vietnamese military forces, or of the territorial disputes over the Spratleys. Vietnam (and more generally the three Indochinese countries) have close informal links with China. Trade flows are largely underrecorded and may be defined as “officially acknowledged smuggling” in interprovincial relations (de Vienne 1995). If one takes into account the fact that a significant share of foreign investment originating from Taiwan, Hong Kong, or Singapore is really Chinese investment, China would be the sixth-largest investor in Vietnam (instead of the twenty-fifth as officially recorded). Thus the similarity between the two countries goes far beyond the fact that they are still Communist countries, or beyond some common features of their economic reform.

Conclusion

The picture of the Vietnam economy in 1998 looks far bleaker than it did three years earlier. It would be too easy to attribute it to the impact of the crisis in Southeast

Asia. Because of its poverty and low level of development, Vietnam remains isolated from external disturbances that propagate through developed networks linking the financial markets. It is affected through its imports and exports because the reforms of the early 1990s resulted in an opening up of the economy. But the main cause of the looming recession is not to be found in the fallout of the Asian turmoil. It is rooted in the delays in structural reforms, and in a political system that is now plagued by struggles among factions defending their vested interests. Although most observers believe that Vietnam will be able to muddle through in the short and even the medium term, the long-term outcome might well prove the relevance of Kornai's statement about the incompatibility between capitalism and socialism.

Notes

The present study builds on a paper prepared jointly with Gérard de Bernis, "Comparison of the industrial strategies and of the types of financing growth between Vietnam and some APEC countries," presented at an international symposium on Growth Strategies and Emerging Markets, Ho Chi Minh City, November 12–13, 1996, organized by the Caisse des Dépôts et Consignations, the Development Center of the OECD, the Centre d'Economie et de Finances Internationales of the University of Aix-Marseille II, and the *Revue Economique*.

1. It became the second in 1996, with 3.05 million tons exported (*Financial Times*, January 8, 1997), and 3.6 million tons in 1997 (*Financial Times*, February 5, 1998).
2. Jeremy Grant, "Hanoi edges closer to tackling graft," *Financial Times* (October 22, 1996); Michael Zielensiger, "Potemkin capitalism in Vietnam after a hopeful start," *International Herald Tribune* (April 3, 1998).
3. Interview with Erik Izraelewicz for *Le Monde*, July 1996.
4. Jeremy Grant, "Hanoi admits rift on reforms," *Financial Times* (January 16, 1998).
5. *Agence-France Presse* (February 22, 1998), quoted in the *International Herald Tribune* (February 23, 1998).
6. *Agence-France Presse* (February 22, 1998).
7. Jeremy Grant, "Vietnam plans trial stock exchange," *Financial Times* (February 11, 1998).
8. Jeremy Grant, "Asian whirlwind could soon prove Vietnam's problem too," *Financial Times* (February 11, 1998).

References

- Adams, F. Gerard, and Libor Krkoska (1996). Lessons of the East Asian transition experience for transformation in Eastern Europe—China and Vietnam vs. the Czech Republic, Hungary, Poland and Russia. *Moct-Most* 6(4): 123–141.
- Andersen, Henny (1995). *Vietnam, Sustainable Growth and the Issue of Capital*. Stockholm School of Economics.
- Chowdury, A., and I. Islam (1993). *The Newly Industrializing Economies of East Asia*. New York: Routledge.

- De Vienne, Marie-Sybille (1995). La Chine et les trois Etats indochinois: économie et politique d'une 'normalisation', 1985-1995. *Le Courrier des pays de l'Est* 399: 44-49.
- Diehl, Markus (1994). Stabilisation in Reforming Socialist Economies: The Case of Vietnam. Paper presented at the 7th World Congress of Social Economics, Verona, August 3-6, 1994. Revised January 1996.
- Drabek, Zdenek (1990). *A Case Study of a Gradual Approach to Economic Reform: The Viet Nam Experience of 1985-88*. Washington, D.C.: The World Bank.
- Freeman, Nick J. (1996). Vietnam: better managing reforms. In *Southeast Asian Affairs 1996*, pp. 384-402. Singapore: ISEAS.
- Funabashi, Y. (1993). The Asianization of Asia. *Foreign Affairs*.
- Gates, Carolyn L. (1995). Enterprise Reform and Vietnam's Transformation to a Market-Oriented Economy. *Asian Economic Bulletin* 12(1): 29-52.
- Gates, Carolyn L. (1996). *Enterprise Adjustment and Economic Transformation: Industrial Enterprise Behaviour and Performance in Vietnam during Stabilization and Liberalization*. ISEAS Working Papers, Institute of Southeast Asian Studies.
- Grant, Jeremy (1996). Doubts raised over health of Vietnamese banking. *Financial Times* (August 21).
- IMF (1996a). *Vietnam, Transition to a Market Economy*. Washington, D.C.: IMF.
- IMF (1996b). *Macroeconomic Experiences of the Transition Economies in Indochina*. Washington, D.C.: IMF.
- Kornai, János (1992). *The Socialist System: The Political Economy of Communism*. Oxford Clarendon Press.
- Krugman, Paul (1994). The end of Asia's miracle. *Foreign Affairs* 73(6): 62-78.
- Nguyen Van Than, and Tran Thi Tuyet Mai. (1995). Employment in rural Areas. *Vietnam's Socio-Economic Development* 2: 39-53.
- Orientations (1996). Orientations and tasks of the five-year Plan (1996-2000). *Vietnam's Socio-Economic Development* 7: 56-80.
- Ton That Nguye (1995). Vietnam: les potentialités du devenir. *Transitions* 36(1-2): 237-268.
- Truong Dinh Hung, and Carolyn L. Gates (1994). Foreign direct investment in Viet Nam: its role and prospects. *Vietnam Social Sciences* 2/40: 10-29.
- Truong, David H. D., and Carolyn L. Gates (1996). Vietnam's Gradualistic Economic Reforms. In Alex E. Fernández Jilberto and André Mommen, eds., *Liberalization in the Developing World: Institutional and Economic Changes in Latin America, Africa and Asia*, pp. 72-95. London and New York: Routledge.
- Van de Walle, D. (1996). *Infrastructure and Poverty in Vietnam*, Washington, D.C.: World Bank.
- World Bank (1993). *The East Asian Miracle: Economic Growth and Public Policy*. Oxford: Oxford University Press.
- World Bank (1994). *Vietnam Public Sector Management and Private Sector Incentives: An Economic Report*. Washington, D.C.: World Bank.
- World Bank (1996). *From Plan to Market (World Development Report 1996)*. Oxford: Oxford University Press.
- World Bank (1997). *Vietnam: Deepening Reform for Growth*. Washington, D.C.: World Bank.

21 The Debate on Understanding China's Economic Performance

Jeffrey D. Sachs and Wing Thyee Woo

1 Introduction

China's rapid growth since 1978 presents various paradoxes that have become the subject of heated debate within the economics profession. Why is it that a country that espouses socialist practices is among the fastest growing countries in the world, when virtually all other socialist economies have collapsed? Broadly speaking, two schools of thought have emerged to interpret the Chinese experience. One school of thought gives great credit to the *evolutionary*, *experimental*, and *incremental* nature of China's reforms. In this view, China has been groping, with considerable success, toward a unique Chinese economic model. A faster approach to reforms, according to the experimentalist school, would have led to more social conflict, instability, and poorer economic policies (because of less experimentation).¹ The other school of thought holds that Chinese institutions are in fact gradually converging with those of nonsocialist market economies, especially those in East Asia. In this view, the faster the convergence, the better will be the outcomes. The convergence school, in which we include ourselves, puts great stress on China's *economic structure* at the start of reforms as a major explanation of rapid growth and of why gradualism was not a barrier to growth.²

In this debate, János Kornai's thinking is invaluable, even though Kornai himself has not focused extensively on China in his own investigations.³ Kornai considered in more detail than any other observer the practical difficulties of the middle ground between capitalist institutions and central planning. He, more than anyone else, debunked the ersatz "market socialism" that came crashing down in Eastern Europe after the revolutions of 1989. We do not seek to put words in Kornai's mouth—his eloquence speaks for itself—but we suspect that he would be deeply suspicious of claims of China's evolution to a successful market socialist model. Speaking for ourselves, we think that China's early successes have been misunderstood by those who see them as a triumph for a long-term Chinese middle way. Moreover, we believe that the next phase of China's reforms is likely to be far trickier than in the past two decades. We do not believe that convergence in institutions between China and the capitalist West is assured. We do, however, suspect that if the process of convergence were to stall or reverse, the economic performance of China would seriously deteriorate.

In our interpretation, the key to China's growth, compared say with the Soviet Union, is that China began its high growth period with a high population density

heavily concentrated in low-wage agriculture, and with a substantial proportion of the population along the coastline (and therefore available for manufacturing export-oriented production). These were the very conditions that were favorable for labor-intensive, export-led growth in other parts of nonsocialist developing East Asia, which began their transition to market-based, export-led growth a decade or two earlier than China. Compared with the Soviet Union (which began the post-communist era with a highly urbanized population heavily concentrated in moribund state enterprise, and typically far from seaports and major Western markets), China faced a much simpler problem of dismantling, or circumventing, the stifling effects of socialist institutions. Moreover, in our view, gradualism has not been a strategy so much as a result of continuing political conflict and other difficulties inherent in setting a policy course in a country of some 1.2 billion people.

The experimentalist school proposes four basic economic propositions concerning China's reforms:

- (E.1) Gradualism in market reforms has been key to China's rapid growth.
- (E.2) China's gradualist strategy is transferable to other economies in transition from central planning.
- (E.3) China's experiments in noncapitalist institutions are proving to be successful in agriculture; township and village enterprises (TVEs) in the rural areas; and state-owned enterprises (SOEs).
- (E.4) China is evolving toward a unique set of economic institutions, as a result of experimentalism in policy design.

The convergence school counters with four alternative theses:

- (H.1) China's rapid growth has come *despite* gradualism, in areas of the economy characterized by radical rather than gradual reforms. China's ability to grow rapidly *despite* gradual reforms reflects China's particular economic structure.
- (H.2) China's gradualist strategy was not and is not transferable to Eastern Europe and the former Soviet Union (hereafter, EEFSU), because of fundamental differences in economic structure.
- (H.3) China's experiments in noncapitalist institutions are proving to be unsuccessful in agriculture; rural, small-scale industry; and state industry, and are therefore in need of further reform towards more typical capitalist institutions.
- (H.4) China is gradually harmonizing its economic institutions with those of East Asian market economies, though the political economy of harmonization is likely to be tricky.

A great deal of the debate regarding China actually surrounds the claims for and against the “shock therapy” or “big bang” reforms in Eastern Europe. After several postcommunist Eastern European economies embarked on radical economic reforms at the start of the 1990s, an intense debate has raged over the relative merits of gradualist versus big bang reforms. Therefore, before proceeding to the Chinese experience itself, it is useful to clarify the real content of big bang reforms. In our usage, big bang reforms include: rapid and comprehensive price and trade liberalization, macroeconomic stabilization, alignment of the official exchange rate to the market rate, ending legal discrimination against all types of non-state enterprises, and an early *commitment* to mass privatization of state-owned enterprises (SOEs), recognizing that actual privatization will take several years in practice.⁴

The avowed long-term aim of big-bang reforms is to create a “normal” capitalist economy, based on private ownership, commercial law, and substantially open trade. In the case of Eastern Europe, normality is typically defined as the economic institutions of the mixed capitalist economies of Western Europe. In operational terms, normality is often defined as membership in the European Union, including adoption of the requisite body of law known as the *acquis communautaire*. In the philosophical approach of “big bang” advocates, the key point is that little institutional experimentation is needed or desired. *The long-run goals of institutional change are clear, and are found in the economic models of existing market-based economies.*

We summarize the differences in outlook across the two schools of thought in table 21.1. Our aim in this chapter is to examine the contrasting interpretations of the China's rapid economic growth by looking at the key questions that divide the two approaches:

1. Is China's rapid growth the result of gradualist policies?
2. Would Chinese-style gradualism have improved performance in the EEFSU? Would big bang reforms have improved the performance in China?
3. Have China's institutional reforms succeeded in agriculture, state enterprises, and rural industry?
4. In the future, will China's reforms be guided by institutional experimentation and innovation, or rather by harmonization with East Asian economies?

The next section of this chapter addresses the first and second question by discussing the main sources of growth in China, and offering some comparisons of experiences in China and EEFSU. The following two sections address the third question by describing the responses of the SOE sector and the rural sector (both agriculture and the TVEs), to China's partial reforms. The fourth question was answered

Table 21.1
Summary of key differences

	Experimentalist school	Convergence school
Desirable speed of reform in a transition economy	Sequential trial-and-error	Rapid and comprehensive liberalization, and commitment to making private ownership the overwhelmingly dominant ownership form
Actual speed of reform in China	Sequential trial-and-error	Mixed radical and gradual reform. Radical liberalization of agriculture, and of international trade in coastal provinces. Slow deregulation of SOEs, and of international trade in interior provinces; and gradual clarification of ownership rights. Hesitant privatization of SOEs
Reasons for gradualism (incrementalism)	Economic experimentation	Political compromise; ideological commitment to state ownership
Sources of rapid growth	Unintended virtuous cycle, and little dislocations from large shifts in policies	Existence of surplus agriculture labor; East Asia pattern of labor-intensive, export-led growth
Outcomes in the SOEs	Substantial improvements in production efficiency	Little technical progress; and overcompensation of SOE personnel and overinvestment that weaken the fiscal situation
Interpretation of the TVEs	Adaptations to China's economic conditions of still-developing markets	Continuing legal restrictions on private ownership, with deleterious consequences for long-term development
Future directions and pace of reform	Policies will change to reflect evolution in material conditions and lessons from continuing experiments	Policies will push China toward a normal private market economy with characteristics similar to other East Asian economies

when the 15th Party Congress decided to privatize 369,000 of China's 370,000 state-owned enterprises (SOEs).

2 The Sources of China's Economic Growth

China's reform strategy after 1978 has best been described (by both schools of thought) as a dual-track approach: the establishment of a market track in parallel to the existing plan track, with the market track (generally) increasing in impor-

tance over time. The dual-track approach started with rapid and comprehensive liberalization of the agricultural sector, but limited liberalization of the other sectors. The impressive growth of the agricultural sector upon marketization led to broader liberalization of the secondary (industrial) and tertiary (service) sectors in the 1980s, with important reforms beginning in 1984. The state-owned enterprises (SOEs), located mainly in urban areas, were liberalized by devolving incrementally to them decision-making power on production, marketing, and investment.

The agricultural reforms at the end of the 1970s rapidly placed a substantial proportion of the economy and workforce effectively outside of the state apparatus. Agriculture accounted for almost 40 percent of GDP at the end of the 1970s,⁵ but, more importantly, no less than 70 percent of the economically active population were peasant farmers. The growth of the non-state sector quickly spread beyond the abolition of the communes themselves. The steady relaxation of the regulations governing the registration and supervision of non-state enterprises since 1984 has caused the non-state sector, particularly in the form of community-owned enterprises located in the rural areas, to grow explosively. It must be emphasized, however, that the SOE sector is not withering away. Its share of total employment was 18 percent in 1978 and 1993—and there were 35 million more SOE workers in 1993 than in 1978.

Dual-track regional development was adopted as the means to integrate China gradually into the world trading and financial systems. In 1980 four southern coastal cities (Shantou, Shenzhen, Xiamen, and Zhuhai) were designated "Special Economic Zones" (SEZs). SEZs were given autonomy to experiment with new institutions, mainly related to the international economy, such as approval of foreign-funded enterprises and discretion over the collection of many taxes. SOEs operating within the SEZs were exempted from many parts of the central plan, labor regulations, and the tax code. The resulting phenomenal growth of the SEZs spurred other regions to demand and receive special privileges as well. The rise in the export-GDP ratio from 4 percent in 1978 to 23 percent in 1994 shows that export-processing has become a major component of China's industrialization strategy.

Sources of Growth in the Reform Period

Several aspects of China's economic growth performance since 1978 are not in dispute. First, the non-state sector (including household agriculture, rural industry, private enterprises, urban collectives, and joint ventures) has grown much faster than the state sector. Therefore, the share of state-owned enterprises in aggregate production has declined markedly. In industry, the share of SOE production fell

from 78 percent in 1978 to 69 percent in 1984, and then to 34 percent in 1994. Second, agriculture provided the major impetus to growth during 1978–1984, but played a much lesser role after 1984. As a summary statement, we can say that agriculture experienced a one-time burst of productivity growth in the first few years after the reversion to the household production system, but then slowed markedly after 1984.

A third undisputed point is that exports have been a leading sector in China's growth, consistent with the patterns of export-led growth seen elsewhere in East Asia. Total exports rose from US\$10 billion in 1978 to US\$149 billion in 1995. The share of exports of goods and services in GDP rose from around 6 percent in 1980 to around 21 percent in 1995. A sharply rising portion of exports comes from the non-state sector, including TVEs, private firms, and foreign investments (sometimes under contracts to, or in cooperation with, state enterprises).

Many China analysts have rebuked the EEFSU economies for undertaking the dislocations of "shock therapy." In our opinion, these analysts have given insufficient weight to the differing economic structures in the two regions. In the former Soviet Union, more than 85 percent of the work force was in nonagricultural state enterprises, compared with around 18 percent of the work force in China (table 21.2). Perhaps 99 percent of the labor force of the former Soviet Union (*including* the 14 percent of the labor force in state farms and collective farms) were entitled to an "iron rice bowl" under the Soviet system as of 1985.

Would it have been possible to engineer a two-track approach when 90 percent of the work force was in the first (i.e., state) track, compared with less than 20 percent in the Chinese case? Our answer is "no," and the best evidence is that gradualism was in fact tried. The Chinese two-track approach was indeed the *avored* model of communist reformers such as Mikhail Gorbachev in the Soviet Union,

Table 21.2

Distribution of employment by type of organization (% of total employment)

	China			Russia	
	1978	1984	1991	1985	1991
State enterprise	18.6	17.9	18.3	93.1	86.1
Collective agriculture	72.0	67.0	63.9	6.0	5.3
Urban collective	5.1	6.7	6.2	n.a.	n.a.
Industrial TVEs	4.3	7.6	10.0	n.a.	n.a.
Private and other	0.0	0.8	1.6	0.9	8.6

Sources: Sachs and Woo (1994).

János Kádár in Hungary, and Wojciech Jaruzelski in Poland in the mid-1980s. The model had obvious appeal. It seemed to offer the prospect of faster growth with minimal social costs or political disruption.

In each case, the Chinese model failed in EEFSU, for two main reasons. First, as in China, tax collections from the state sector declined after the operational autonomy of SOEs was increased, a consistently observed pattern that result in part from the fact that state enterprises can easily use the new discretion to skim state income and to hide state profits, and in part from the fact that non-state enterprises provide competition for the state sector. The decline in tax revenues was much harder to manage in the EEFSU, because huge social expenditure commitments and subsidies to preserve guaranteed jobs in the state sector—in both cases covering nearly 100 percent of the population—could not be financed out of the falling tax revenues.

Second, in both China and EEFSU, most state-sector workers did not voluntarily leave their “iron rice bowls” after legalization of the non-state sector. The non-state sector in China grew up as a result of new entrants to the labor force and workers leaving agricultural employment to take jobs in industrial or service sectors. In EEFSU, there was no such large pool of non-state workers to join the new enterprises. Labor force growth was essentially nil, and the non-state sector was virtually nonexistent. Even agricultural workers on the state farms and the collective farms in EEFSU had their own “iron rice bowls,” so that unlike Chinese peasants, they were mostly unwilling to leave their heavily subsidized jobs.

Table 21.2 makes clear that the Soviet Union, and Soviet-type economies in Eastern Europe, had no reserve of labor outside of the SOE sector that could provide the engine of growth for a new non-state sector. The growth of the non-state sector required the decline of the state sector almost as a matter of accounting, inasmuch as the state sector completely dominated the economy. Only by ending the subsidization of the state sector was it possible to free resources for the new non-state sectors in the economy. Two-track gradualism could not work in that context.⁶

In short, gradual liberalization in the EEFSU unleashed financial destabilization, but *without* promoting the growth of a new private sector because the state-enterprise sector was not cut in the gradualist phase. Conversely, recent studies have shown that *collapse* of production in the EEFSU economies after 1989 cannot plausibly be related to the speed of reforms. Rather, output declined because of vastly overextended heavy industry. When cheap energy sources from the Soviet Union ended—both because of falling energy production in the Soviet Union and rising prices charged by the Soviet Union (and then Russia) to the post-Soviet economies—and when the Soviet military-industrial sector declined, every economy

in EEFSU faced a sharp industrial contraction. But it was precisely the fast-reforming countries that first and most vigorously reversed the industrial declines through the growth of new industrial and service sector activities (see Aslund, Boone, and Johnson 1996).

Suppose that China had in fact pursued more rapid liberalization of the economy, including a harder budget constraint on state enterprises and a faster unification of product markets and the market for foreign exchange. How much larger would have been the dislocations in the economy? We cannot answer this crucial question with any precision, but it is instructive to look next door at the case of Vietnam. During 1985–1988, Vietnam implemented a gradual reform strategy that did not address serious macroeconomic imbalances. The program failed: inflation and import of rice accelerated while growth performance remained unchanged. In 1989 Vietnam enacted an Eastern-European style “big bang,” including across-the-board price liberalization, a 450 percent devaluation to unify the exchange market, and a tight credit policy. The collective farms were returned to family farms with long-term leases. Growth accelerated, inflation ended, agricultural productivity soared (turning Vietnam into a rice exporter in 1989), and small, non-state enterprises proliferated. The “big bang” did not cause an output decline in Vietnam as in Eastern Europe. The difference in outcome lies in Vietnam’s economic structure in 1989, in which 77 percent of the labor force was engaged in agricultural activities. As an overwhelmingly agricultural economy, Vietnam enjoyed the same gains as China from liberalization of agricultural, and the flow of peasants to the nonagricultural sector. Strong market-oriented reforms (macroeconomic stabilization and liberalization), not gradualism per se, tended to accelerate this shift.

3 The SOE Sector under Reform

Until September 1997, *market socialism* was the guiding force behind SOE reform, and this explains why the general reform direction had been the steady expansion of the operational autonomy of the SOEs with almost no serious discussions (until very recently) of privatization as a reform option. From 1979 onward, managers received in piecemeal fashion the rights to make decisions about bonuses, how and what to produce, pricing, marketing, and investment. In parallel with this expansion of managerial autonomy was the steady decontrol of prices, with the prices of final goods being gradually liberalized since 1979, and the prices of inputs liberalized via the dual-track system since February 1985.

Assessments of the above decentralizing reforms on SOE performance have differed widely. Adherents of the experimentalist school accord great success to the

state enterprise reforms in China, while adherents of the convergence school tend to see chronic failures in SOE reforms. On the experimentalist side, Jefferson and Rawski (1994, p. 58) have concluded that

reform has pushed China's state-owned enterprises in the direction of "intensive" growth based on higher productivity rather than expanded resource consumption. We observe a consistent picture of improved results—higher output, growing exports, rising total factor productivity, and increased innovative effort—against a background of gains in static and dynamic efficiency that reflect the growing impact of market forces.

However, China's own leadership has been much more pessimistic. Vice-premier for economy Zhu Rongji announced in 1996 that the SOE sector was deeply plagued with problems. According to him,

The current problems of SOEs are: excessive investments in fixed assets with very low return rates, resulting in the sinking of large amounts of capital; low sales-to-production ratio giving rise to mounting inventories. The end result is that the state has to inject an increasing amount of working capital through the banking sector into the state enterprises.⁷

Zhu's pessimistic verdict represents the dominant view of Chinese economists and officials throughout the reform period—a view shared by foreign economists of the convergence school. According to Naughton (1995), though, who continues to defend the results of the state enterprise reforms, this negative assessment is the result of ignorance on the part of Chinese observers and of ideological prejudice on the part of some foreign observers:

Focusing on profitability, [state bureaucrats] see the erosion in state sector profits as a profound crisis of the state sector. Without good measures of total factor productivity, they conclude that state sector performance is deteriorating. Foreign observers, hearing the cries of alarm from the state planners, shake their heads knowingly as they perceive still further evidence that state ownership is intrinsically inefficient. Neither party sees that the difficulties are the result of an ultimately beneficial transition to a different type of economy, and are entirely compatible with gradually improving efficiency.

The essence of Naughton's argument is that this united front of Chinese officials and foreigner observers against China's SOEs was built on ignorance of economic theory ("competition from new non-SOEs erodes SOE profits") and of the technically sophisticated literature on TFP estimation. The recent decision of the 15th Party Congress in September 1997 to privatize China's SOEs suggests that Naughton's views have lost support within the Chinese leadership, despite the obvious risks of the forthcoming process of privatization. In our view, Naughton grossly underestimated the growing Chinese dissatisfaction with the SOE sector because he did not give due weight to three interrelated phenomena:

1. The lag in total factor productivity growth of the SOEs relative to the non-state sector, and in absolute terms (e.g., Woo, Hai, Jin, and Fan 1994).
2. The SOEs' overcompensation of their employees and their "thirst for investment" have created a fiscal crisis for the state, and destabilized the economy macroeconomically (e.g., Fan and Woo 1996).
3. The increasing misappropriation of SOE profits by insiders has generated social uneasiness.

The Basic Problem with Socialist Reforms

The logic behind socialist reforms is flawed in that once managers had the discretion over expenditure, they could use accounting subterfuges to transfer the profits to themselves and the workers rather than to share the profits with the state.⁸ As Kornai has so often stressed, such behavior was supported by the manager's knowledge that the enterprise faced a soft-budget constraint; reported losses would be met not by unemployment or closures, but by soft loans, tax breaks, and budget subsidies. Not only is it beyond the capacity of the government to check accounting or management abuses when they are widespread, but auditors are corruptible as well. Because decentralizing reforms are normally accompanied by the lifting of price controls, claims of higher costs being responsible for low or negative profits are hard to disprove. These financial manipulations help to explain why an increasing number of China's state enterprises are losing money despite a booming economy that private firms are profiting from tremendously.

The investment expenditure of state enterprises, similarly, has soared in the reform period because the state enterprises know that resulting profits can often be privatized through inflated costs, and that any losses will be socialized by state subsidies given to avert unemployment. The justifications for the investments are hard to fault: an unprofitable enterprise needs technical renovation to render itself competitive, and a profitable enterprise should expand production to make more profits. Whenever the People's Bank of China caved in to the demand for investment loans as well as for bailouts, the result has been high inflation. The record shows that it has generally been hard for the People's Bank to resist the political pressure for easy money.

China's state enterprises have also engendered social instability beside price instability. The inability of the government to know the true financial state of the enterprises allowed their managers to embezzle the assets, inflaming popular opinion against the government.

The Road toward Privatization

None of the preceding SOE reforms fundamentally altered the ownership structure of the enterprises, though they did significantly affect the control structure. A fundamental change in official philosophy about SOE reform occurred at the end of 1993 when the Central Committee identified the ambiguity of property rights to be an important cause of the unsatisfactory performance of SOEs, and decided that

Large and medium-sized State-owned enterprises are the mainstay of the national economy. . . . [For them] it is useful to experiment with the corporate system. . . . As for the small State-owned enterprises, the management of some can be contracted out or leased; others can be shifted to the partnership system in the form of stock sharing, or sold to collectives and individuals.⁹

By the end of 1995, this decision had been formulated into the slogan of “holding on to the large SOEs, and freeing the small SOEs (*zhua da, fang xiao*).” The debate on SOE reform in China had progressed beyond whether privatization was necessary to debate over the definition of “large” and the optimal form of privatizing “small” SOEs. News reports of full-scale sales of small and medium SOEs became increasingly frequent. The best known example was Zhucheng city in Shandong province, which started privatizing SOEs in 1992 when two-thirds of its SOEs were losing money or just breaking even. Almost 90 percent of county-supervised SOEs in Zhucheng were already privatized by the middle of 1995.

With the spreading of privatization throughout the country, asset-stripping by managers accelerated to astronomical levels. The result was that SOE sector became a net drain on the state in 1996—SOE losses exceeded the sum of sales taxes collected by the SOEs and the income taxes paid by profitable SOEs. The abysmal financial performance of the SOE sector forced the Chinese leadership to confront directly the SOE problem. In September 1997, the Party announced a new two-component enterprise reform strategy.

The first component is the establishment of a shareholding system for all but the largest one thousand state enterprises. In this shareholding system, the state need not be the majority, or even the biggest, shareholder in every state enterprise. The first component, in short, is the conversion of “state socialism” to “people’s socialism”—which is hard to distinguish from “people’s capitalism.” The second component of the new reform strategy combines the remaining one thousand state enterprises into conglomerates. The claim is that this consolidation will generate economies of scale and enable easier supervision of state assets. This reform component is inspired by the perceived dynamism of Japanese *zaibatsus* and South

Korean *chaebols*, and fits in well with the policy predilections of the conservative wing of the Party and the interests of the heavy industry ministries.

A Preliminary Assessment of the 1997 Enterprise Reform Program

There is an increasing danger of a social explosion if the government cannot control asset-stripping by the managers between now and when the enterprises are privatized. This is the Eastern European phenomenon of managerial privatization when the state enterprise system enters into the “end-game stage.” The social revulsion at such massive stealing could be politically destabilizing, especially if workers are also being laid off during this period.

We doubt that the amalgamation of the already large state enterprises into bigger industrial combines would improve efficiency. Furthermore, although the number of units that the state would supervise is reduced, it is not clear that this would necessarily reduce the illegal transfer of enterprise assets to managers and their friends because each unit is now so much bigger than before. The government is, in fact, creating enterprise groups that will be too large to be allowed to fail, a development that is likely to be costly to the budget in the future.

The two components of the new reform strategy captures very well the essence of the entire economic reform effort since 1978: two steps forward, and one step backward. China's gradual pace of reform reflects the fact that policies are the products of compromises among competing interest and ideological groups.

4 Rural Reforms

Rural reforms have included fundamental changes in agriculture—the reversion to household farming after the dissolution of the commune system—as well as deep changes in nonagricultural rural sectors, especially the rise of the TVEs. Even though the quality of rural life has improved markedly under these reforms, there are serious debates, and doubts, about the extent and adequacy of the rural reforms, and the future prospects for the rural sector.

Agriculture

Agriculture is another quintessential case of partial reform. On one hand, the end of the commune system led to significant gains in efficiency and individual freedom in the countryside. On the other hand, the reforms stopped well short of true private ownership of farm land.

There was a jump in agriculture yield (kilos per hectare) growth after 1978, but it was followed by relative stagnation after 1985. The data suggest that the impressive agricultural growth in the early years of the agriculture reform was a one-shot improvement in productivity that followed the liberalization of the agricultural sector and the introduction of the household responsibility system for land tenure. A simple extrapolation exercise indicates that the big achievement of the 1978 agricultural was to return rice and wheat yields to their underlying trends that were suppressed by the stringent collectivist agriculture practices of the 1958–1977 era.

Three factors have contributed importantly to the agriculture slowdown after 1985. The first factor is farmers' uncertainty about future land-use rights. Despite the 1984 government decision that farmers could get leases up to fifteen years, Prosterman, Hanstad, and Li (1996) found in their field work that "local officials have not implemented this policy to any significant degree. . . . [In] many villages, representatives from the collective take back all the land in the village every three to six years and reallocate the plots [to adjust for changes in household size]." The result is that farmers have refrained from making the many small long-term improvements (e.g., digging wells and small feeder drains, applying more organic fertilizer) in the land that would have increased grain yield.

The second important factor for agriculture stagnation is that state procurement prices after the early 1980s have not been raised in line with the increases in input prices. A third factor contributing to the post-1985 slowdown in agricultural productivity growth has been the large reductions in state investment in agricultural infrastructure (e.g., irrigation works) in the years after 1979. The level of real investment in agricultural infrastructure in 1994, for example, was only 58 percent of the 1979 level.

In summary, it appears that China's agriculture sector is under duress because of incomplete deregulation, incomplete privatization, and unequal social policies. The stagnation of the rural sector could be a source of future political instability, as evidenced by the rural unrest in 1992 and 1993.¹⁰ Unless agriculture growth is greatly increased, it is likely that rural-urban migration will accelerate, adding to the 100–150 million rural-urban migrants that already constitute China's so-called floating population.

The Township and Village Enterprises

We must point out that there are two common usages of the term TVE that can be potentially confusing, the official usage in statistical collection and the academic usage in discussion of ownership type. The official statistical usage (coverage) has broadened over time. Prior to 1984, TVE referred to township-owned (*xiang-ban*)

and village-owned (*cun-ban*); and from 1984 onward, TVE statistics also include joint-owned (by several persons or families, *lian-ban*) and individual-owned (by one person or family, *geti*). The present official statistical usage gives the impression of TVEs being overwhelmingly private in nature because 87 percent of TVEs in 1994 were *geti*. This impression should be resisted because *geti* produced less than 27 percent of TVE output, and less than 19 percent of *industrial TVE* output.¹¹

However, most academic discussions on the ownership structure of TVEs implicitly use a narrower definition that covers only the enterprises that are registered formally (and increasingly falsely, in our opinion) as township-owned and village-owned. This implicit narrow definition explains why Naughton (1994) and Walder (1995) categorically described TVEs as "local government-owned." Unless otherwise noted, we will adhere to this narrow definition of TVEs as public-owned in the following analysis on the "ownership nature of TVEs."

As in other areas of China's reform, the success of TVEs has generated an intense debate. The experimentalist school views them as an important and highly successful institutional innovation, melding market incentives with public ownership. The convergence school, by contrast, views them as a partially successful halfway house to real private ownership. Whereas the experimentalist school emphasizes the special fit of the TVEs with China's undeveloped economic conditions, the convergence school emphasizes serious problems ahead unless China now moves to real privatization of the TVEs.

There are three main types of TVEs. The first type is known as the Jiangsu model because the local authorities in Jiangsu exercised tight controls over the TVEs (e.g., participating in production and investment decisions, and regulating wages and labor mobility across TVEs). The second type of TVE form is known as the Zhejiang model because the local governments in Zhejiang province, although a significant shareholder in many TVEs, normally refrain from intervening in the production, dividend, and personnel decisions of the TVEs, provided that the enterprise makes an annual contribution to the village funds. The Zhejiang TVEs resemble leased companies, with the difference that their managers could be removed by the local officials.

The third TVE form is true private enterprises masquerading as TVEs. In this case, the entire capital of the enterprise is from an individual or a small group, and the enterprise pays a fee to the local authority in order to be allowed to register itself as a TVE, a charade that is popularly referred to as "wearing the red cap." The main reasons for the desire to disguise the true ownership are lower tax rates, easier approval procedures, fewer restrictions on the size and operations of the enterprise, and shelter against possible reversal in the political fortunes of the reformers.

The terminological haze has thickened in the 1990s with the additional easing of restrictions on the registration of firms as TVEs, making the coexistence of true TVEs and red-capped private enterprises a common phenomenon in many places. Given the varieties of TVEs, the vagueness about their ownership and control, and their evolving nature, it is therefore natural that different authors have emphasized different "basic" characteristics of the TVEs, often without acknowledging their great diversity over time and space. For example, Nee (1996) regards TVEs as informal joint ventures between the state and the private sector, often with "extensive informal privatization of collective-owned assets and firms," whereas Walder (1995) views TVEs as "under a form of public ownership no different from the large urban state sector."

Explaining the Ownership Form

The TVE ownership structure is highly unusual by international standards. In most East Asian countries with rural industry, such as Indonesia and Thailand, ownership of small enterprises is private, often within a family. By contrast, TVE ownership is collective, at least officially. Much has been written to explain the appearance of collective-owned enterprises in rural China. For example, some members of the experimentalist school have argued that collective ownership is an effective way to raise capital funds for rural enterprise. According to Naughton (1994), "Banks are ill-equipped in the early stages of transition to process small-scale lending applications and assess risks. Local government ownership in China played a crucial role in financial intermediation. Local governments could better assess the risks of start-up businesses under their control . . . and serve as guarantors of loans to individual TVEs." Some members of the experimentalist school have even interpreted the TVE record as definitive proof against the conventional wisdom that private ownership is the natural ownership form of small-scale enterprises, and argued that what mattered for efficiency is not ownership but competition in product and factor markets (Nolan 1993).

Advocates of institutional convergence are skeptical of the experimentalist school's functionalist explanation of TVE ownership form, especially of its emphasis on the state's superiority in financial intermediation. Taiwan's small and medium private enterprises exhibited dynamic growth in the 1960–1985 period even though they were heavily discriminated against by the wholly state-owned banking system. This is because informal financial markets (curb markets) appeared "spontaneously" to cater to the needs of small Taiwanese firms. The power of market forces (when tolerated by the local authorities) to induce financial institutional innovations was also recently seen in Wenzhou city in Zhejiang Province when economic

liberalization began in 1979. Liu (1992) reported that "ninety-five per cent of the total capital needed by the local private sector has been supplied by "underground" private financial organizations, such as money clubs, specialized financial households and money shops."

Small private enterprises have flourished throughout East Asia, as well as in the transition economies of Eastern Europe. In Poland, for example, the number of small businesses grew from 700,000 in 1989 (the eve of Poland's radical reforms) to 1,800,000 in 1993 (Sachs 1993). In 1993, both the Polish small, *privately owned* Polish businesses and the Chinese TVEs (including *geti*) employed around 20 percent of the labor force, with one difference being that Poland achieved most of this employment transformation in just four years compared to fifteen years in China.

In the view of the convergence approach, an adequate general theory for TVE ownership structure should be based on two main considerations. First, private ownership was heavily regulated and discriminated against in many areas until recently. Although individual ownership (*geti*) was given constitutional protection in 1978, private ownership (*siren-ban*, e.g., shareholding) was given constitutional protection only in 1987. Therefore, (registered) collective ownership of rural industry arose as the primary response to the profitable niches created by central planning because of the severe disadvantages faced by enterprises registered as privately owned.

In short, according to the convergence school, the "market failures" identified by the experimentalist school are not caused by inefficiencies intrinsic to a private market economy (such as externalities and public goods). These so-called market failures are actually created by ideologically motivated constraints imposed by the state. Specifically, the banks have extended more loans to TVEs than to private enterprises because of state directives, and *not* because of the TVEs being intrinsically more efficient or because of the local banks' recognition that the local governments were better assessors of risks than themselves.

The second basic consideration for a general TVE theory is that the collective ownership of TVEs initially reflected the low labor mobility in the countryside, which resulted largely from the household registration system that tied the peasants to the land. Community ownership was plausible when community members expected to remain in the same place indefinitely, and there was also no complicating factor of inward migration into rural areas.

Future Prospects of the TVEs

The convergence school sees four basic problems with the TVEs. The first, and most obvious, is that collective ownership invites political interventions by the local government in the workings of local enterprises, to the detriment of efficiency and fair-

ness.¹² The second problem with collective ownership is that it defeats risk diversification. When a community puts its wealth only into the narrow range of enterprises in the community, its residents can end up losing everything—their jobs and their savings.

A third problem is that collective ownership limits the scale of operations of the enterprise. Currently, a TVE can grow as a result of new investments by the community (including reinvestment of profits) or through bank loans. It is difficult, however, to get outsiders to invest in the TVE, because the property rights of the outside investors would not be well defined or well protected. The fourth problem is that collective ownership limits the market for managerial control. Suppose that a rural entrepreneur has a good idea for a new enterprise. In a normal market economy, he might be able to raise his own money to start the business, or to purchase an existing business. Both of these options are currently difficult in China, because of the heavy bias toward collective ownership. Conversely, suppose that an existing enterprise has a bad manager, but one who is favored for political reasons by the local government. In a normal market setting, an outside buyer might approach the owners of the business and make a takeover bid, replacing the manager after buying the enterprise. This is unlikely with the collective ownership of the TVEs.

In addition to these problems with the TVE ownership form, a recent development has greatly increased the pressure on the TVEs to “clarify” their property rights. The output expansion of many coastal TVEs in southern China has forced them to rely increasingly on migrant labor from the poorer provinces. The original inhabitants in these richer provinces want to prevent the new residents from having an automatic share in the dividends of the collective-owned enterprises, and so some areas have converted the collective TVEs into “shareholding cooperatives” by corporatizing the TVEs and dividing the shares among themselves. The fact that the government has not clamped down on this decollectivization of TVEs has been viewed as implicit approval, and this has accelerated the conversion of TVEs to shareholding cooperatives.

5 Conclusions and Future Prospects

We have examined China's economic development in the context of the debate between the experimentalist school and the convergence school to present the different interpretations. These two broad interpretations of China's growth have vastly different implications for future reforms in China and elsewhere. The experimentalist school expects that China will continue to liberalize and innovate

in a tentative, incremental manner in order to minimize the cost of any single policy experiment's going awry. By contrast, the convergence school would expect China to continue to move closer to the institutions of China's capitalist neighbors in East Asia. As partial reforms lead to "contradictions" in the economy (such as continued losses in the SOEs, or stagnation in agricultural productivity), events will prompt China to deepen the reforms and "normalize" its institutions.

Perhaps the best test of the two approaches is whether China's policy choices are in fact leading to institutions harmonized with normal market economies or to more distinctive innovations. In our view, the recent trend, quite strongly, has been toward harmonization rather than innovation. In our understanding, this trend toward harmonization is the result of internal pressures as well as explicit external pressures, such as China's quest to join the World Trade Organization.

The year 1994, in particular, was a period of rapid reform. In January 1994, China ended its unusual dual currency system (of *renminbi* and foreign exchange certificates) by unifying the exchange rate. By 1996 this had culminated in the effective convertibility of the yuan for current account purposes. In 1994 China also began a major overhaul of its tax system. Tax farming (revenue contracts with provincial governments and SOEs) was replaced by setting up central government tax offices in the counties, introducing an enterprise income tax that is (more) uniform across enterprise types, and relying increasingly on the value-added tax for revenue. In summer 1994 China issued its first set of company laws. Special policy banks were set up to enable the existing state banks to move to lending that was based entirely on commercial principles. What became very clear over the year was the tolerance of the central government toward the conversion of TVEs into share-holding cooperatives and the removal of red caps by pseudo-TVEs.

By mid-1995, the phrase of "holding on to large SOEs and freeing the small SOEs" had the concrete meaning of transferring some 90,000 small industrial SOEs to the non-state sector by sales, leases, or mergers. In short, full privatization, and not merely partial privatization (known euphemistically as diversification of ownership, which is offering of a small portion of the SOE's equities on the stock market), has begun in China. The government also extended the maximum period for leases on farm lands to thirty years, up from fifteen years. The first private bank opened at the end of 1995.

China committed itself at the 1995 APEC summit meeting in Osaka to reduce its average tariff rate by a third in 1996. This step toward tariff convergence, as part of China's bid membership in the World Trade Organization, was completed by August 1996. China's innovation of a dual stock market where a firm offered A-shares to Chinese citizens and B-shares to foreigners was effectively ended in mid-1996 when

the restrictions against purchase of B-shares by Chinese citizens were relaxed. Finally, in September 1997 China announced that it would privatize all but the one thousand largest SOEs.

All of these changes are instances of China's harmonizing its economic institutions to those of its capitalist neighbors and trading partners, rather than of China's experimenting to create institutions with "Chinese characteristics." These post-1993 institutions in China are the results of China's studying and adopting market institutions that have proven to be beneficial in foreign settings, and not mainly the results of China's having invented new institutions through internal experience. The policy trends, it seems to us, represent a growing acknowledgment by the Chinese government that decentralizing reforms alone will not work, and that China requires deeper integration into world markets on internationally accepted terms, as well as a much clearer set of private-property institutions.

Finally, we do not want to leave the impression that harmonization will be easy. We have seen in the former Soviet Union that the absence of a tradition of rule of law, combined with the lack of accountability in politics, can push reforms far off track. There remain important vested interests in the state-sector in China that will militate against needed reforms. New forms of corruption are likely to accompany privatization, as insiders manipulate the privatization process in nontransparent ways. Civil society is surely too weak in many realms to provide an effect check on arbitrary power and corrupt practices. Thus, the road ahead is still fraught with uncertainty and obstacles. Nonetheless, we believe that China is on the road to a deeper harmonization and integration with the advanced market economies. We also suggest, based on China's own recent history as well as worldwide experience, that China will have the chance for sustained growth and a process of income convergence with the richer nations under such a long-term strategy.

Notes

1. Distinguished members of the experimentalist school include Peter Nolan (1993), Barry Naughton (1995), and Justin Lin, Fang Cai, and Zhou Li (1994).
2. Members of the convergence school includes Michael Bruno (1994), Gang Fan (1994), and Geng Xiao (forthcoming).
3. Among Kornai's voluminous writings on market socialism, perhaps the most comprehensive overview is chapter 21 of *The Socialist System* (1992). Among other things, this chapter demonstrates that China's state enterprises are subject to the same soft-budget constraint that Kornai documented in the Eastern European cases. Moreover, in discussing the possibility of a state-run capital market in China, Kornai has this to say: "There is no way to tell whether a state capital market will be established in countries that still follow the path of market socialism, for example China. If it is, experience will provide better grounds for a verdict. Without practical experience there is only guesswork to back any attempt to eval-

uate the proposal. I cannot conceal my doubts. It has not proved possible to simulate a real market by artificial means even in transactions relating to goods" (p. 503).

4. By these standards, the only transition countries to have actually implemented a big-bang reform are the Czech and Slovak Republics, Estonia, Poland, and Slovenia.

5. Measured using 1990 relative prices because relative prices in 1978 were regulated to favor industrial products.

6. In Sachs and Woo (1994), we summarized the argument as follows: "Differing performance primarily reflects different economic structures prior to reform: China was a peasant agricultural society, EEFSU was urban and overindustrialized. . . . China faced the classic problem of normal economic development, the transfer of workers from low productivity agriculture to higher productivity industry. In EEFSU, the problem is structural adjustment: cutting employment in inefficient and subsidized industry to allow new jobs in efficient industry and services."

Several economists have challenged our depiction of China as a predominantly agricultural economy in 1978 and of post-1978 growth as mainly normal economic development (e.g., Lin, Cai, and Li 1994; Lardy, personal communications). They interpreted China in 1978 as an overindustrialized economy because industry accounted for 48 percent of GDP in 1978 and 47 percent in 1994—no sign of industrial growth being higher than aggregate output growth. In our opinion, their interpretation is misguided. Most important, the argument over flows of workers and political economy of reform rests on the allocation of labor, not GDP. By this count, there is no doubt (nor disagreement) that China was substantially agrarian and rural. Second, official estimates of the GDP share of industry greatly inflate the industrial share, because industrial prices were artificially inflated relative to agricultural prices. Therefore, one step of reforms involved the step-by-step freeing of repressed agricultural prices. For example, if we use constant 1990 prices to weight the various sectors of the economy, the industry share of GDP was 37 percent in 1978 and 51 percent in 1994.

7. "Guo you qiye sheng hua gaige ke burong huan" (No time shall be lost in further reforming state owned enterprises), speech at the 4th meeting of the 8th People's Congress, *People's Daily, Overseas Edition* (March 11, 1996).

8. For example, extra compensation such as additional housing and holiday trips could be reported as investment expenditure and sales promotion expenses respectively.

9. "Decision of the CPC Central Committee on issues concerning the establishment of a socialist market economic structure," *China Daily*, Supplement (November 17, 1993).

10. See *Far Eastern Economic Review* ("Now the Peasant Revolt: Get Off Our Backs"), July 15, 1993; and *Japan Economic Newswire* ("Beijing Abolishes 37 Rural Taxes"), June 21, 1993.

11. *Xiang-ban*, *cun-ban*, and *lian-ban* are classified as "collective-owned enterprises" (*jiti*). Of total TVE output in 1991, *xiang-ban* accounted for 37 percent, *cun-ban* for 30 percent, *lian-ban* for 6 percent, and *geti* for 27 percent. Output from industrial TVEs accounted for 75 percent of total TVE output in 1991. Of total industrial TVE output, *xiang-ban* accounted for 41 percent, *cun-ban* for 34 percent, *lian-ban* for 6 percent, and *geti* for 19 percent. Data are from 1992 *TVE Yearbook*.

12. There are now a growing number stories in the official press about the negative interventions of local officials. For example, *China Daily* (February 18, 1994) reported: "But ambiguous ownership or property rights in the collective economy coupled with the long standing integration of government administration and enterprises management [have] led to some problems. Among these were incidences where some local officials embezzled or diverted collective enterprise funds or properties to other purposes, and the rights of enterprises or employees were repeatedly infringed upon" ("Rural firms set 3rd reform wave").

References

Aslund, Anders, Peter Boone, and Simon Johnson (1996). How to stabilize: lessons from post-communist countries. *Brookings Papers on Economic Activity* 1: 217–313.

- Bruno, Michael (1994). Our assistance includes ideas as well as money. *Transition* 5(1): 1–4.
- Fan, Gang (1994). Incremental changes and dual track transition: understanding the case of China. *Economic Policy* 19 (supplement): 99–122.
- Fan, Gang, and Wing Thyee Woo (1996). State enterprise reform as a source of macroeconomic instability. *Asian Economic Journal* 10(3): 207–224.
- Jefferson, Gary, and Thomas Rawski (1994). Enterprise reform in Chinese industry. *Journal of Economic Perspectives* 8(2): 47–70.
- Kornai, János (1992). *The Socialist System*. Princeton, N.J.: Princeton University Press.
- Lin, Justine Yifu, Fang Cai, and Zhou Li. (1994). China's Economic Reforms: Pointers for Other Economies in Transition. Policy Research Working Paper 1310, World Bank.
- Liu, Yia-Ling (1992). Reform from below: the private economy and local politics in the rural industrialization of Wenzhou. *China Quarterly* 130: 293–316.
- Naughton, Barry (1994). Chinese institutional innovation and privatization from below. *American Economic Review* 84(2): 266–270.
- Naughton, Barry (1995). *Growing Out of the Plan: Chinese Economic Reform, 1978–1993*. Cambridge: Cambridge University Press.
- Nee, Victor (1996). Changing mechanisms of stratification in China. *American Journal of Sociology* 101(4): 908–949.
- Nolan, Peter (1993). *State and Market in the Chinese Economy: Essays on Controversial Issues*. London: Macmillan.
- Sachs, Jeffrey (1993). *Poland Jumps to the Market Economy*. Cambridge, Mass.: MIT Press.
- Sachs, Jeffrey, and Wing Thyee Woo (1994). Structural factors in the economic reforms of China, Eastern Europe and the Former Soviet Union. *Economic Policy* (April): 102–145.
- Walder, Andre (1995). Local governments as industrial firms: an organizational analysis of China's transitional economy. *American Journal of Sociology*, 101(2): 263–301.
- Woo, Wing Thyee, Wen Hai, Yibiao Jin, and Gang Fan (1994). How successful has Chinese enterprise reform been? Pitfalls in opposite biases and focus. *Journal of Comparative Economics* (June): 410–437.
- Xiao, Geng (1997). *Property Rights and Economic Reform in China*. (In Chinese.) Beijing: China Social Science Press.

22 Stability and Growth: Commentary on a Commentary

Robert Solow

I first heard the name of János Kornai as the author (with T. Lipták) of a paper outlining a novel two-level planning algorithm (in *Econometrica*, January 1965, pp. 141–169). It had an interesting angle, if I remember correctly after all these years. The familiar translation of Lange-Lerner ideas into an algorithm had the center transmitting prices to the production units, which then offered quantities to the center, with iterations to follow. Kornai-Lipták had the center sending quantities to the production units, which then responded with prices.

Today, of course, he is known everywhere in the world, originally as a deep internal critic of the overcentralized, overextended, underdisciplined command economies, and now as the most widely respected observer and critic of postcommunist Hungarian economy and society. I felt some temptation to go back to the work of the early Kornai—there are two earlier papers in *Econometrica*—as a topic for this occasion, which is sentimental as well as scientific; but in the end it struck me as more appropriate and more satisfying to pay homage by commenting on his contemporary concerns.

In 1994 or 1995, Kornai contributed a long article to the leading newspaper in Hungary, published under the general heading “Kornai’s Critique of the Government.” The article was, as he says, an appraisal of the economic program of the new government led by the Hungarian Socialist Party. An English translation appeared in 1995 in *Acta Oeconomica* (47[1–2]: 1–38), under the more explicit title “Lasting Growth is the Top Priority: Macroeconomic Tensions and Government Economic Policy in Hungary.” My intention is to comment on some of the main points of that article, as a general macroeconomist who knows next to nothing about the Hungarian economy, now or earlier. It follows that I will be thinking less about specifically Hungarian questions than about broad issues of macroeconomic judgment and policy.

The “macroeconomic tensions” Kornai has in mind are inflation (just under 20 percent per year in 1994), unemployment (11–12 percent in 1994, down from 12–13 percent in 1993), the budget deficit, the current account deficit, and the continuing decline in production. (Real GDP fell by about 20 percent between 1989 and 1993.) As the title of the *Acta Oeconomica* paper suggests, Kornai emphasizes the then central importance of setting in motion a process of continuing growth of real output. It is a valid goal itself, he says, and moreover *everything* is easier to accomplish when there are more goods and services to go around year after year. Just to take one example: a necessary condition for faster economic growth is bound to be a shift in the composition and destination of output from consumption to

investment; and this would be far easier to achieve, in a not very rich country if consumption only needed to grow more slowly than output while investment grows more rapidly, than it would be if the only way to make room for more investment were to contract consumption in absolute terms. The same observation is equally relevant to policies of redistribution. Kornai would like to see Hungary shed the remnants of paternalism and move toward direct need-based redistribution. It is obviously easier to get agreement on this step, and easier to carry it out, if the aggregate consumption pool is growing.

Kornai is surely right about that. The thought leads to perhaps his most important criticism of the (then) current trend in economic policy discussions in Hungary (and not only in Hungary). There is a common tendency to talk about the achievement of economic policy goals one at a time. Specifically, given the ongoing inflation, not to mention the fiscal and current-account deficits, one is not surprised to hear that many in the government and outside of it took the view that *first* it was necessary to stabilize the economy—reduce the inflation, cut the budget deficit, and restrain imports—and *then* it would be time to restart economic growth. “The time is not ripe” for a growth policy, to coin a phrase.

Kornai is resolutely against this philosophy. He sees, first, that growth is just too important in the Hungarian context, to be put on the back burner; second, that a contracting GDP will make it far more difficult to achieve the other stabilization goals; and third, that more years of stagnation or decline will only make it harder to get growth going again if the time ever comes to do so.

There is a lesson here for the large Western European economies that have pursued just that policy to their own great disadvantage. Neither in Hungary nor in Europe is the correct alternative merely to say growth first, and then stabilization. There is no substitute for simultaneous pursuit of both goals. Of course, perfect balance is not achievable. Sometimes demand will run ahead of potential and sometimes demand will fall a bit short. But the explicit and credible intent should be to move ahead on both fronts. The proper reply to central bankers, who usually say that they will lower interest rates after they see fiscal consolidations, is: Fine, you can operate with a one-week lag.

Kornai makes it very clear that he is not recommending a *net* fiscal stimulus at the moment. His immediate goal is not simple expansion, despite the high unemployment rate, and certainly not a contractionary policy (which is the more immediate danger). He wants the government to aim at cutting back consumption just to the extent that it can encourage or induce increases in investment (and net exports). Such statements are often met with the charge of fine-tuning. But that is a debating trick; policy must aim at a target, even if everyone knows there will be ex-post

errors. One could phrase Kornai's proposal this way: whenever the potential for growth of real output appears to be there, encourage it to happen; and allocate the incremental output more to investment and less to consumption.

One has to say that Kornai makes a convincing case for the general policy line he advocates. He is trying to thread a path between irresponsible "dynamization" on one side, and the impulse, on the other wise, to squeeze the economy first, allowing unemployment and excess capacity to increase but pleasing the IMF, the Deutsche Bundesbank and others. (It is fascinating that a Socialist Party government should be so bent on pursuing the squeeze-first policy; maybe a distrust of democratic capitalism is shared by Marxism and monetarism. Or perhaps the attraction that ultra-orthodoxy exercises on communist and ex-communist governments comes from the habit of neglecting the interests of ordinary people, exactly as with right-wing governments.)

Perhaps it is enough to say that the "dynamization" policy is ruled out by its own history, and the "squeeze first" policy is ruled out by economic common sense. What is lacking in Kornai's presentation, it seems to me, is the sort of accounting of targets and instruments that Tinbergen taught us to do. Even without writing down a determinate model, which would be foreign to Kornai's intentions, one could list the targets: real GDP, the rate of inflation, the ratio of investment to consumption, the current-account balance relative to GDP, and perhaps certain distributional indicators (to be discussed later). Then one would need some indication of a collection of policy instruments that is available to the government and adequate to the task. One bonus from this approach is often a clarification of the number and variety of independent policy instruments actually at the disposal of the government. (If I seem a little wistful on this subject, it is because such an accounting might improve the rationality of macro-policy discussion in the United States. Here we tend to load every possible target onto the single instrument of monetary policy.)

I really have a substantial, not a methodological, point in mind here. Consider an important example. Kornai asserts that (real) wages are too high in Hungary and must come down. He may have it in mind that wages already exceed the marginal product of labor and constitute a major obstacle to further investment. But that is not his main point. He emphasizes something else: the fact that wages tend to feed consumption, so that the vital shift from consumption to investment can only occur through a fall in wages.

He has no illusions about the outrage that will be aroused by the suggestion that wages are too high in a country burdened by working-class poverty. He reflects, nevertheless, about *how* real wages can be curbed. (Because the inflation will continue for a while, even if at a decelerating pace, it is only necessary to arrange a slow

growth of nominal wages.) There are three possibilities to be considered. One is to have a deep recession for as long as it takes to force wages down. Another is the imposition of effective administrative wage controls. The third is voluntary restraint. Without being in the least naïve, Kornai opts for the third route. He urges that the Socialist Party may be uniquely able to take the lead in this process, especially if the sequence from reduced consumption to higher investment to faster productivity increase to eventual real-wage growth can be presented as the coherent policy in aid of which wages must temporarily fall.

Here I think a little ingenuity in the design of policy instruments might pay off. The life of the policy-maker is never easy when allocation and distribution get mixed together. The distributional losers are sure to be in opposition. Allocational goals may be in the general interest, but distributional means tend to be zero-sum. This suggests at least a gesture at disentangling them. If the main purpose is to curb consumption, then one could try a form of forced saving. Some part of wages—maybe a larger part for those with higher wages—could be paid in the form of non-negotiable bonds, or even equities, with the funds directed into infrastructure or private investment, perhaps through mutual funds. In either case, the bonds would accumulate market-based interest, and perhaps turn into equities at a later date. I do not know if such a scheme would be viable in Hungary, given history and habits. I mention it primarily to illustrate the clarification that can come from counting targets and instruments (with due attention to identities).

It is useful to look at Kornai's discussion of the current-account deficit with the same considerations in mind. To begin with, he rejects the idea that the proper way to move toward balance is by a policy of domestic contraction to reduce imports and create a strong incentive to find export markets. That is just another example of squeeze-first, and it has the disabling disadvantage that it would probably make lasting growth harder to achieve. Instead, Kornai endorses the carrying out of an overdue devaluation of the forint. He sees clearly that a once-for-all depreciation is unlikely to be sufficient, if only because the ongoing inflation can decelerate generally at best. So he advocates frequent or even continuous adjustment of the exchange rate. It is not clear whether he is thinking of floating the forint or making a series of administered changes.

The point I want to emphasize is that, in this instance, a fresh instrument comes along with a fresh target. And that is not all: Kornai also favors the active use of tariffs and other commercial-policy devices within the GATT rules. He is no protectionist, but he does not want to forgo any available tool at a critical time. I have no opinion about the workability of an effective tariff policy under Hungarian conditions; I am still thinking about the Tinbergenian accounting of targets and instru-

ments. If, in fact, both the exchange rate and commercial policy are available as instruments, it should be possible to accomplish a bit more than just improvement of the current account. The obvious angle is that the tariff instrument could contribute both to the improvement of the trade balance and to the promotion of investment. The first depends mainly on the level of the tariff structure, whereas the second could be accomplished by favoring imports of capital goods over imports of consumer goods. That is because tariffs need not be uniform. The combination of devaluation and tariff policy is, in addition, a net gain in the number of instruments.

This is another instance of the gain in policy-making that can come just from across-the-board attention to the number and variety of targets and instruments. The actual formulation of a policy mix is of course a quantitative matter and requires some sort of model, as well as some preliminary trial and error. As Kornai would have said, the details of policy are not the business of a single researcher; but I am suggesting that a goal-by-goal approach can be improved on, even at the qualitative level.

When I say that János Kornai's analysis and prescription make sense to me, I do not mean that I think they are just right for Hungary in 1995 or today. Only another close student of the Hungarian economy could make that judgment. I mean that his ideas apply across a much broader field, including some of the small and large countries of Western Europe.

At the end, Kornai muses aloud about the likely reception of his ideas. He thinks of his prescriptions as having little or no party-political content; they are not especially right or left in implication. He says he is speaking as a "professional," and indeed he is. But he is also quite aware that some parts of the political spectrum will at least be attracted to his recipe, and others will not. He is inclined to explain this by the fact that Hungarian political parties have their own groups of professional economic advisors, just as in the West; any serious policy ideas will be filtered by them without being passed on to the political process. He claims to be puzzled by this: why should political parties have their own "tame" professional economists, thus tending to narrow the range of economic ideas that will play a role in the parties' policy choices, and to group them in "packages" suitable for each party? It is a good question, whose answer may be complicated. Part of the answer may have to do with hidden affinities between ideas; but the rest may be a matter of powerful personalities, of force of habit, of "transaction costs." I suspect that a stable two-party system may be especially likely to develop in that direction, if only by virtue of continuity. Should I wish that for Hungary?

There is no sign that Kornai's ideas are now out of date in the Hungarian context. Real GDP rose in 1994 and 1995, but only by 2.9 and 1.5 percent, leaving it still well

below the 1989 peak. Correspondingly, the unemployment rate fell slightly in those years, but was still 10.6 percent in 1995, with no visible tendency for 1996 (the year in which I write) to be an improvement. Consumer-price inflation improved to 18.8 percent in 1994 but then rose sharply to 28.2 percent in 1995. The nominal devaluation of the forint just about kept pace with the price level, so there was no significant real depreciation. Nevertheless the current account improved very slightly; but imports of machinery and equipment were no larger than imports of consumer goods, not a very good sign. (Each was about 20 percent of total imports; raw material imports were 42 percent of the total.) If lasting growth was the top priority in 1994, it remained the top priority in 1996—and Kornai's message remains relevant.

23

The Surprising Success of Gradual Price Liberalization in Hungary

Wim Swaan

Price liberalization has widely been considered as a key issue in the transformation process from a centrally planned economy to a market economy. It has been in the center of the debate on shock therapy versus gradual reform (Lipton and Sachs 1990; Murrell 1992*a*, 1992*b*, 1993; Brada 1993). Hungary has generally been considered as an example of a gradual approach, not only in price reform, but in its general approach to postsocialist transformation. This chapter focuses on the issue of price reform in a narrow sense and will not go into the details of stabilization and restructuring in a wider sense. By way of introduction the background of systemic transformation in Hungary is presented in the next section.

Compared to the attention the topic of price liberalization has received, surprisingly little is known on what could be called the black box of price reform: the instruments of how prices were actually liberalized at the micro level. This institutionalist aspect is of particular importance when assessing gradualist approaches to price liberalization. As I will show, gradual price liberalization in Hungary took a form quite different from what is mostly implicitly assumed: it did not consist of a previously announced gradual increase of prices, but rather of a gradual, quick shift of pricing responsibilities from the authorities to the firms.

In assessing the success of a process of price liberalization four criteria are of relevance:

1. The speed of liberalization, both in absolute terms and relative to the original aims.
2. The level of inflation, both in absolute terms and relative to the original aims.
3. The absence of setbacks.
4. The fall of domestic output and the speed and degree of recovery.

On the basis of these criteria, gradual price liberalization in Hungary in the period 1988–1991 can be considered to be relatively successful.

The heritage of socialist reform obviously contributed to the relative success of gradual price liberalization in Hungary. Although relative prices had come somewhat closer to market prices during socialist reform, this in itself can only partly explain the success of gradual price liberalization. Despite terminological changes, prices had been far from free and huge deviations from market prices had continued to exist until 1988, in particular in the field of consumer prices. At least as important was the learning process that enabled both managers and officials to develop pricing capabilities. A more detailed discussion of price reforms under reformed

socialism in Hungary has been given elsewhere (Swaan 1989, 1990, 1991, 1993; see also Bodócsi and Cinkotai 1991).

By the time full price liberalization was politically feasible, a sophisticated set of instruments of price regulation had developed that enabled a gradual but relatively quick liberalization. This provided an opportunity to ease the combined impact of market failure and government failure, which unavoidably occurs during postsocialist transformation. Firms could acquire part of the necessary capabilities to set prices and operate on markets before being fully exposed to market forces. Moreover, the major reshuffling of relative prices required under postsocialist transformation could be loosely coordinated, unlike a situation where prices are liberalized all at once. This, however, required a certain minimum of administrative capabilities of the officials of the Price Office, which would have been inconceivable without the preceding reform process.

1 Price Liberalization and Systemic Transformation in Hungary

Unlike in most other socialist countries, the beginning of the transformation process in Hungary is difficult to date. There was not one major event that triggered the process of change. By around mid-1988, reformers within the Communist Party initiated a process of economic and political reform. The reformist wing in the party tried to strengthen its position by mobilizing society against the orthodox currents, which very quickly, however, turned against itself as well. At the end of 1988 free political parties were formed. During 1989 these parties negotiated democratic elections, which were held in spring 1990 and which led to the defeat of the reformist communists (Bruszt 1990).

Economic liberalization paralleled the process of political liberalization. In July 1988 the Communist Party endorsed a program of reform, including a partial liberalization of prices and imports. Initially, this was envisaged to take place within the framework of a regulated socialist economy. In actual practice, however, a gradual shift took place to full liberalization, not in the least because the Communist Party in the meantime lost political control. By the summer of 1989, the Price Office had already developed plans to transform itself into an Competition Office, which eventually took place by the end of 1990.

A very typical aspect of Hungarian reform in the period 1988–1990 was that it was not implemented according to a theoretical blueprint or even a central strategy. The various elements of reform were developed in partial isolation and were only loosely coordinated. In fact, quite a number of decisions regarding political and

economic reform during 1989 and 1990, including those on price reform, were actually no more than a legalization after the event of a situation that had already been shaped in practice. Thus, the larger part of price liberalization was already completed by the time of the first democratic elections.

Before going into the details of the process of price liberalization, it might be useful to briefly summarize the other elements of systemic transformation. More detailed accounts can be found in Hare and Révész (1992), Derviş and Condon (1994), World Bank (1995), Kornai (1995, 1996).

Privatization. After 1982 increased legal opportunities were created for small-scale private activities. With the new Company Law of 1989 and the start of large-scale privatization, full opportunities for the private sector were created. In Hungary state companies have been sold one by one. As compared to massive privatization through voucher schemes, this has had the advantage of attracting massive investments by strategic investors, in particular from abroad, but it has had the disadvantage of delaying restructuring in companies that could not be sold off easily—precisely the ones where restructuring was needed most.

Import liberalization. Parallel to price liberalization, imports were gradually liberalized, extending from around 40 percent of imports in manufacturing industry in 1989 to around 80 percent in 1992 (Kornai 1995, p. 21). This is generally considered as having been *too* successful: similar to what happened in Czechoslovakia and Poland, import liberalization was not sufficiently coupled with tariff adjustments, thereby leading to an opening up and competition for domestic producers unprecedented in trade liberalization elsewhere in the world (Gács 1994).

Taxes and subsidies. Value-added tax and personal-income tax were introduced in 1988. Total subsidies decreased from 17.08 percent of GDP in 1988 to 6.95 percent in 1993 (World Bank 1995, p. 167).

Fiscal policy, reform of the budget, and welfare expenditures. This has been by far the weakest aspect of Hungarian transformation. Reform of the budget and welfare expenditures was postponed until spring 1995. This postponement not only contributed to continuous budget deficits and (to a lesser extent) to trade balance deficits, but it also unnecessarily created public expectations on the possibility of maintaining a high level of welfare expenditures.

Monetary policy and exchange rate policy. The monetary authorities succeeded in maintaining sufficient control over the money supply. This formed an important contribution to the feasibility of price liberalization. Exchange rate policy has been

open to question. During the first years of the transformation process, the forint was actually revaluated in real terms, which is particularly notable as compared to competitors in the region, such as Poland and the Czech Republic, where economic transformation started with a major devaluation.

Banking and credit. Banking reform has been unnecessarily delayed and became only effective by the middle of the decade. Although foreign banks were allowed to set up representative offices, it was explicit policy of the first democratic government to keep the major existing banks in domestic (state) property. Because of the delay of the banking reform, repeated bailouts were necessary. The bankruptcy law of 1992, on the other hand, was extremely strict.

Transformation in Hungary was preceded by extended reforms during the socialist era. In 1968 directive planning instructions had been abolished, yet these were replaced by other forms of strict central control. Despite formal changes and various reforms in later years, the system basically continued as a centrally planned economy until 1989 (for overviews and analysis see Kornai 1986 and 1992).

2 Gradual Price Reform: The Shift of Pricing Responsibilities

Apart from the gradual reduction of subsidies and organizational issues such as the dissolution of the Price Office at the end of 1990, an important aspect of price liberalization in Hungary was the gradual shift of responsibilities in price setting from the price authorities to the firms. This last element deserves special attention, for it has been specific for the way price liberalization took shape in Hungary.

Basically four steps can be distinguished: official prices, the obligation to report intended price increases, official price consultations, and informal instructions. All steps were based upon instruments that had been developed during the process of socialist reform between 1968 and 1988. Clearly, during 1989 and 1990 not all products and enterprises had to undergo all stages, and some firms indeed started in the middle, based on the responsibility they had already built up under previous reform.

1. *Direct determination by the authorities* The official prices, as prices directly determined by the authorities were called, were divided into fixed prices, maximum prices, and loosely limited prices, but in practice most prices were set at the implied maximum. In 1968, when directive planning instructions had been abolished, the share of official prices was 77 percent. It further declined to 58 percent in 1977 and 38 percent in 1987. Formally all prices *not* belonging to the category of "official prices" were called "free prices." However, as will become clear from what follows,

“free prices” were actually far from being “free.” All forms of intervention to be discussed later in fact relate to these “free prices.”

2. *Obligation to report intended price increases for specified products* As a first step to liberalization, many producers had been subjected to the obligation to report intended price increases for specified products. This instrument had already been widely applied since 1973, and had grown in importance during the 1980s.

Partly it served the function of signaling to the authorities where decisions on price increases versus subsidy increases ought to be taken. Although formally price increases could be implemented after the period of suspension was over, a variety of instruments were applied to “convince” enterprises not to do so. Apart from simply increasing subsidies on the condition that prices would not be increased, enterprises could be threatened with rigorous auditing, or with reclassification of the product into the category of official prices (for a detailed discussion, see Swaan 1989).

With liberalization proceeding, however, in particular in 1990, but partly also during 1989, the latter methods were less and less applied, and suspensions mainly served the role of coordinating a process of simultaneous price adaptation throughout the economy.

3. *Official price consultations* A further step to liberalization were the so-called price consultations. This instrument was very intensively applied in 1988 (when it was introduced) and in 1989. They were still loosely applied in 1990, and abolished from 1991. The consultations consisted of a very detailed review, which was then approved by the Price Office. In part, these consultations were a formal legalization of the widespread practice of the Price Office and other organs to give informal instructions, for instance by conveying “expectations” on “reasonable prices.” This legalistic aspect was reflected in the fact that formal documents were signed and that a summary of the agreement was published in the official gazette of the Price Office (a detailed, quantitative analysis of the price consultations, backed by case studies, can be found in Swaan 1991).

4. *Informal instructions* As a final step before total liberalization, “only” the general rules on price formation were applied, extended with all kinds of informal instructions. Up until 1988 this involved considerable intervention. From 1985 the Price Office even formally gained the right to suspend any price increase that had come to its attention. To this end a chain of informational obligations was created, so as to assure that all price increases would indeed come to the attention of the Price Office before they were actually implemented.

During 1989 and 1990 substantial changes occurred. In 1989 a considerable number of enterprises for the first time experienced hardly or no informal intervention from the authorities. From 1990 onward, prices not subjected to any special form of intervention (most prices by that time) were indeed "free prices," for the first time since the introduction of this category in 1968.

Table 23.1 presents detailed data on how in the field of consumer-products price formation shifted from full central determination to market prices.

Table 23.1

Intervention in price formation, selected product groups of final consumption,^(a) 1977–1992 (percentage of turnover)

		Official prices	Obligation to report intended price increases ^(b)	Official	Total
Foodstuffs	1977	75	n.a.	—	n.a.
	1987	64	n.a.	—	n.a.
	1988	30	27	5	62
	1989	21	16	10	47
	1990	9	0	0	9
	1991	6	2*	—	8
	1992	0	2*	—	2
Durable consumer goods	1977	47	n.a.	—	n.a.
	1987	14	n.a.	—	n.a.
	1988	0	32	10	42
	1989	0	26	9	35
	1990	0	15	5	20
	1991	0	—	—	0
	1992	0	—	—	0
Pharmaceutical products	1977	98	n.a.	—	n.a.
	1987	91	n.a.	—	n.a.
	1988	70	21	0	91
	1989	0	91	0	91**
	1990	0	54	0	54**
	1991	0	—	—	50**
	1992	0	—	—	50**
Electrical energy, district heating, other household energy	1977	100	0	—	100
	1987	100	0	—	100
	1988	100	0	0	100
	1989	100	0	0	100
	1990	100	0	0	100
	1991	89	—	—	89
	1992	60	—	—	60
Total final consumption	1968	77	—	—	77
	1977	58	n.a.	—	n.a.
	1985	43	17	—	60

Table 23.1 (continued)

	Official prices	Obligation to report intended price increases ^(b)	Official	Total
1987	38	21	—	59
1988	22	21	5	48
1989	19	12	7	38
1990	16	6	1	23
1991	9	0*	—	9
1992	7	0*	—	7

Sources:

Total data:

- 1968: Béla Csikós-Nagy, *A magyar árpolitika, az 1979/80 évi árrendezés* (Hungarian price policy, the price revision of 1979/80), Budapest: Közgazdasági és Jogi Könyvkiadó, 1980, p. 54.
- 1977: Béla Csikós-Nagy, *Árpolitikánk időszzerű kérdései, 1985–1988* (Topical questions of price policy in Hungary, 1985–1988), Budapest: Közgazdasági és Jogi Könyvkiadó, 1985, pp. 86–87.
- 1985–1991: András Bodócsi, untitled manuscript.
- 1992: János Kornai, Eliminating the shortage economy: a general analysis and examination of the developments in Hungary, *Economics of Transition* 3(1), 1995, p. 24.

Sectoral data:

- 1977: Béla Csikós-Nagy, *Árpolitikánk időszzerű kérdései, 1985–1988* (Topical questions of price policy in Hungary, 1985–1988), Budapest: Közgazdasági és Jogi Könyvkiadó, 1985, pp. 86–87.
- 1987: András Bodócsi, untitled manuscript.
- 1988, 1989: András Bodócsi, *Az árrendszer reformjának előrehaladása* (Progress of reform in the price system). Mimeo, Budapest, December 1988, tables 2/c and 2/d.
- 1990–1991: András Bodócsi, untitled manuscript.
- 1992: János Kornai, Eliminating the shortage economy: a general analysis and examination of the developments in Hungary, *Economics of Transition* 3(1), 1995, p. 24.

Notes:

n.a.: Not available.

—: Not applied.

(a) The figures concern the combined impact of intervention in producer prices and consumer prices.

(b) The figures reported here refer only to the obligation to report intended price increases of specified products directly to the Price Office. They are exclusive of the general obligation to report price increases as part of the informational system that existed between 1985 and 1989 to facilitate informal instructions (see text). The data are also exclusive of the obligation to report intended price increases if the profit margin of the enterprise exceeded a specified limit, in the order of 5–10 percent of turnover. The latter obligation was abolished in 1988.

* Estimated by the author. For 1991 and 1992 the obligation to report intended price increases was restricted to edible oil, margarine, and paprika (foodstuffs), and writing and printing paper (cultural articles).

** For pharmaceutical products only, retail profit margins have been subjected to a maximum since 1989. For 1989 and 1990 this was combined with the obligation to report intended increases of producer prices. From 1991 onward maximized retail profit margins were the only instrument applied.

Data for 1993 are identical to those of 1992, except for the obligation to report intended price increases in cultural articles, which were abolished in that year.

As an example, consider how liberalization was implemented step by step in foodstuffs. In 1988 the share of directly determined, "official" prices decreased from 64 percent to 30 percent. To a large extent, however, this was coupled with the introduction of the obligation to report intended price increases (which was applied to a total of 27 percent of foodstuffs). In addition to this, market leaders in food retail were subjected to price consultations, both in 1988 and 1989. In 1989 further liberalization took place. Market leaders in beer and soft drinks, which had in 1988 been subjected to the obligation to report price increases and which had been subjected to centrally determined maximum prices in earlier years, were now part of the system of price consultations, in addition to some other market leaders of food industry. Consultations with these firms were pretty tough as compared to those with other firms (information based on various issues of *Magyar Közlöny* and *Ár-szabályozás*, the official gazette of the Hungarian government and the Price Office, respectively). By 1990 almost full liberalization took place, including import liberalization, while foreign direct investment was making headway.

Concerning the economy as a whole, price regulation was by 1991 restricted to a small group of products. The prices of water, of the larger part of household energy, and of some services continued to be determined by the government. For pharmaceutical products retail profit margins have been subjected to a maximum, but otherwise they have been left free. The obligation of reporting intended price increases, finally, was maintained until 1993 for paprika and writing and printing paper, until 1995 for edible oil and margarine, and was by 1996 not applied to any product at all.

3 Impact on Output and Prices

Having discussed the institutional changes in price regulation, let us consider the impact of liberalization on prices and output. Unfortunately, no data are available that are compatible with the sector classification the Price Office used with respect to final consumption as presented in table 23.1. Data on producer price regulation, import liberalization, inflation and output are presented in table 23.2. It should be noted, however, that due to peculiarities of the socialist reform process until 1987, producer price liberalization was less spectacular than consumer price liberalization.

It will not come as a big surprise that price liberalization is coupled with price increases. For no single year or single sector, however, do price increases exceed 50 percent. As can be expected, the combination of import liberalization, price

liberalization, and price increases went together with a fall in the volume of domestic output. The drop in domestic output, however, did not occur entirely parallel to the liberalization process. In some branches it happened only with delay, which is probably strongly related to CMEA disintegration; in other branches the fall in output even preceded price liberalization.

4 The Success of Price Liberalization

In the first section the following criteria were mentioned in judging the success of price liberalization:

- The speed of liberalization, both in absolute terms and relative to the original aims.
- The level of inflation, both in absolute terms and relative to the original aims.
- The absence of setbacks.
- The fall of domestic output and the speed and degree of recovery.

On the basis of these criteria, price liberalization in Hungary can be considered to be relatively successful in three ways:

1. It was relatively successful compared to price liberalization in other postsocialist countries.
2. It was even more successful compared to postsocialist countries that at times had tried to realize gradual price liberalization.
3. It can be considered to be one of the most successful aspects of the transformation process in Hungary itself.

Consider each in turn.

1. Prices were liberalized gradually, but quickly, over a period of around three years (1989–1991). Gradual liberalization in this field was not subjected to political setbacks. Institutional liberalization was even more or less completed before the first democratic election in spring 1990, with only a major change of energy prices to be implemented in 1991. Compared to other postsocialist economies, inflation remained very low in the period of liberalization, not exceeding 50 percent in any one year, with 35 percent for 1991 as the highest rate of inflation of the period.

2. Hungarian price liberalization is particularly successful if compared to other countries where plans were developed to introduce gradual price liberalization

Table 23.2
 Producer price liberalization, import liberalization, and its impact on output and inflation, selected branches of manufacturing industry, 1978-1992

	Official prices	Obligation to report intended price increases	Official price consultations	Total direct intervention in prices	Import liberalization	Price increase of domestic industrial sales	Domestic output growth (volume)	Export growth (volume)	Total output growth (volume)	Share of ruble exports in total sales	Share of exports in total sales
	Share in turnover (in percentages)										Share in total sales (in percentages)
Mining	1978	75	n.a.	—	n.a.	—	—	—	—	—	—
	1987	80	n.a.	—	n.a.	—	—	—	—	—	—
	1988	75	0	0	90	2.0	1.4	0.5	1.4	2.2	3.7
	1989	75	0	0	75	2.7	-3.6	6.9	-3.3	2.1	3.8
	1990	75	0	0	75	7.3	-11.6	-8.3	-11.4	2.0	4.4
	1991	52	—	—	50	31.0	-14.1	-21.4	-14.4	1.5	3.8
Engineering	1992	25	—	—	25	48.3	-7.0	-52.6	-8.8	—	2.1
	1978	14	n.a.	—	n.a.	2.1	-17.1	22.2	-16.1	—	n.a.
	1987	0	n.a.	—	n.a.	—	—	—	—	—	—
	1988	0	15	15	30	4.5	7.6	-0.7	4.0	29.7	41.2
	1989	0	14	8	22	-1.4	0.4	2.6	1.2	25.4	37.2
	1990	0	7	6	13	13.8	-7.7	-0.2	-4.6	27.5	42.7
Food industry	1991	0	—	—	0	15.4	-7.5	-24.2	-14.8	18.4	39.0
	1992	0	—	—	0	30.2	-24.4	-31.7	-27.9	—	37.6
	1978	13	n.a.	—	n.a.	14.4	-29.8	-13.3	-23.1	—	n.a.
	1987	8	n.a.	—	n.a.	—	—	—	—	—	—
	1988	3	32	0	35	5.4	-0.8	5.2	0.2	6.1	17.9
	1989	3	14	5	22	4.8	0.5	9.2	1.8	5.6	16.6
	1990	0	0	0	0	19.3	-1.2	4.9	0.1	5.1	22.5
	1991	3	2*	—	5	30.8	-3.1	-10.0	-4.7	3.6	21.9
	1992	0	2*	—	2	19.0	-3.5	-1.9	-3.1	—	24.1
						13.0	-4.4	-21.5	-8.0	—	n.a.

Total industry	1978	30	n.a.	—	n.a.	—	3.7	3.8	2.6	3.5	12.2	24.1
	1987	22	n.a.	—	n.a.	—	—	—	—	—	—	—
	1988	20	6	4.1	55	—	4.1	-0.6	6.0	0.8	10.0	22.4
	1989	19	23	6	48	43	15.4	-4.6	0.7	-3.9	10.8	27.7
	1990	19	15	2	36	71	24.2	-6.8	-16.5	-9.5	6.8	25.7
	1991	13	1	—	14	72	31.9	-16.6	-21.9	-17.9	—	24.4
	1992	8	1	—	9	77	9.7	-15.9	-8.0	-14.0	—	n.a.

Sources:

Price regulation:

1978: Béla Csikós-Nagy, *Árpolitikánk időszerű kérdései, 1985–1988* (Topical questions of price policy in Hungary, 1985–1988), Budapest: Közgazdasági és Jogi Könyvkiadó, 1985, p. 76.

1987: András Bodócsi, untitled manuscript.

1988, 1989: András Bodócsi, Az árendszert reformjának előrehaladása (Progress of reform in the price system). Mimeo, Budapest, December 1988, tables 2/c and 2/d.

1990, 1991: András Bodócsi, untitled manuscript.

1992: János Kornai, Eliminating the shortage economy: a general analysis and examination of the developments in Hungary, *Economics of Transition* 3(1), 1995, p. 24.

Data for total industry are computed by the author on the basis of the weights of gross output by branch (source: Iparstatistikai évkönyv [Yearbook of Industrial Statistics] 1978, 1987, 1988, Gazdaságstatistikai Évkönyv [Yearbook of Economic Statistics] 1990, 1991, 1992).

Import Liberalization:

1989–1990: János Gács, A liberáló második éve (The second year of liberalization), *Figyelő* 35(1), January 3, 1991, p. 9.

1991–1992: János Kornai, Eliminating the shortage economy: a general analysis and examination of the developments in Hungary, *Economics of Transition* 3(1), 1995, p. 21.

Data for 1989–1990 and 1991–1992 are not comparable.

Domestic Price Increases:

Iparstatistikai évkönyv (Yearbook of Industrial Statistics) 1987, 1988, Gazdaságstatistikai Évkönyv (Yearbook of Economic Statistics) 1990, 1991, 1992.

Output and Export Growth:

Iparstatistikai évkönyv (Yearbook of Industrial Statistics) 1987, 1988, Statistikai Évkönyv (Statistical Yearbook) 1989, 1991, 1992.

Share of Ruble Exports and Total Exports in Sales:

Statistikai Évkönyv (Statistical Yearbook) 1987, 1988, 1989, 1990, 1991.

Notes:

n.a.: Not available.

—: Not applied.

* Estimated by the author. For 1991 and 1992 the obligation to report intended price increases was restricted to edible oil, margarine, and paprika (food industry), and writing and printing paper (light industry).

(such as Romania, Russia, and the Ukraine). These plans either were not implemented in the first place, or when they were implemented, they ended up in the continuation of price controls, combined with high inflation.

3. In the longer run, performance of the Hungarian economy has not been unequivocally successful. Inflation stabilized at a relatively high rate (20–25 percent) and output has been recovering very slowly. To be sure, results are not too bad if compared to other postsocialist economies. As regards output recovery, only Poland has been doing significantly better, and stable inflation rates below 10 percent were still rare by 1995 (for the Czech Republic, Slovakia, Slovenia, and Croatia, but not for Poland). Nonetheless, in absolute terms performance has been disappointing in Hungary, in particular in view of the favorable starting point of the transformation process in Hungary and the expectations connected to that.

Although the relatively disappointing performance can definitely be related to the gradual character of the Hungarian transformation process as a whole (see Kornai 1996), it is hard to relate it to the gradual character of price liberalization in a narrow sense. Disappointed performance has rather been connected to the postponement of budgetary reforms, in particular as regards welfare expenditures, and the delays and setbacks in banking reforms, including the consolidation of bad debts. This kind of delays is an inseparable aspect of the political economy of gradual reform (see, for instance, Kornai 1996, who presents a detailed account of Hungarian gradualism).

Yet, in spite of this, it is noteworthy that in the field of price liberalization no such delays and setbacks occurred. Indeed, price liberalization ranks among the most successful aspects of Hungarian transformation. It was implemented quickly, the goals were reached as expected, inflation was low, and there were no setbacks. The only other aspects of Hungarian transformation that were comparably successful are monetary policy, import liberalization, and the attraction of foreign direct investment, although in particular in the former two fields policy suffered from certain shortcomings.

5 The Heritage of Socialist Reform

The relative success of gradual price liberalization in Hungary can obviously be attributed to the reform process that had already started in 1968. However, the *way* that socialist reform contributed to later successful liberalization deserves special attention.

5.1 No Free Prices

Despite the fact that a substantial part of prices were officially called “free prices,” the price of hardly any product was free in the sense that it would have fallen outside the control of the authorities or that it would have been determined by market forces. In view of the prevailing regulatory practice in the period 1968–1987, it would have been more appropriate to speak of “regulated prices” instead of “free prices.” Above it was already noted that the so-called free prices were intensively regulated by such instruments as:

- The obligation to report intended price increases;
- The official price consultations;
- Informal instructions, combined with a comprehensive informational system to enable intervention.

Other regulations applied to the “free prices”:

- The Laws on Unfair Prices. These were very vague, but the threat of heavy sanctions was very effective in constraining price increases. The fundamental problem of the laws was that they were indispensable to counteract for monopolistic enterprise behavior, but at the same time completely obstructed any movement toward a somewhat more market-oriented business price policy.
- The so-called Competitive Price System, which was in force from 1980 until 1987. Despite its name, this was basically an administrative system that aimed to create a link between domestic prices and world market prices through a very complicated system of regulation. This was not coupled with any real import competition, however, and in practice the system ended up in endless bargaining and boundless exemptions, implying that the original aim was not reached. This is documented by both institutional studies (Swaan 1989, pp. 21–28, 40–42; Bodócsi and Cinkotai 1991) and econometric studies (Csoór and Mohácsi 1985; Simon 1987). The system was formally abolished in 1987, with the exception of the pricing of raw materials, where world market prices continued to be a (weak) guideline for domestic prices.
- Up until 1988, the cornerstone of the reformed price system was the principle that prices should be determined directly by the authorities in at least one stage of the production process, either at the start (producer prices of raw materials and basic intermediate products) or at the final stage (consumer prices). In the other stages, prices were “free” in the sense that they were regulated by the instruments described above: the obligation to report intended price increases, the laws on unfair prices,

the rules of the so-called competitive price system, the official price consultations, and last but not least all kinds of informal instructions. Following the terminology that Kornai (1986, p. 1701; see also Antal 1985, pp. 133–146) applied to socialist reform in Hungary, one might say that direct bureaucratic control over prices was replaced by a system of indirect, but comprehensive bureaucratic control.

An ingenious system was created, providing detailed control over business policy with less efforts of the authorities. Compared to the orthodox system of periodic price revisions which were out of date by the time they were implemented, this system had the important advantage that price changes could be implemented more easily.

5.2 Slightly Smaller Deviation from Market Prices

The fact that almost all prices remained under (indirect) control of the authorities of course does not exclude the possibility that the authorities would have brought prices closer to market prices. The available evidence suggests that this indeed had happened, although to a very limited extent only. The prices of sensitive consumer goods, such as meat, were gradually adjusted, in particular from 1978 onward, and were less subsidized than in other socialist countries. On the other hand, however, consumer prices of energy and of consumer services, such as rents, were hardly changed and remained extremely subsidized, similar to the practice in other centrally planned economies.

Figures on relative price changes following price liberalization indeed suggest that under reformed socialism prices in Hungary deviated substantially from market values, but slightly less than in other socialist countries.

1. In the period 1988–1991 relative producer price changes in twenty-one branches of manufacturing industry were very high, giving an indication of the huge deviation from market prices in the period before 1988. The changes were, however, even higher in Poland than in Hungary (mean absolute difference 11.5 percent in Hungary and 18.1 percent in Poland; standard deviation 15.5 in Hungary and 25.4 in Poland; OECD 1993, pp. 93–94). These figures give only a weak indication, as changes in producer prices were less dramatic than in consumer prices and as services and agriculture are not included.

2. The most significant difference from other centrally planned economies was probably in the field of energy and raw materials that were imported from the CMEA. As a part of the so-called competitive price system domestic producer prices of CMEA imports followed world market prices from 1980 onward, with the dif-

ferences flowing into the state budget. As a consequence, the impact of the huge shift of prices in Soviet trade, which occurred with the disintegration of the CMEA in 1991, had much less of an impact in Hungary than in other member states. Nonetheless, as the system of domestic prices had suffered from substantial weaknesses (exemptions and the like), domestic producer prices of CMEA imports increased considerably in 1991. For fuels and energy, domestic prices increased by 51 percent, which is much lower than the 165.2 percent price rise of CMEA imports, yet at the same time considerably higher than the average producer price increases of 31.9 percent for the whole economy (see Gács 1995, pp. 175–177). This delayed increase of energy prices also explains why producer price increases in 1991 were actually higher than in 1989 and 1990, the years of domestic price liberalization (see table 23.1).

It ought to be noted, however, that both indicators only concern producer prices. Particularly in the field of energy and fuel, consumer prices developed more or less isolated from producer prices.

5.3 Administrative Capabilities of State Organs

Although socialist price reform between 1968 and 1988 did not quite lead to a system of free prices or market prices, it provided a learning process for both managers involved with price setting and officials involved with price regulation. This learning process has strongly contributed to the success of gradual price liberalization between 1989 and 1991. Let us consider first the administrative capabilities of the state organs, inasmuch as this provided the framework of the process.

By the time full price liberalization became feasible, a very sophisticated set of instruments of price regulation had developed. Although most of these instruments have at times also been applied in regulated market economies (such as Belgium, France, and Sweden), none of these countries had simultaneously applied such a wide range of instruments as happened in Hungary. This broad set of instruments had developed over time, mostly by trial and error: many of the instruments originally came about as ad hoc countermeasures against unexpected and undesired enterprise behavior. The immediate availability of these instruments in 1989 was indispensable for the gradual shift of pricing responsibilities to the enterprises. Even if anybody had thought to create price reform along these lines by design, it would have been totally impossible to create the required sets of regulatory instruments overnight.

Moreover, application of these instruments requires a minimum of capabilities of the officials involved with implementing them, which is unlikely to have been

available in orthodox planned economies. Although the capabilities of lower level officials, in particular of auditors and of officials in the branch departments of the Price Office continued to pose problems (Swaan 1989, 1990), by the 1980s the leading officials in the Price Office had developed the capabilities required to coordinate a process of gradual liberalization toward market prices. They understood the working of markets, were aware of other measures required apart from price liberalization (such as import liberalization and antitrust policy), and were able to skillfully apply the instruments at their disposal. Obviously, this would have been impossible without the process of learning by doing they went through during the 1980s. Indeed, price liberalization was in 1988 and 1989 partly initiated by leading officials from the Price Office. This might be astonishing in view of the fact that this rendered their own position redundant. In fact, it was forward-looking behavior. The initiatives to liberalize prices were combined with successful endeavors to transform the Price Office into a Competition Office, which in the end took over a considerable part of its personnel, including the director.

5.4 Pricing Capabilities and the Shift of Responsibilities to Firms

Price-setting is probably not the most difficult task the management of a firm needs to master, in particular not if compared with technological and organizational capabilities, or the capability to tailor the firm's product mix to the demand of potential purchasers. Nonetheless, price-setting does require some skillfulness, which is not self-evident or given by nature.

The relative importance of acquiring pricing capabilities may be illustrated with the results of case studies of the impact of price consultations at Hungarian firms undertaken in the summer of 1989 (Swaan 1991). A standard question in the interview was how the firm's price policy would differ if it were left completely free. This was intended as to get an impression of the actual degree of intervention going on. Basically, however, the answers did not point to the problem of intervention, but to the importance of pricing capabilities. Part of the managers answered that the absence of any price regulation would not make much difference: they knew the tricks of how to handle the Price Office. This typically concerned firms that traded their own products on foreign markets and that on the domestic market too were accustomed to some responsibility of their own in price setting. Other managers were less specific in their answer. One manager frankly told that he would not have the faintest idea what to change when prices were left free. This was a firm in food industry. Not only had its prices been strictly regulated, but it also had suffered from strong intervention in the determination of the product mix. Firms like this one simply had no idea how demand would respond to a combined change of the prod-

ucts offered and their prices, at a time when supply at all consumer markets was changing and when real incomes were declining. By gradually increasing the pricing responsibility of the firm, it could develop the necessary capabilities. The sophisticated system of price regulation as it had developed by 1989 proved to be very instrumental in this.

6 Conclusion: Prices, Capabilities, and Coordination

The way price liberalization took place in Hungary has a number of interesting implications for our understanding of the transformation of postsocialist economies.

First, shock therapy appears not to be the only way of effectively liberalizing prices, contrary to what has frequently been argued (see, for instance, Lipton and Sachs 1990; Blanchard et al. 1991). The reason for the success of gradual price liberalization in Hungary was *not* in the first place that relative prices were already close to market values. Rather, socialist reform in the period 1968–1988 appeared to have been instrumental in the creation of a very sophisticated set of instruments of price regulation without which gradual liberalization would have been difficult to conceive. Moreover, it contributed to the development of pricing capabilities within firms.

A frequently heard argument against gradual price liberalization is that it would lead to intertemporal speculation: if a path of gradual price adjustment were to be announced, the argument goes, producers and traders would hold back output, thereby jeopardizing the feasibility of the program (Blanchard et al. 1991, pp. 29–30; Van Wijnbergen 1992). In Hungary this type of intertemporal speculation occurred only to a very limited extent, if at all. By gradually increasing the responsibility of the enterprises, instead of following a centrally determined path of price adjustment, the potential revenues of intertemporal speculation were reduced, while at the same time pricing capabilities of the firms were enhanced. Moreover, as regards raw materials, where intertemporal speculation is most likely to occur because of easy storage possibilities, pricing practice had been different in Hungary. This was probably the only field where the deviation from market prices was substantially lower in Hungary compared to other socialist economies, due to artificial attempts to apply world market prices to CMEA imports from 1980 onward.

Another, related argument against gradual price liberalization is that commitment to it cannot be sustained over a longer period. Here, too, the Hungarian case is a special one. Liberalization of the institutional framework of price control was more or less completed by the time of the first democratic election in 1990. Quite surprisingly, it largely took place behind the scenes, while the political debate in the

period 1988–1990 focused on other issues. As I pointed out, earlier, the problem of sustained commitment to gradual reform has in the end been quite considerable in Hungary, in particular as regards budgetary and fiscal reform. Nonetheless, price liberalization in a narrow sense was not at all jeopardized by it.

As regards price liberalization in a narrow sense, the approach followed in Hungary has important coordination advantages as compared to sudden “cold turkey” liberalization as undertaken in Poland in 1990 and Russia in 1992. The grand reshuffling of relative prices as required under postsocialist transformation gives rise to short-run coordination problems between firms, if prices are liberalized all at once. Firms have to make (wild) estimates of how input prices will develop, and how the simultaneous changes of output prices will affect demand. Because firms will prefer to be on the safe side, prices may initially be set too high, leading to a much higher rate of inflation in the first months of the stabilization program than expected. This is reinforced by the initial lack of pricing capabilities and the temporary continuation of routinized enterprise behavior as it developed under socialist planning (cf. Swaan and Lissowska 1992). As long as monetary and fiscal policy succeed in being sufficiently strict (as it was in Poland but not in Russia), this iterative process will settle at a lower inflationary rate within six to twelve months. In Hungary, the short-run coordination problem of a simultaneous change in relative prices could be partly avoided, as the process was implicitly coordinated by the Price Office.

The foregoing two points related Hungarian price liberalization to the general arguments made in the debate on shock therapy versus gradualism. The third implication considers the role of prices in postsocialist transformation, and in economic coordination in general. Although price liberalization is far from a trivial problem, as is witnessed by the problems various countries have had in implementing it, it nonetheless seems to be among the less complicated aspects of the transformation process. As I have argued here, price liberalization probably was the most successful aspect of transformation in Hungary. Similar conclusions seem to hold for Poland, which followed an entirely different approach under quite different circumstances.

Although price liberalization is a necessary condition for the emergence of markets, the Hungarian case (and indeed the Polish case, too) show that it is far from a sufficient one. The issues of ownership change and the development of legal and other supporting institutions appear to have been even more problematic.

The problem of postsocialist transformation seems to fit well with existing evidence that prices are just one aspect through which markets are coordinated. In fact, it is costly to create markets that clear through prices only (cf. Carlton 1989). Most

markets therefore rely on other mechanisms as well, such as reputations and networks of business contacts. Prices do not convey all information required, and actors should also rely on other forms of (tacit) knowledge on which to base their trading and producing decisions. This kind of argument has especially been put forward in Austrian or evolutionary approaches, which emphasize the dynamic, decentralized aspects of the market process, and which consider the metaphor of the Walrassian auctioneer as an incomplete model of the functioning of markets (see Langlois 1986, 1994; Thomsen 1992; and, among others, Hayek 1945 and Schumpeter 1934, 1939). In postsocialist economies the development of market-related capabilities, the development of networks and reputations, and the development of supporting institutions in general will require considerable time and effort (see Swaan 1997).

The findings of this chapter provide a counterexample for the argument that *generally* shock therapy or cold-turkey approaches would be the only feasible way to liberalize prices. In this sense it provides support for the view that gradual transformation may allow actors to adapt to new circumstances, thereby leading to a less drastic disruption of the economy (see, for instance, Murrell 1992a, 1992b; see also Swaan and Lissowska 1992). The Hungarian experience does *not*, however, provide a *general* argument in favor of gradual price liberalization. On the contrary, the way gradual price liberalization took place in Hungary would have been difficult to imitate in other postsocialist countries. Particularly, the administrative capabilities of the Price Office (and other state organs) have been crucial in the process.

In general, the problem of postsocialist transformation is the combined occurrence of (huge) market failure and (huge) government failure (cf. Kuznetsov 1993; Radosevic 1994; Eatwell et al. 1995, chapter 2). Whereas proponents of shock therapy emphasize the importance of particular kinds of government failure (credible commitment, sustainability), proponents of gradualism emphasize particular kinds of market failure (delayed adjustment of individuals and firms, time and efforts needed for institutional development). An appropriate policy choice, however, requires awareness of both approaches and consideration of the relative significance of them in each policy issue. Although, as has been argued, Hungarian transformation has been characterized by the lack of an explicit, coherent policy strategy, the policy choice actually made regarding price liberalization implicitly recognized both the presence of short-time market failures and the availability of appropriate administrative capabilities to compensate partially for this.

Notes

Research for this paper was supported by the Hungarian National Science Foundation OTKA, grant No. F 018297. I am particularly grateful to András Bodócsi for his continued support in providing data and

detailed institutional information on price regulation, and furthermore to János Gács for his comments on an earlier version. Responsibility for the paper remains my own.

The research leading to this paper has in many ways been inspired by János Kornai. Apart from the works quoted, his major books (1959, 1980, and 1992) have been very important in understanding the functioning of a socialist economy. I also benefited enormously from his personal advice, in particular during my stay in the early 1980s at the Institute of Economics of the Hungarian Academy of Sciences, which ten years later happened to become my employer. In addition, János Kornai's work had an important indirect influence through the bridge it formed to both Western and Hungarian institutionalism. Because I was educated at a university that was dominated by orthodox neoclassical and post-Keynesian approaches to economics, it was only through the reading of the works of János Kornai on the subject of planned economic systems, as taught by Michael Ellman, that I became first acquainted with institutionalist approaches to economics.

References

- Antal, László (1985). *Gazdaságirányítási és pénzügyi rendszerünk a reform útján* (The system of economic guidance and the financial system in Hungary on the road of reform). Budapest: Közgazdasági és Jogi Könyvkiadó.
- Blanchard, Olivier, Rudiger Dornbusch, Paul Krugman, Richard Layard, and Lawrence Summers (1991). *Reform in Eastern Europe*. Cambridge, Mass.: MIT Press.
- Bodócsi, András, and János Cinkotai (1991). Az állami árírányítás módszerei a '80-as években (Methods of price control in the 1980s). In László Gábor and Júlia Szalai, eds., *A drágaságról, adalékok az infláció politikai szociológiájához és gazdaságtanához* (Contributions to the political sociology and economy of inflation), 289–349. Budapest: MTA Szociológiai Intézet.
- Brada, Josef C. (1993). The transformation from communism to capitalism: how far? how fast? *Post-Soviet Affairs* 9(2): 87–110.
- Bruszt, László (1990). 1989: the negotiated revolution in Hungary. *Social Research* 57(2): 365–387.
- Carlton, Dennis W. (1989). The theory and the facts of how markets clear: is industrial organization valuable for understanding macroeconomics? In Richard Schmalensee and Robert D. Willig, eds., *Handbook of Industrial Organization*, vol. 1. Amsterdam: North-Holland.
- Csoór, Klára, and Piroska Mohácsi (1985). Az infláció fő tényezői, 1980–1984 (Main factors of inflation, 1980–1984). *Gazdaság* 19(2): 21–39.
- Derviş, Kemal, and Timothy Condon (1994). Hungary—partial successes and remaining challenges: the emergence of a “gradualist” success story? In Olivier Blanchard, Kenneth A. Froot, and Jeffrey D. Sachs, eds., *The Transition in Eastern Europe*, vol. 1, *Country Studies*, 123–153. Chicago: University of Chicago Press.
- Eatwell, John, Michael Ellman, Mats Karlsson, D. Mario Nuti, and Judith Shapiro (1995). *Transformation and Integration: Shaping the Future of Central and Eastern Europe*. London: Institute for Public Policy Research.
- Gács, János (1994). Trade liberalization in the CSFR, Hungary and Poland: rush and reconsideration. In János Gács and Georg Winckler, eds., *International Trade and Restructuring in Eastern Europe*, 123–151. Heidelberg: Physica Verlag.
- Gács, János (1995). The effects of the demise of the CMEA and the USSR on output in Hungary. In Robert Holzmann, János Gács, and Georg Winckler, eds., *Output Decline in Eastern Europe: Unavoidable, External Influence or Homemade?*, 161–191. Dordrecht: Kluwer Academic.
- Hare, Paul, and Tamás Révész (1992). Hungary's transition to the market: the case against a “big bang.” *Economic Policy—A European Forum* 14: 227–256.
- Hayek, Friedrich A. (1945). The use of knowledge in society. *American Economic Review* 35(4): 519–530.

- Kornai, János (1959). *Overcentralization in Economic Administration: A Critical Analysis Based on Experience in Hungarian Light Industry*. Oxford: Oxford University Press.
- Kornai, János (1980). *The Economics of Shortage*. Amsterdam: North-Holland.
- Kornai, János (1986). The Hungarian reform process: visions, hopes and reality. *Journal of Economic Literature* 24(4): 1687–1737.
- Kornai, János (1992). *The Socialist System: The Political Economy of Communism*. Princeton, N.J.: Princeton University Press.
- Kornai, János (1995). Eliminating the shortage economy: a general analysis and examination of the developments in Hungary. *Economics of Transition* 3(1): 13–37; 3(2): 149–168.
- Kornai, János (1996). Paying the bill for goulash-communism: Hungarian development and macro stabilization in political-economy perspective. *Social Research* 63(4): 943–1040.
- Kuznetsov, Evgenii (1993). How can an economy adjust to simultaneous market and government failure? Lessons from the Soviet Union, contemporary Russia and countries of late-late industrialization. *Communist Economies & Economic Transformation* 5(4): 473–497.
- Langlois, Richard, ed. (1986). *Economics as a Process: Essays in the New Institutional Economics*. Cambridge: Cambridge University Press.
- Langlois, Richard (1994). The boundaries of the firm. In Peter J. Boettke, ed., *The Elgar Companion to Austrian Economics*, pp. 173–178. Aldershot: Edward Elgar.
- Lipton, David, and Jeffrey Sachs (1990). Creating a market economy in Eastern Europe: the case of Poland. *Brookings Papers on Economic Activity* 1: 75–147.
- Murrell, Peter (1992a). Evolutionary and radical approaches to economic reform. *Economics of Planning* 25(1): 79–95.
- Murrell, Peter (1992b). Evolution in economics and in the economic reform of the centrally planned economies. In Christopher Clague and Gordon C. Rausser, eds., *The Emergence of Market Economies in Eastern Europe*, pp. 35–53. Oxford: Blackwell.
- Murrell, Peter (1993). What is shock therapy? What did it do in Poland and Russia? *Post-Soviet Affairs* 9(2): 111–140.
- OECD (1993). *OECD Economic Surveys: Hungary*. Paris: Organization for Economic Co-operation and Development.
- Radošević, Slavo (1994). Strategic technology policy for Eastern Europe. *Economic Systems* 18(2): 87–116.
- Schumpeter, Joseph A. (1934). *The Theory of Economic Development*. Oxford: Oxford University Press.
- Schumpeter, Joseph A. (1939). *Business Cycles*. New York: McGraw Hill.
- Simon, András (1987). An input-output analysis of prices in the Hungarian economy between 1981–1985. *Acta Oeconomica* 38(1–2): 17–33.
- Swaan, Wim (1989). Price regulation in Hungary: indirect but comprehensive bureaucratic control. *Comparative Economic Studies* 31(4): 10–52.
- Swaan, Wim (1990). Price regulation in Hungary, 1968–87: a behavioural-institutional explanation. *Cambridge Journal of Economics* 14(3): 247–265.
- Swaan, Wim (1991). Prices and market behaviour in Hungary in the early stages of the transition to a market economy. *Soviet Studies* 43(3): 507–533.
- Swaan, Wim (1993). Behaviour and Institutions under Economic Reform: Price Regulation and Market Behaviour in Hungary. Ph.D. dissertation, University of Amsterdam.
- Swaan, Wim (1997). Knowledge, transaction costs and the creation of markets in post-socialist economies. In Paul G. Hare and Junior R. Davis, eds., *Transition to the Market Economy*, vol. 2, 53–78. London: Routledge.

Swaan, Wim, and Maria Lissowska (1992). Enterprise behaviour in Hungary and Poland in the transition to a market economy: individual and organizational routines as a barrier to change. In Wolfgang Blaas and John Foster, eds., *Mixed Economies in Europe: An Evolutionary Perspective on their Emergence, Transition and Regulation*, 69–102. Aldershot: Edward Elgar.

Thomsen, Esteban F. (1992). *Prices and Knowledge: A Market-Process Perspective*. London: Routledge.

Van Wijnbergen, Sweder (1992). Intertemporal speculation, shortages and the political economy of price reform. *Economic Journal* 102(415): 1395–1406.

World Bank (1995). Hungary: Structural Reforms for Sustainable Growth. Report No. 13577-HU, June 12.

The Publications of János Kornai

Books

Overcentralization in Economic Administration. Oxford: Oxford University Press, 1959. Second edition, 1994. Also in Hungarian (1957, 1990).

Mathematical Planning of Structural Decisions. (With Tamás Lipták and Péter Wellisch.) Amsterdam: North-Holland, 1967; Budapest: Akadémiai Kiadó, 1967. Second edition, 1975. Also in Hungarian (1965), Slovak (1966), German (1967), and Polish (1969).

Anti-Equilibrium. Amsterdam: North-Holland, 1971. Second edition, 1975; third edition, 1991. Also in Hungarian (1971), Romanian (1974), German (1975), Japanese (1975), Polish (1977), and Croatian (1983).

Rush versus Harmonic Growth. Amsterdam: North-Holland, 1972. Also in Hungarian (1972), Czech (1977), Spanish (1977), and Chinese (1988).

Economics of Shortage. Amsterdam: North-Holland, 1980. Also in Hungarian (1980, 1982, 1989), Czech (1981–1982), French (1984), Polish (1985), Chinese (1986), and Russian (1990).

Growth, Shortage, and Efficiency. Oxford: Basil Blackwell, 1982; Berkeley and Los Angeles: University of California Press, 1982. Also in Hungarian (1982), Estonian (1985), Chinese (1986), and Polish (1986).

Contradictions and Dilemmas. Budapest: Corvina, 1985; Cambridge, Mass.: MIT Press, 1986. Also in Hungarian (1985), Vietnamese (1988), and Estonian (1992).

The Road to a Free Economy: Shifting from a Socialist System—The Example of Hungary. New York: W. W. Norton, 1990; Budapest: HVG Kiadó, 1990. Also in Hungarian (1989), Russian (1990), Czech (1990), Slovak (1990), French (1990), Italian (1990), Spanish (1991), Polish (1991), Ukrainian (1991), Estonian (1992), Japanese (1992), Serbian (1992), Tamil (1992), Sinhalese (1992), and Chinese (1994).

Vision and Reality, Market and State: New Studies on the Socialist Economy and Society. Budapest: Corvina, 1990; Hemel Hempstead and New York: Harvester-Wheatsheaf, 1990; New York: Routledge, 1990. Also in Hungarian (1989).

The Socialist System: The Political Economy of Communism. Princeton, N.J.: Princeton University Press, 1992; Oxford: Oxford University Press, 1992. In Hungarian (1993), German (1995), French (1996), and Bulgarian (1996).

Highway and Byways: Studies on Socialist Reform and Postsocialist Transition. Cambridge, Mass.: MIT Press, 1995. Also in Hungarian (1993).

Struggle and Hope: Essays on Stabilization and Reform in a Post-Socialist Economy. Chaltenham, England: Edward Elgar, 1997. Also in Hungarian (1996).

Az egészségügyi reformról (On the reform of the health system). Budapest: KJK, 1998. In Hungarian.

Edited Volumes

Non-Price Control. János Kornai and Béla Martos, eds. Budapest: Akadémiai Kiadó, 1981; Amsterdam: North-Holland, 1981. Also in Hungarian (1981).

La Voie Hongroise. Analyses et expérimentations économiques. (The Hungarian way: economic analyses and experiments.) János Kornai and Xavier Richet, eds. Paris: Calmann-Lévy, 1986. In French.

Articles

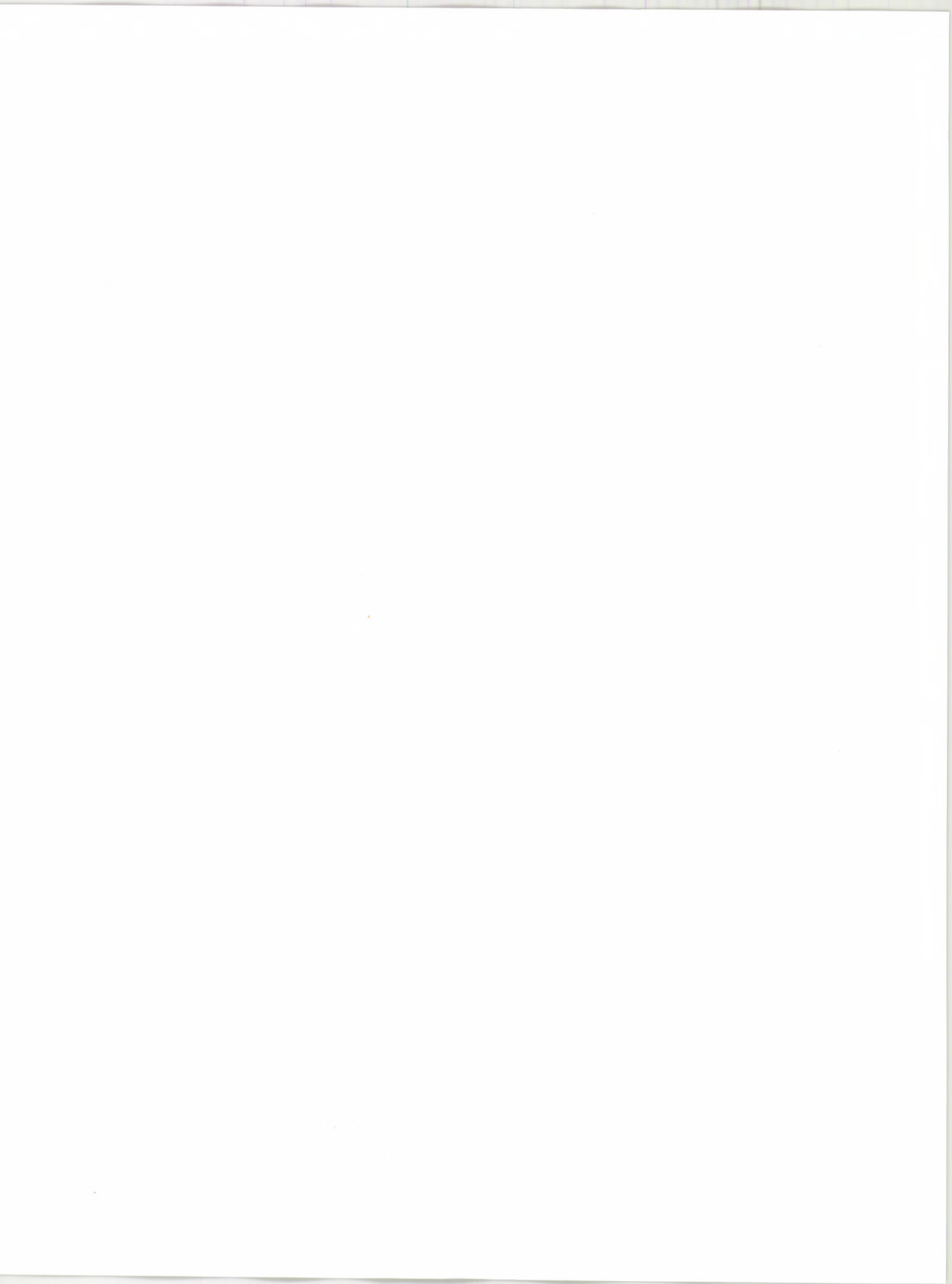
(1962). Mathematical investigation of some economic effects of profit sharing in socialist firms. (With Tamas Lipták.) *Econometrica* 30(1): 140–161.

- (1962). The application of the input-output table to determine the optimum development program of the aluminium industry. (With Béla Martos.) In Ottó Lukács, ed., *Input-Output Tables, Their Compilation and Use*, pp. 224–234. Budapest: Akadémiai Kiadó. Also in Hungarian.
- (1963). The determination of the optimum investment plan for an industrial sector by the use of linear programming. *Eastern European Economy* 1(4): 44–56. Also in Hungarian and Polish.
- (1965). Two-level planning. (With Támas Lipták.) *Econometrica* 33(1): 141–169. Also in Russian and Polish.
- (1965). Mathematical programming as a tool in drawing up the five-year economic plan. *Economics of Planning* 5(3): 3–18. Also in Russian and Serbian.
- (1966). Experiments in Hungary with industry-wide and economy-wide programming. (With Béla Martos.) In Bruno de Finetti, ed., *Mathematical Optimization in Economics*, pp. 169–194. Rome: Centro Internazionale Matematico Estivo.
- (1967). Hungary: the programming model of the national economy. In *Macroeconomic Models for Planning and Policy-Making*. Geneva: United Nations.
- (1967). Mathematical programming of long-term plans in Hungary. In Edmond Malinvaud and M. O. L. Bacharach, eds., *Activity Analysis in the Theory of Growth and Planning*, pp. 211–231. London: Macmillan; New York: St. Martin's Press.
- (1967). Application of an aggregate programming model in five-year planning. (With Zsuzsa Ujlaki.) *Acta Oeconomica* 2(4): 327–344. Also in Hungarian and Russian.
- (1969). Man-machine planning. *Economics of Planning* 9(3): 209–234. Also in Hungarian and Spanish.
- (1969). Multi-level programming: a first report on the model and on experimental computations. *European Economic Review* 1(1): 134–191. Also in Hungarian and Italian.
- (1970). A general descriptive model of planning processes. *Economics of Planning* 10(1/2): 1–19. Also in Spanish.
- (1971). Economic systems theory and general equilibrium theory. *Acta Oeconomica* 6(4): 297–317. Also in Hungarian, French, and Japanese.
- (1971). Plan sounding. (With Zsuzsa Dániel, Anna Jónás, and Béla Martos.) *Economics of Planning* 11(1/2): 31–58. Also in Hungarian.
- (1972). Macrofunctions computed on the basis of plan models. (With Zsuzsa Dániel and Judit Rimler.) *Acta Oeconomica* 8(4): 375–406. Also in Hungarian.
- (1973). Thoughts on multi-level planning systems. In Louis M. Goreux and Allan S. Manne, eds., *Multi-Level Planning: Case Studies in Mexico*, pp. 521–551. Amsterdam: North-Holland.
- (1973). Autonomous control of the economic system. (With Béla Martos.) *Econometrica* 41(3): 509–528. Also in Hungarian.
- (1973). Some intersectoral and intertemporal choice problems: Hungarian experience in long-term planning. In H. C. Bos, H. Linnemann, and P. de Wolff, eds., *Economic Structure and Development*, pp. 201–214. Amsterdam: North-Holland. Also in Spanish.
- (1973). Economic system theory and general equilibrium theory. In *Criticism of Contemporary Western Economic Theories*, pp. 7–32. Budapest: Institute of Economics, Hungarian Academy of Sciences. Also in Hungarian.
- (1975). Models and policy: the dialogue between model builder and planner. In Charles R. Blitzer, Peter B. Clark, and Lance Taylor, eds., *Economy-Wide Models and Development Planning*, pp. 13–31. London: Oxford University Press.
- (1975). Mathematical programming models in industrial development planning. *Industrialization and Productivity* (1975): 22.
- (1976). The measurement of shortage. *Acta Oeconomica* 16(3/4): 321–344. Also in Hungarian and Japanese.

- (1976). Pressure and suction on the market. In Judith Thornton, ed., *Economic Analysis of the Soviet Type System*, pp. 191–215. Cambridge: Cambridge University Press. Also in Hungarian, Russian, and Japanese.
- (1977). Decentralized control problems in Neumann-economies. (With András Simonovits.) *Journal of Economic Theory* 14(1): 44–67. Also in Hungarian and Russian.
- (1978). The normal state of the market in a shortage economy: a queue model. (With Jörgen W. Weibull.) *Scandinavian Journal of Economics* 80(4): 375–398. Also in Hungarian and Russian.
- (1979). The oeuvre of Kenneth J. Arrow. *Acta Oeconomica* 23(1–2): 193–203. Also in Hungarian.
- (1979). Resource-constrained versus demand-constrained systems. *Econometrica* 47(4): 801–819. Also in Hungarian, Portuguese, Chinese, and Spanish.
- (1979). Appraisal of project appraisal. In M. J. Boskin, ed., *Economics and Human Welfare: Essays in Honour of Tibor Scitovsky*, pp. 75–99. New York: Academic Press. Also in French.
- (1980). The dilemmas of a socialist economy: the Hungarian experience. *Cambridge Journal of Economics* 4(2): 147–157. Also in Hungarian, German, Polish, Japanese, Estonian, Chinese, Russian, and Spanish.
- (1980). “Hard” and “soft” budget constraint. *Acta Oeconomica* 25(3/4): 231–246. Also in Hungarian, Portuguese, French, and Chinese.
- (1981). On the difficulties and deficiencies of mathematical research in Hungary. *Acta Oeconomica* 26(1/2): 175–198. Also in Hungarian and Estonian.
- (1981). Some properties of the Eastern European growth pattern. *World Development* 9(9/10): 965–970. Also in Chinese.
- (1982). Adjustment to price and quantity signals in a socialist economy. *Économie Appliquée* 35(3): 505–524. Also in Japanese and Chinese.
- (1983). The health of nations: reflections on the analogy between the medical sciences and economics. *Kyklos* 36(2): 191–212. Also in Hungarian, Portuguese, Chinese, Russian, and Spanish.
- (1983). Equilibrium as a category of economics. *Acta Oeconomica* 30(2): 145–159. Also in Hungarian, Japanese, and Estonian.
- (1983). Paternalism, buyers’ and sellers’ market. (With Jörgen W. Weibull.) *Mathematical Social Sciences* 7(2): 153–169. Also in Hungarian and Japanese.
- (1983). Comments on the present state and prospects of the Hungarian economic reform. *Journal of Comparative Economics* 7(3): 225–252. Also in Hungarian, Japanese, Chinese, and French.
- (1984). Reproduction of shortage on the Hungarian car market. (With Zsuzsa Kapitány and Judit Szabó.) *Soviet Studies* 36(2): 236–256. Also in Hungarian, Japanese, and Chinese.
- (1984). Descriptive-explanatory models of the socialist economy: review of a research direction. *System Research* 1(2): 135–143.
- (1984). Softness of the budget constraint—an analysis relying on data of firms. (With Agnes Matits.) *Acta Oeconomica* 32(3/4): 223–249. Also in Hungarian, Japanese, and Chinese.
- (1984). Bureaucratic and market coordination. *Osteuropa Wirtschaft* 29(4): 316–319. Also in Hungarian, Japanese, Chinese, Estonian, and Russian.
- (1985–1986). Investment, efficiency and shortage: a macro-growth model. (With András Simonovits.) *Matekon* 22(2): 3–29. Also in Hungarian.
- (1985). On the explanatory theory of shortage: comments on two articles by K. A. Soós. *Acta Oeconomica* 32(3/4): 145–164. Also in Hungarian.
- (1985). Gomulka on the soft budget constraint: a reply. *Economics of Planning* 19(2): 49–55. Also in Hungarian and Japanese.
- (1986). The soft budget constraint. *Kyklos* 39(1): 3–30. Also in Hungarian, Japanese, and Spanish.
- (1986). State-owned firm, bureaucracy and market: Hungarian experience. *Third Chintaman Deshmukh Memorial Lecture*. Bombay: Reserve Bank of India.

- (1986). The Hungarian reform process: visions, hopes and reality. *Journal of Economic Literature* 24(4): 1687–1737. Also in Hungarian, Polish, Chinese, Estonian, Russian, and Spanish.
- (1986). The Chinese economic reform—as seen by Hungarian economists. (With Zsuzsa Dániel.) *Acta Oeconomica* 36(3/4): 289–305. Also in Hungarian and Chinese.
- (1987). The dual dependence of the state-owned firm in Hungary. In G. Tidrick and Chen Jiyuan, eds., *China's Industrial Reform*, pp. 317–338. New York and Oxford: Oxford University Press. Also in Chinese.
- (1988). Individual freedom and reform of the socialist economy. *European Economic Review* 32(2/3): 233–267. Also in Hungarian, Japanese, Russian, and Spanish.
- (1989). On the responsibilities of economic theorists, advisers and politicians. *New Hungarian Quarterly* 30(114): 170–177. Also in Hungarian and Russian.
- (1989). Some lessons of the Hungarian experience for the Chinese reformers. In Peter van Ness, ed., *Market Reforms in Socialist Societies: Comparing China and Hungary*, pp. 75–106. Boulder, Colo.: Lynne Rienner Publishers.
- (1990). The affinity between ownership forms and coordination mechanisms: the common experience of reform in socialist countries. *Journal of Economic Perspectives* 4(3): 131–147. Also in Hungarian, Croatian, Spanish, Russian, and Italian.
- (1990). My days as a naive reformer. *New Hungarian Quarterly* 31(119): 120–128. Also in Hungarian.
- (1991). Stabilization and economic transition in Hungary: the next two years. In Jaime de Melo and André Sapir, eds., *Trade Theory and Economic Reform, North, South, and East: Essays in Honor of Béla Balassa*, pp. 307–326. Oxford: Basil Blackwell.
- (1992). The principles of privatization in Eastern Europe. *De Economist* 140(2): 153–176. Also in Hungarian and Russian.
- (1992). The postsocialist transition and the state: reflections in the light of Hungarian fiscal problems. *American Economic Review* 82(2): 1–21. Also in Hungarian, Polish, Czech, German, and Slovak.
- (1993). Postsocialist transition: an overall survey. *European Review* 1(1): 53–64. Also in Hungarian, German, and Russian.
- (1993). Market socialism revisited. In Grethe B. Peterson, ed., *The Tanner Lectures on Human Values*, volume 14, pp. 3–41. Salt Lake City: University of Utah Press. Also in Hungarian.
- (1993). The Soviet Union's road to a free economy. In Grethe B. Peterson, ed., *The Tanner Lectures on Human Values*, volume 14, pp. 42–68. Salt Lake City: University of Utah Press. Also in Hungarian.
- (1993). The evolution of financial discipline under the postsocialist system. *Kyklos* 46(3): 315–336. Also in Hungarian, Russian, German, and Slovak.
- (1993). Transformational recession: a general phenomenon examined through the example of Hungary's development. *Économie Appliquée* 46(2): 181–227. Also in Hungarian, German, Czech, Polish, Russian, and French.
- (1994). Transformational recession: the main causes. *Journal of Comparative Economics* 19(3): 39–63.
- (1995). Eliminating the shortage economy: a general analysis and examination of the developments in Hungary. *Economics of Transition* 3(1): 13–37; 3(2): 149–168. Also in Hungarian, German, and Russian.
- (1995). Lasting growth as the top priority: macroeconomic tensions and government economic policy in Hungary. *Acta Oeconomica* 47(1/2): 1–38. Also in Hungarian, Russian, and German.
- (1995). The dilemmas of Hungarian economic policy. In Béla K. Király and András Bozóki, eds., *Lawful Revolution in Hungary 1989–94*, pp. 323–349. New York: Columbia University Press.
- (1996). Hardening of the budget constraint under the postsocialist system. *Japan and the World Economy* 8: 135–151.
- (1996). Paying the bill for goulash-communism: Hungarian development and macro stabilization in political-economy perspective. *Social Research* 63(4): 943–1040. Also in Hungarian, Czech, Slovak, German, and French.

- (1997). Adjustment without recession: a case study of Hungarian stabilization. In Salvatore Zecchini, ed., *Lessons from the Economic Transition: Central and Eastern Europe in the 1990s*, pp. 123–152. Dordrecht: Kluwer Academic. Also in Hungarian.
- (1997). The political economy of the Hungarian stabilization and austerity program. In Mario I. Blejer and Marko Skreb, eds., *Macroeconomic Stabilization in Transition Economies*, pp. 172–203. Cambridge: Cambridge University Press. Also in Hungarian and Hebrew.
- (1997). The reform of the welfare state and public opinion. *The American Economic Review* 87(2): 339–343.
- (1997). Reform of the welfare sector in the post-communist countries: a normative approach. In Joan Nelson, Charles Tilly, and Lee Walker, eds., *Transforming Post-Communist Political Economies*, pp. 272–298. Washington, D.C.: National Academy Press. Also in Italian and Croatian.
- (1997). Reforming the welfare state in postsocialist societies. *World Development* 25(8): 1183–1186.
- (1998). The citizen and the state: reform of the welfare system. *Emergo* (Winter 1998): 2–14.
- (1998). Legal obligation, non-compliance and soft budget constraint. In Peter Newman, ed., *The New Palgrave Dictionary of Economics and the Law*. New York: Macmillan. Also in Hungarian.
- (1998). The concept of the soft budget constraint syndrome in economic theory. *Journal of Comparative Economics* 26(1): 11–17.



Eric S. Maskin is the Louis Berkman Professor
of Economics at Harvard University.

András Simonovits is Scientific Advisor at the
Hungarian Academy of Sciences.

"*Planning, Shortage, and Transformation* gathers together work by an impressive number of outstanding economists who have given considerable thought to comparative systems and the organization of production and exchange activities. The book provides many interesting insights on the microeconomic determinants of transition and economic development."

Philippe Aghion, University College London and Harvard University

The MIT Press | Massachusetts Institute of Technology | Cambridge, Massachusetts 02142 | <http://mitpress.mit.edu>

MASLH 0-262-13357-1

