

Essays on Labour Markets in Russia and Eastern Europe

A thesis submitted to the Department of Economics, Oxford University
in partial satisfaction of the requirements for the degree of

Doctor of Philosophy

Maxim Vyacheslavovich Bouev
St. Antony's College
University of Oxford

Hilary Term, 2006



Памяти Петра и Альбины

Essays on Labour Markets in Russia and Eastern Europe

Copyright 2006

by

Maxim Vyacheslavovich Bouev

Approximate word count: 84,000 (including appendices, footnotes, figures, tables, and the list of references)

Essays on Labour Markets in Russia and Eastern Europe

by

Maxim Bouev

Doctor of Philosophy in Economics

St. Antony's College

University of Oxford

Hilary Term 2006

Abstract

This thesis is concerned with various aspects of transitional labour reallocation either between different labour market states, or between less and more efficient enterprises, or between formal and informal sectors. The possibility of irregular employment opportunities receives special attention in this work. The substantive material is arranged in three independent essays.

The first, empirical study portrays the most important trends in labour reallocation in Russia, and presents analyses of two types. First, transition probabilities are studied, and some determinants of worker flows are identified using a multinomial logit modelling. Second, a survival analysis of the non-employed is conducted to reveal possible causes of growing stagnancy of unemployment and inactivity. The findings are contrasted with the stylised theory of labour reallocation in transition (Aghion and Blanchard, 1994). The directions in which theoretical modifications should be attempted in future research are suggested. The second and the third essays draw upon some of these suggestions and are aimed at making a contribution on the theoretical front.

The second essay puts forward a development of the seminal model of transition from planned to market economy by Aghion and Blanchard (1994). We introduce an informal sector to show that its presence can generate the dynamics qualitatively different from the types considered in the previous literature on the topic. It is argued that convergence to qualitatively different steady states can help explain varying transitional experiences of East European countries and the former Soviet Union republics. Attention is drawn to policy implications of the model, in particular to the creation of conditions favourable for the development of the new private sector as opposed to informal private initiative.

Finally, the third essay takes the issue of coexistence of formal and informal sectors in transition further to see if such duality is possible in the long run, and to discuss the role of the government in creating preconditions for it. The study draws on the standard framework of Pissarides (2000) of search in the labour market. It demonstrates that a long-run equilibrium with both formal and informal economies is possible under very mild assumptions. It is also shown that labour market imperfections can create a situation when reduction in informality may be detrimental to economic welfare.

Although the foci of the essays differ, the issues raised therein are closely knit so that many threads can be drawn together. In the concluding chapter we discuss the main areas to which this thesis contributes, summarise the main findings, and make some suggestions for future research.

Contents

List of Figures	v
List of Tables	vii
1 Introduction	1
1.1 The Way of Thinking About Transition	4
1.2 Issues Tackled in This Thesis	8
1.2.1 "10 Years of Worker Reallocation in Russia: An Outlook on Distinctive Patterns and Implications for Theory" .	9
1.2.2 "Diverging Paths: Transition in the Presence of the Informal Sector"	11
1.2.3 "State Regulations, Job Search and Wage Bargaining: A Study in the Economics of the Informal Sector"	14
2 10 Years of Worker Reallocation in Transitional Russia: An Outlook on Distinctive Patterns and Implications for Theory	18
2.1 Introduction	18
2.2 Data	24
2.2.1 Description	24
2.2.2 Variables and Categories	26
2.2.3 Survival Time Data	31
2.3 Empirical Analysis	33
2.3.1 Summary Statistics	34
2.3.2 Aggregate Worker Flows	39
2.3.3 Logit Modelling	53
2.3.4 Survival Analysis	72
2.4 Concluding Remarks	102
2.A Description of Variables Used in the Analysis	107
2.B Methodology (Survival Analysis)	109
2.B.1 A Simple Hazard Model	110
2.B.2 Proportional Hazard	110
2.B.3 Survival is Continuous but Spell Lengths are Interval Censored	111
2.B.4 Sampling Issues	111
2.B.5 Unobserved Individual Heterogeneity	112

2.B.6	More Heterogeneity: the Split-Population Model	113
2.B.7	The Split-Population Model with Frailty	114
2.B.8	Dual Risks Models	114
2.B.9	Interval Baseline Hazard and Identification of Specific Distributions	119
2.B.10	Summarising the Assumptions	124
3	Diverging Paths: Transition in the Presence of the Informal Sector	126
3.1	Introduction	126
3.2	Background	130
3.2.1	Optimal Speed of Transition Literature	130
3.2.2	Expanding the Seminal Model	143
3.3	A Dynamic Model of Transition with an Informal Sector	146
3.3.1	Main Assumptions	146
3.3.2	Workers	151
3.3.3	The Old Sector	152
3.3.4	The New Sector	152
3.3.5	The Informal Sector	153
3.3.6	Unemployment	155
3.3.7	Wages and Job Creation in the New and Informal Sectors	155
3.3.8	Dynamics	163
3.3.9	Simulations	178
3.3.10	Further Discussion and Policy Implications	190
3.3.11	Departures from Basic Model	193
3.4	Concluding Remarks	196
3.A	New and Informal Sector Mark-ups: Cost of Separation Interpretation	200
3.B	Derivation of Basic Results	201
3.B.1	Wages and Job Creation in the New and Informal Sectors	201
3.B.2	Steady States	203
3.C	Behaviour of Job Creation Functions	207
3.D	The Effect of the Informal Sector on the Timing of Transition	209
4	State Regulations, Job Search and Wage Bargaining: A Study in the Economics of the Informal Sector	211
4.1	Introduction	211
4.2	The Informal Sector: A Glimpse of the Literature	217
4.2.1	Long-run Informality	217
4.2.2	Restrictive Assumptions	219
4.2.3	A Need for Labour Markets	221
4.3	A Model of Informal Employment	222
4.3.1	The Main Idea	223
4.3.2	Matching Technology	225
4.3.3	Formal and Informal Jobs	226

4.3.4	Workers	229
4.3.5	Wage Determination	230
4.3.6	Steady State Equilibria	231
4.3.7	Discussion	248
4.4	Welfare and Policy Making	263
4.4.1	Implications for Policies	263
4.4.2	Implications for Welfare	272
4.5	Concluding Remarks	296
4.A	Proofs, Main Technical Results, and Further Analysis	301
4.A.1	Some Core Elements of the Model	301
4.A.2	Proof of Propositions and Analysis of the Bifurcation Point	304
4.A.3	Welfare and Policy Making	307
5	Conclusions	322
	Bibliography	331

List of Figures

1.1	Simulations of the seminal Aghion and Blanchard (1994) model	6
2.1	Employment rate (% of working age population; source: RLMS 1992-2002)	40
2.2	Unemployment rate (% of labour force; sources: EBRD, 2000, 2003; RLMS 1992-2002)	40
2.3	Inactivity (% of working age population; source: RLMS 1992-2002)	41
2.4	Random draws model	74
2.5	Types in the split-population model	77
2.6	Polynomial baseline interval hazards for the unemployed transiting to employment and inactivity	91
2.7	Polynomial baseline interval hazards for the inactive transiting to employment and unemployment	92
2.8	Piecewise baseline interval hazards for the unemployed transiting to employment and inactivity	93
2.9	Piecewise baseline interval hazards for the inactive transiting to employment and unemployment	94
2.10	Empirical survivor functions for the unemployed and the inactive	122
3.1	Divergence of CEEC and CIS countries: real GDP growth (CEEC including Central and Eastern Europe and the Baltic states, excluding South Eastern Europe and the Balcans; source: EBRD, 2003)	137
3.2	Unemployment in CEEC and CIS (excluding Azerbaijan, Kyrgyz Republic, Tajikistan, and Uzbekistan; sources: EBRD, 1999, 2000, 2003, 2004)	137
3.3	Flows of labour in the model	150
3.4	Simulations: a case of the lagging reformer - the economy converges to the steady state without the new sector	182
3.5	Simulations: a case of the advanced reformer - the economy converges to the steady state without the informal sector	183
3.6	Simulations: behaviour of average wages and transition probabilities	184
3.7	Informal sector in selected European transition countries (source: Johnson <i>et al.</i> , 1997)	186
3.8	Informal sector in selected FSU countries (source: Johnson <i>et al.</i> , 1997) . .	186
3.9	Dynamics of the share of informal economy in GDP of selected transitional countries – estimates obtained by various methods (courtesy of Ed Feige, source: Feige and Urban, 2003)	187

3.10 Simulations: a case of the switching reformer 188

3.11 Timing of transition 210

4.1 Dependence of employment values on model’s parameters 236

4.2 No shadow sector equilibrium 240

4.3 Shadow sector equilibrium 240

4.4 Stable interior equilibrium 241

4.5 Multiple equilibria 241

4.6 Existence conditions for model’s equilibria 245

4.7 The effect of an increase in τ or k_f 290

4.8 The effect of a rise in m or F 291

4.9 Private surplus and interior equilibrium 293

List of Tables

2.1	Initial size of the RLMS samples	27
2.2	Summary statistics for RLMS samples	35
2.3	Summary statistics for RLMS samples (some household characteristics) . .	36
2.4	Accession probabilities for employment (E), unemployment (U) and inactivity (I) in Russia and Estonia	43
2.5	Separation probabilities for employment (E), unemployment (U) and inactivity (I) in Russia, CEE and the US	44
2.6	Churning in employment for the working wage population in Russia and selected CEE states	45
2.7	Relative stagnancy of the two non-employment states in Russia	45
2.8	Coverage by unemployment benefits	45
2.9	Multinomial Logit models of transitions from employment	55
2.10	Multinomial Logit model of transitions from employment	56
2.11	Multinomial Logit models of transitions from non-employment	57
2.12	Two-destination hazard regression models of transitions from unemployment (polynomial hazard)	85
2.13	Two-destination hazard regression models of transitions from inactivity (polynomial hazard)	86
2.14	Two-destination hazard regression models of transitions from unemployment (piecewise constant hazard)	87
2.15	Two-destination hazard regression models of transitions from inactivity (piecewise constant hazard)	88
4.1	Effect of parameter changes in the stable interior equilibrium	289

Acknowledgements

As any piece of research conducted over a long period of time this work, while being a product of my own efforts, still owes a lot to the ideas, suggestions, opinions, comments etc. of a great number of people. After all, while being a yardstick of the scientific maturity of a candidate, any doctorate thesis should incorporate knowledge and achievements of the scientific community to which the candidate belongs and in which she grows to be capable of conducting independent research.

During my work on the essays included in this volume, at different times I benefited from communication with the following people.

The initial idea to work on labour markets in Eastern Europe, and in particular informal employment, dates back to 1996-1997, the years of my involvement in the Economics Education and Research Consortium (EERC-Russia) project with Vladimir Matveenko and Katya Vostroknutova (see Matveenko et al., 1998). At the same time, discussions and hot arguments with Katya about the informal sector and its significance for economic development of my country helped a lot with stating the problems and finding avenues for possible investigation.

The first substantial essay in this volume, an empirical analysis of labour reallocation in Russia, owes much to suggestions by and discussions with Antoni Chawluk, Måns Söderbom, Luis Correia and Dmitry Kulikov. Stephen Jenkins, Tom Mroz, Laura Henderson, and Jurgen Doornik were also extremely helpful with their willingness to reply to my (sometimes numerous) emails.

The second essay on the optimal speed of transition has been developed from my MPhil thesis written here, at the University of Oxford. In the course of my work Richard Jackman, Tito Boeri, Micael Castanheira, Guido Friebel, Carol Scott Leonard, Demid Golikov, and Ksenia Arsentieva helped me a lot either by providing with useful references, recommendations of simulation software, or just by being ready to answer my questions. Participants of the Gorman Workshop, Oxford, in May 2001, and of the seminars at the European Bank for Reconstruction and Development in May 2004, and the Centre for Economic and Financial Research (CEFIR) in Moscow, June 2004, were invaluable by being critical and drawing my attention to potential weak spots of the work. In the EBRD discussion I am grateful for crucial comments to Willem Buiter and Simon Commander. In Moscow Ksenia Yudaeva, Irina Denisova, and Daniel Hammermesh provided especially useful criticism. Apart from that, Ed Feige was very kind to allow me to include a few pictures from his article (Feige and Urban, 2003) into the text of the thesis, while Wendy Carlin helped with ideas for improvement of the chapter for possible future publication.

Finally, the third essay on the labour market approach to formal-informal segmentation of the economy has benefited from comments by participants of Gorman Workshop, Oxford, in May 2002, the Conference on Unofficial Economies in Transition in Zagreb, Croatia in October, 2002, and the Bank of Finland Institute for Economies in Transition - CEFIR Conference on Russian Macroeconomic Issues in Moscow, April 2005. At Gorman Workshop the comments by Purichai Rungcharoenkitkul, Wei Yang Cheong, and John Muellbauer were insightful. Discussions with Klarita Gërxhani, Ed Feige, and Agota Scharle were very helpful in Zagreb. Irina Denisova's comments made a very good point in Moscow.

Many results incorporated in these three essays build upon their earlier versions that have been reviewed and published by the William Davidson Institute, University of Michigan, Ann Arbor, USA (see Bouev, 2001, 2002, 2004, 2005), for which I am very grateful.

Now comes the most important acknowledgement. It should have been placed first but goes here to emphasise a contribution of another person, whose help has been invaluable with putting together and streamlining the ideas and suggestions made at the aforementioned seminars and by a great number of different people. This work would have just been impossible without the fantastic guidance and advice of Margaret Stevens, my supervisor. The confidence and courage with which I am going into the viva owes a lot to her spending hours on reading, commenting, proofreading this text and its earlier versions. She has always been a great person to work with.

A Personal Note

Apart from the acknowledgement of the direct advice and help with concrete issues on the scientific front I would like to thank many other people whose involvement in my life has also made for this thesis having been finally written.

To start with, I express my gratefulness to Alexander Bukhvalov for showing me in the mid 1990s that mathematics is not necessarily purely abstract but can be applied too, in particular, to economics. Then Vladimir Matveenkov's long guidance and encouragement to do independent research was a good school during my years at the European University at St. Petersburg, Department of Economics. Byung-Yeon Kim convinced me that applying to Oxford is not a no-go but rather a must, while Carol Scott Leonard has always been helpful with her advice not only on the academic issues but also on the general rules of survival in Oxford and St. Antony's College in particular. I thank David Hendry, John Quah, Gavin Cameron, and Kevin Roberts for their belief in my knowledge during my MPhil viva in advanced econometrics in 2001, and Margaret Stevens for enlisting Bent Nielsen's help in preparation for that. Looking at the other side of the tuition process, a considerable part of my development in Oxford undoubtedly owes to communication with many undergraduate PPE and E&M students who I taught at Somerville, Exeter, New, and Mansfield Colleges, and the Department of Economics.

I am deeply grateful to N M Rothschild & Sons Ltd. for their role that they have been playing in my life. In particular I would like to thank Sir Derek Thomas and Sir Evelyn de Rothschild for their sagacity, financial (and not only) support and encouragement. The bank and the attitude to life of many people who I met there provided me with a number of wonderful examples urging to look beyond my usual ambits, and created a new kind of personality in my innermost depths.

The St. Antony's College community is gratefully thanked for making the College my home from home, while my colleagues and desk neighbours in the graduate workroom at the Department of Economics in 2003-2005 are acknowledged for being great company.

I thank Alexander Komarov, Mikhail Anissiforov, Vladimir Donskoi, Ales Patrusau, Olga Nekrassova, Andrey Bychkov, Sara Roberts, Anna Ivanova, Demid Golikov, Dmitry Kulikov, Yuri Makassiouk and Oleksiy Omelyanchuk for - apart from everything else, of course - constantly inspiring me to work harder till this piece of acknowledgements. I thank Sara's parents for proofreading and urging me to submit.

Я благодарю Катю за то время, когда мы собирали рекомендации и подавали документы в Оксфорд и другие университеты, за наши планы, которые мы строили, переехав жить в Европу, за ее терпение, которое много для меня значило во время моего приключения под названием MPhil. Кроме того, ты знаешь, что я благодарен тебе за гораздо большее.

I thank Vlad and Sara for their staunch friendship and opening to me new horizons, day after day, in many aspects of life.

Я благодарю Аню за 2003-й год, за имейлы из Перу, и за ту последнюю решимость закончить этот тезис, которая появилась у меня во время общения с ней. Спасибо тебе и за многое другое. Я благодарен Марине и Глебу Ивановым за интересную компанию, которую они нередко мне дарили в мои последние год-полтора в Оксфорде.

Besides, I thank all my friends not mentioned here by name for their support and making my time in Oxford such an incredible adventure!

Наконец, и самое главное. Я благодарен моей семье, моим родителям, Юрию и Елене, за вашу любовь, безоговорочную поддержку, ободрение, утешение в трудные времена, и мужество, с которым вы встретили расставание с сыном, уезжавшим за границу на долгие 6 (а ведь будет даже больше) лет. Хотя туманы Оксфорда и суровые ветры родной Невы и разделяют 2000 километров, я чувствую, что вы всегда рядом. Этот тезис - ваш тоже. Однако, посвящаю его я памяти Петра и Альбины, моих дедушки и бабушки, которые всегда хотели видеть меня закончившим работу над диссертацией, упрочившим традицию, раз установленную в нашей семье.

Июль 2006 года

Chapter 1

Introduction

Transition in formerly communist countries of Central and Eastern Europe and the former Soviet Union is a unique process in many respects. It is a watershed in the lives of many people, who grew up under one economic and political system, learned one system of values, but suddenly and in many ways unexpectedly found themselves needing to cope with a completely new, fledgling environment, new values, a new way of living. Transition is a bridge, supposedly linking two relatively steady lands, two destinations, command and market economies, one of which has already been rejected by history. While the basic set of rules of survival in both lands was or is known, nowadays for many the question is how and what does it take to make the trip across the bridge? Transition is an unmatched opportunity for many social scientists to use this natural experiment to verify their theories, reject or modify them, come up with new insights, and develop new ideas of how to govern this exodus happening in what a conservative American politician has recently alluded to as new Europe,¹ and the vicinity of its borders.

¹This refers to the briefing on the war in Iraq given by Donald H. Rumsfeld, the US Secretary of Defense, at the Foreign Press Center, United States Department of Defense, on Wednesday, 22 January 2003. The transcript is available at http://www.defenselink.mil/transcripts/2003/t01232003_t0122sdfpc.html

Even if one abstracts from the etymological meaning itself of the word “transition”, the allegory of the people’s journey is not accidental here. The first theoretical attempts (Burda, 1993; Aghion and Blanchard, 1994; Blanchard, 1997) to create a model of economic transition in Eastern Europe have implicitly drawn upon the idea of exodus by envisaging the transformation as reallocation of labour. That can be seen from many angles as either changes in people’s labour market status from employment to unemployment and then back to employment, or movement of the workforce between old, inefficient enterprises owned by the state and newly established, efficient firms managed by private owners. In more recent years the topic of resource reallocation has been expanded to include also redeployment of capital (Castanheira and Roland, 2000; Castanheira, 2003) and its focus has shifted somewhat away from movements of the labour force. However, the way of thinking about the transition was strongly influenced by the early models rooted in the search and the so-called flow approaches to labour markets (Blanchard and Diamond, 1992; Pissarides, 2000).

This thesis is concerned with various aspects of labour reallocation in transition. The issues that the study takes up are generally centred on workforce shifts either between different labour market states (employment, unemployment, or inactivity), or between less and more efficient enterprises, or formal and informal sectors. The latter topic, namely the possibility of irregular employment opportunities, receives special attention in this work. According to many recent observations, informal economies have lately grown in size and significance all over the world, but especially in some of the countries of economic transition (see Lackó, 2000; Schneider and Enste, 2000). That growth has been accompanied by an increase in the extent of informal labour markets (Schneider, 2000, 2001). Obviously, this

fact should have a bearing on the reallocation of labour, especially in circumstances of transitional turmoil, but, surprisingly, it has been either underestimated or completely ignored in previous research.

Three independent essays constitute the substantive material of the thesis. The first, empirical, provides a description of workforce movement in Russia between employment, unemployment and inactivity. Among other things, it tries to establish a general pattern of movements, and see if in the main it is consistent with the picture drawn by the basic theory of transition (such as given in, e.g. Blanchard, 1997). Implications for improvement of the theoretical fundamentals are suggested. The second and the third essays draw upon some of these suggestions and are aimed at making a contribution on the theoretical front. The second piece of writing puts forward a revised version of the Aghion and Blanchard (1994) model of labour reallocation in transition, where the informal sector is introduced and plays a crucial role. The model is seen as a modification of the benchmark theory and is used to explain differences in development experienced by, on the one hand, Central and East European (CEE) states, and, on the other hand, by countries of the former USSR such as Russia or Ukraine. Finally, the third essay takes the issue of coexistence of formal and informal sectors in transition further to see if such an equilibrium is possible in the long run, and what role the government can play in (avoiding) such a dichotomy. It should be noted that while the foci of the essays are different and independent from one another, the issues raised therein are closely knit so that many threads can be drawn together.

In the rest of this chapter we provide the reader with a background useful for better understanding of our general motivation to explore various facets of transitional labour

reallocation. The objectives pursued in each of the essays are also outlined. We do not, however, aim to review the extensive literature available on the topic, as separate surveys related to particular problems in focus can be found in further chapters. Instead, the backdrop is set here by introducing the reader to the view on transitional labour reallocation created with the help of the benchmark model by Aghion and Blanchard (1994). We present, rather simplistically, crucial elements of the framework, briefly discuss issues raised by the theory, and point to its implications and predictions. Then we move on to giving short previews of each essay, by cross-linking their topics to the ideas and problems set out in the background section.

1.1 The Way of Thinking About Transition

”The picture of transition drawn by the model is stark, but strikes me as the right one.”

Olivier Blanchard (1997, p.99)

In the first half of the 1990s, great minds of economics, and especially its fledgling branch to be called the economics of transition, saw the two main forces shaping transformation as reallocation - from old regime activities to new, market-type ones, from state firms to new private firms - and restructuring - of existing old regime enterprises. We leave aside the latter question, important as it is for the general success of transition, but reaching beyond the scope of this collection of essays. As regards the former issue - the reallocation, the simplistic picture of transition drawn in many studies (see, *inter alia*, Burda, 1993; Atkeson and Kehoe, 1996; Blanchard, 1997; etc.) was that of the growing new private sector, restructuring of old regime firms owned by the state, and developing unemployment. These developments have prominently featured in the seminal model of

transition by Aghion and Blanchard (1994).

The model describes a closed economy, with two producing sectors and a homogeneous labour force. The essential assumption of a simple version of the model is that the old sector declines at some constant rate γ , while the new sector grows according to the rate at which it creates identical jobs and recruits homogeneous unemployed workers, say p_n . The rate of job creation, p_n , is proportional to firms' profits which depend on new sector wages, w_n . The latter are derived from efficiency wage considerations and, importantly, depend on the level of unemployment, U , so that $p_n = p_n \left(w_n(U) \right) = p_n(U)$. The character of the dependence of p_n on U is not monotonous, however, with p_n increasing in U at low unemployment and decreasing in U when the level of unemployment is high. Then the patterns of workforce reallocation are represented by three basic differential equations:

$$\begin{aligned} \frac{dO}{dt} &= -\gamma O, \\ \frac{dU}{dt} &= \gamma O - p_n(U) U, \\ \frac{dN}{dt} &= p_n(U) U. \end{aligned} \tag{1.1}$$

The first equation describes the development of the sector of old regime enterprises with obsolete capital and management. Over time it is sooner or later bound to restructure, to become more efficient and turn into a new type of business organisation. It can go bankrupt and close too. Whether it is actual restructuring through reformation of old regime practices, or complete closure, it leads to a reduction in the labour force, O , employed in firms of the old type. This reduction happens at rate γ . Workers laid off in the old sector become unemployed and start looking for work. For simplicity, the old

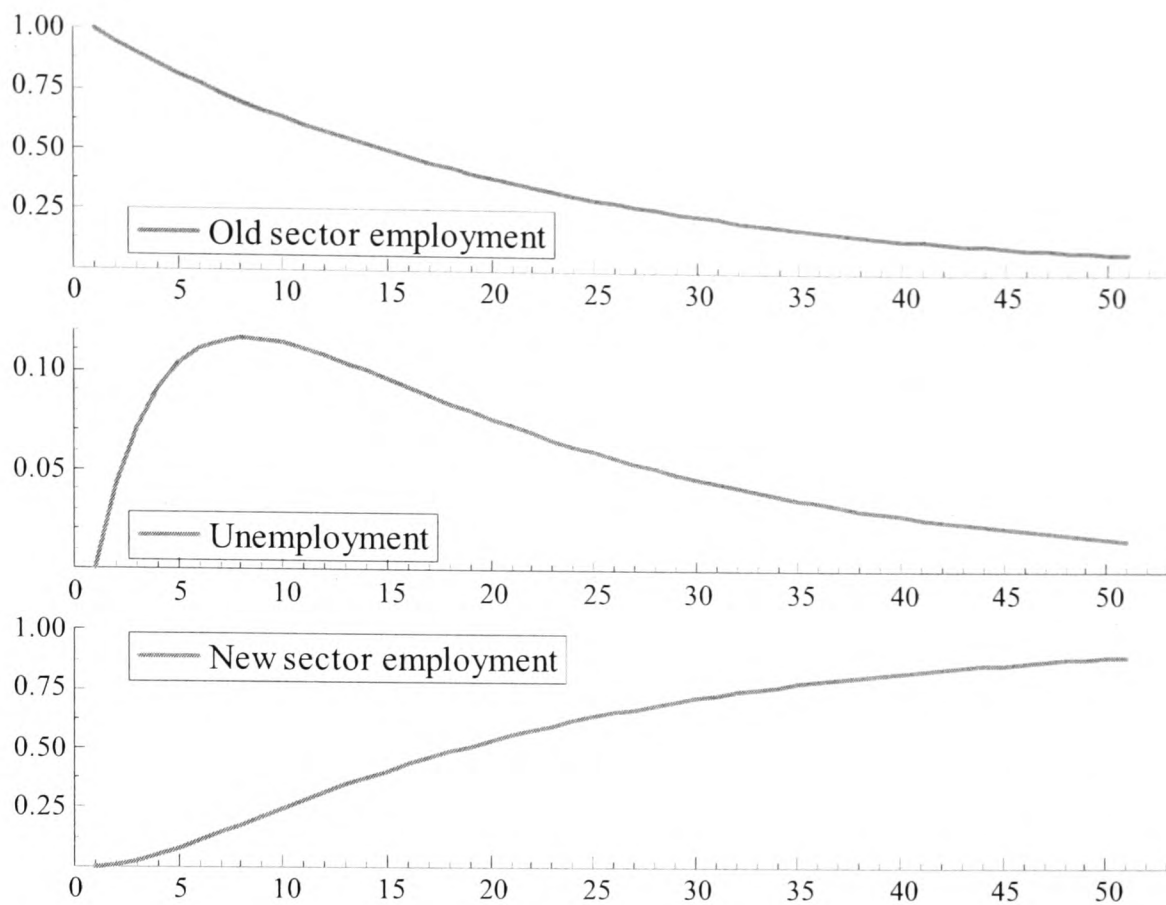


Figure 1.1: Simulations of the seminal Aghion and Blanchard (1994) model

sector in this model only sheds labour, it does not hire it, so the only destination that workers pursue is the growing sector of private initiative, new firms. As hires in the new sector are made at rate p_n , the difference between the inflow from the old sector, γO , and the outflow into the private sector, $p_n U$, drives development of unemployment (second equation). Interestingly, as p_n depends on U , the growth of unemployment due to layoffs in the old sector can be reversed as p_n becomes sufficiently large so that $\gamma O < p_n U$. Finally, in the absence of dismissals in the new sector, the third equation describes the evolution of privately employed labour, N .

Anton Chekhov once² said that "Brevity is the sister of talent", and so it is in using

²Letter to A.P.Chekhov, 11 April 1889. See Anton Chekhov, *Letters on the Short Story, the Drama and other Literary Topics*, selected and edited by Louis S. Friedland (New York: Minton, Balch & Co., 1924), pp.170-80.

mathematics. The simple model by Aghion and Blanchard (1994) had a remarkable success in replicating the initial labour market trends in transition economies at the beginning of the 1990s, and seemed to provide a good basis for making predictions of future developments. Indeed, let us look at Fig.1.1, where numerical simulations of the equations (1.1) are shown.³ The top panel predicts what would happen to the old sector employment - it starts to dwindle away to nothing. The middle panel presents the way unemployment was to develop over the course of transition - the evolution has an inverted-U shape. Finally, the bottom panel suggests what should happen to the new sector - it takes off to absorb from unemployment all the labour shed in the old sector. Generally speaking, this is what transition is about and this is what in many ways was to happen in Central and Eastern Europe as well as in countries of the former Soviet Union.

So far so good. The seminal model created a view of transformation which is seen as a homogeneous process (no heterogeneity of the labour force or firms within sectors is assumed) with unemployment being a transitional phenomenon (it is not persistent, and the spare labour is eventually used up in the new activities). The directions of labour reallocation are, thus, from employment to unemployment, and then back to employment, from old to new types of enterprises.

As transition has advanced beyond the first few years, however, it now becomes obvious that the path it has followed does not concur well with the predictions of the benchmark model. In particular, in reality direct flows between different types of employment have prevailed over the shifts through unemployment, unemployment itself has become persistent and stuck close to double-digit levels, while the overall transition experience has

³The values used for γ and p_n are 0.05 and 0.3, respectively. The horizontal axis measures time in arbitrary units.

been not uniform and varied across different countries, sectors of economy, and groups of individuals. Other aspects of reality, such as a drop in active participation in the labour market, or development of the informal sector, vying for labour with the new officially registered enterprises, have not been covered by the seminal model at all. Those are just a few stylised facts that provide a good motivation for more research on labour reallocation in transition. What are the distinct features of that reallocation? What modification to theory can those features suggest? In what parts the basic theory has to be reconsidered and modified? What are the possible departures? What would the implications for economic policy be?

1.2 Issues Tackled in This Thesis

The collection of three essays in this volume is aimed at making a contribution to the exciting topic of labour reallocation in transitional countries. Each substantive chapter implicitly or explicitly tackles one or more questions raised in the end of the previous subsection. Although not all of the essays can be directly pitted against the basic model, all of them are motivated by weaknesses of that simplistic picture of transition; all of them one way or another present pieces of research into the areas where the stylised theory is feeble. The non-technical reviews below set the stage for each study in relation to the basic view of transition just outlined.

1.2.1 "10 Years of Worker Reallocation in Russia:

An Outlook on Distinctive Patterns and Implications for Theory"

As has been noted above, the benchmark theory gives a view of homogeneous transition with major flows of labour being between employment and unemployment. A number of empirical studies in Central and East European countries have shown that this is not quite the case in reality. Students of the Russian transition have arrived at a somewhat similar conclusion using the data mainly from the first part of the 1990s. In our first essay placed in Chapter 2 of this thesis we expand on previous empirical work by analysing the data from the Russian Longitudinal Monitoring Survey of households covering the first ten years of transition from 1992 to 2002. The set of issues we are primarily interested in clarifying here is as follows.

First, we ask the general question of how different is the Russian reality from the picture drawn by the benchmark theory. In particular, we want to profile the most important patterns of labour reallocation in Russia and see whether they are in accord with theoretical predictions and the experience of other countries in transition. As the data allow analysis of the entire decade, it is also interesting whether the trends established in the Russian labour market at the beginning of the reform period maintain ten years down the transition road.

Second, we aim to learn more about the role of inactivity, i.e. the state of the labour market that describes jobless individuals not actively searching for work. This labour market status is ignored in the Aghion and Blanchard (1994) model, but is probably important for the overall labour reallocation. Moreover, studying inactivity is interesting

because the inactive may well comprise people employed in the informal economy, whose recent growth in transitional countries has already been briefly mentioned above.

Third, the redistribution of labour in transition has probably been non-uniform and affected differently various groups of workers. Establishing whether it is indeed so as applied to the Russian experience is another objective pursued in the first essay. A better insight into the determinants of labour force movement between labour market states can throw light on processes driving the transformation.

Fourth, little previous research has been devoted to the issue of persistence of non-employment both in Eastern Europe and in Russia. At the same time, there is a substantial evidence on growing long-term unemployment in many countries of the region (see, e.g. Boeri, 2000b). Unfortunately, no insights can be drawn from the benchmark model of transition. Instead, on the one hand, equations (1.1) predict that unemployment is bound to disappear as transition advances, which can hardly be reconciled with its increasing stagnancy. On the other hand, for a job-seeker in the Aghion and Blanchard (1994) model the average rate of exit from unemployment, p_n/U , is a decreasing function of the former. Thus, as unemployment in the benchmark model subsides in the course of transition, workers on average spend less time searching for a new job. Clearly, setting out from here there should be no such a problem as long-term unemployment. So, in the first essay in this volume we investigate possible causes of both long-term unemployment and long-term inactivity in the Russian labour market.

Finally, another objective of the essay is to suggest avenues along which the benchmark theory should be improved.

The study of worker mobility in the essay consists of three elements. The first one

involves examination of average transition probabilities between labour market states over year-long periods in order to identify the most significant flows across employment, unemployment and inactivity. In the second part multinomial logit models are used to see how these patterns change with various personal and household characteristics as well as general economic conditions. Finally, a set of recent econometric techniques not previously applied to the Russian labour market is used in duration analysis of non-employment.

Results of the first essay single out the directions in which further investigation will be particularly welcome, both as empirical efforts aimed at supplementing our knowledge of different aspects of labour reallocation and as endeavours either strengthening the basic theory or developing new departures. We maintain that heterogeneity of the labour force, job-to-job shifts and circulation of labour between unemployment and inactivity, and the presence of irregular employment opportunities are important characteristics of transition. A better theory of labour reallocation should be able to take into account at least some of these issues.

1.2.2 "Diverging Paths:

Transition in the Presence of the Informal Sector"

In the second, theoretical, essay we incorporate into the benchmark model of transition some of the suggestions put forward in the first study. In particular, in Chapter 3 we deal with the question of the importance of the informal economy, while explaining the absence of uniformity of transition paths across different groups of countries. Generally, the essay is concerned with the following three problems.

One issue is a diverging experience of transition in Central and East European coun-

tries (CEEC), on the one hand, and in the Commonwealth of the Independent States (CIS), on the other hand. It is interesting that while the two regions started the transformation process in broadly similar initial conditions, CEEC have made a better progress towards restructuring of the state sector and have been quicker than the CIS to create an environment favourable for the development of the private sector. At the same time, development of unemployment and evolution of output in the two groups of countries have followed two different trajectories especially distinct in the first half of the 1990s. Although a number of explanations have been put forward in the literature, the theoretical studies of transition spawned by the Aghion and Blanchard (1994) paper have been unable to provide a convincing rationale behind many noticeable dissimilarities in focus. This has motivated us to suggest another vision of the problem.

We seek explanation in the presence of the informal economy and its effects on the development of the formal counterpart. While the benchmark model of transition considers only two sectors, namely the formal state and the formal private, the third, informal sector has become a crucial feature of transformation in many countries in Eastern Europe and the former Soviet Union. Although the informal economy is also a domain of private economic initiative, it is effectively different from the new private sector in the Aghion and Blanchard (1994) model, as it evades governmental regulations, especially taxation discussed in that work. The significant extent of the informal economy in Russia has been confirmed in many recent empirical studies. That fact as well as a crucial role played by inactivity in the reallocation process (the finding of our first essay) have urged us to investigate how informality affects the course of transition envisaged by the basic theory.

The direct impact of informality is that informal enterprises often compete for resources

with formal firms. At the same time, the presence of the informal sector may create externalities which indirectly affect the viability of the formal private sector. From a theoretical point of view, the heart of both effects can be seen in the rates of job creation in the formal and informal sectors. The importance of creation of conditions favourable for the development of the private sector has been mentioned in many studies. However, even in two-sector theoretical models of transition little attention has been paid to the investigation of this issue. Instead, in developing policy recommendations followers of Philippe Aghion and Olivier Blanchard have mainly paid heed to the effects that the government may have on the rate of decline of the old sector, represented in equations (1.1) by parameter γ . A special debate has been opened about the choice of optimal γ , while an impact on the new sector job creation rate, p_n , has hardly been discussed. In our second essay we aim to demonstrate that on the policy-making front more attention should be given to the development of proper policies towards new firms, especially when the presence of the informal economy is taken into account.

Thus, to close the gap in theory, we build a model of transition in the presence of the informal sector, where reallocation of labour takes place from old sector employment to unemployment, between unemployment and new sector employment (in both directions), and between unemployment and irregular (informal) employment (also in both directions). It is shown that qualitatively different outcomes of transformation depend on a stance adopted by the government towards the new sector. In particular, we demonstrate that transition, as the reallocation of resources to formal firms of a profit-oriented type, can fail if the pressure of various factors on new formal firms outweighs the pressure on the informal sector. In such a case it is the informal firms that absorb faster the labour shed by old

regime enterprises, and that eventually dominate the post-transition economy. We argue that the convergence to qualitatively different equilibrium destinations helps explain the varying experience of CEEC and countries of the Commonwealth of Independent States.

In the model that we discuss in Chapter 3, the major flow of labour reallocation goes either one way - from the old sector to the new formal one, or the other - from the old sector to the informal sector. Although unemployed *workers*, randomly searching for a new job, can end up in either a new formal or an informal firm, we leave aside the question of *entrepreneurs* choosing a sector in which it is more profitable to open a new vacancy. For simplicity the development of new formal and informal economies is seen as being driven only by profits in respective sectors reinvested in the same sectors. It is demonstrated that the co-existence of the formal and informal economies is likely to be a transitional phenomenon, while the presence of both in the eventual steady state is highly unlikely. For that co-existence to take place in the end of transition a number of different conditions have to precisely counterbalance each other (see the so-called knife-edge condition in that chapter) which is hardly realistic. Instead, the model suggests that it is more probable that in the end one sector should crowd out another. We take this issue of co-existence further in the third essay which focuses solely on the dynamic distribution of labour between the formal and informal sectors.

1.2.3 "State Regulations, Job Search and Wage Bargaining:

A Study in the Economics of the Informal Sector"

In the seminal paper on informal economies in transition, Johnson *et al.* (1997, p.161) note that

"Our work can be thought of as complementing Olivier Blanchard's (1997) analysis of transition economies, which highlights the creation of new private firms by entrepreneurs as the engine of growth. We focus on the political and institutional determinants of the entrepreneurial response, and in particular, on the allocation of resources between the official and the unofficial sectors."

It is hardly possible to think of a better quote to suit to the third essay presented in Chapter 4 of this volume. Here we generally take up the following challenges.

First, the third essay is aimed at completing a picture of labour reallocation in transition by looking into its distribution between formal and informal firms alone, while leaving aside the question of reallocation from the old to the new sector. This aspect of reallocation is under-represented in our second essay reviewed above, and is completely overlooked in the basic theory of Aghion and Blanchard (1994).

Second, we ask the question of whether the co-existence of formal and informal firms can be long-term, i.e. whether or not it is a transitional phenomenon as the second essay suggests. The previous theoretical modelling provides no satisfactory rationale behind long-run mixed equilibria with both formality and informality. Thus, this issue is investigated here with the help of a relatively simple model which avoids many restrictive assumptions made in other studies of the informal sector.

Third, in studying formal-informal segmentation we want to provide a better account of the interaction of firms and workers in the labour market. Surprisingly, all other work carried out on the same topic has paid nearly no attention to that aspect of segmentation, with labour markets often modelled as perfectly competitive and always clearing.

Fourth, the third essay looks into the problem of whether wages can be a channel transmitting effects of various governmental regulations and leading to the labour market split into the formal and informal parts. It is widely held that one of the best-known

effects of the informalisation process in the world in general, and transition economies in particular, is to reduce the burden of governmental regulations and labour costs. However, the issue of wage bargaining, as a mechanism through which preconditions for formal-informal duality emerge, has never been addressed.

Fifth, in the third essay we want to elaborate on some issues of labour market imperfections, also present in the second essay that may provide incentives to hide businesses in the shadow economy. In particular, the model in Chapter 3 incorporates possibilities of rent sharing between workers and firms which affect job creation rates in the new formal and informal sectors. However, the rent sharing aspect is not modelled explicitly there, neither do we look in the second essay at the actual conditions that affect firms' and workers' choice in wage bargaining. Chapter 4 should close this gap in theory.

Finally, in Chapter 4 we also aim to provide welfare implications of various governmental policies aimed at a reduction in informality. At the same time we ask the question of whether the elimination of the informal sector is actually an objective that governments should be concerned with in the first place when dealing with segmented labour markets.

These challenges are met with the help of a standard search-in-the-labour-market model à la Pissarides (2000) with two sectors. It demonstrates that a long-run equilibrium with both formal and informal economies is possible under very mild assumptions when labour markets are characterised by wage bargaining and time consuming search. It is shown that imperfections in the labour market can create a situation when reduction in informality may be detrimental to economic welfare. In general, the model provides a theoretical framework for studying reallocation of jobs and workers between new formal and informal businesses, and it is much more attentive to mechanisms of interaction

between firms and workers in the labour market than all the preceding work on the topic. It, thus, addresses a facet of the reform process which is missing both in the benchmark model of transition and in its some more elaborate versions. Therefore, it presents an important basis both for making policy recommendations, and for possible modelling of transitional redistribution of labour.

Martin Raiser, one of the well-known students of economic transformation in Eastern Europe, has preceded one of his articles⁴ by the paraphrased Buddhist wisdom that reads "Transition is a bridge, therefore do not dwell upon it." Every introduction is a bridge too. So it is time to move on to the substantive contents.

⁴See Raiser (1995).

Chapter 2

10 Years of Worker Reallocation in Transitional Russia: An Outlook on Distinctive Patterns and Implications for Theory

2.1 Introduction

Reallocation of labour in transition is a *sine qua non* for the overall success of reforms (Vodopivec, 2002). The legacy of the communism era was the labour force predominantly concentrated in low-productive state enterprises. Thus, one of the main objectives of economic transformation has been to redeploy this mass of workers to more effective businesses, either emerging through a restructuring of old regime establishments or being created *de novo*. Not surprisingly, the topic of labour reallocation has attracted intense

attention both from theoretical and empirical points of view. Theoretical studies focusing on the issue were spawned by the early papers by Burda (1993), and Aghion and Blanchard (1994). They drew a homogeneous picture of transition and attached special importance to growing unemployment as a supply of labour to be tapped into by the businesses of a new, market economy, type. However, empirical studies of Central and East European countries (CEEC), proliferating since the beginning of the 1990s, have shown that reality is more complicated than the basic theory has predicted: significant unemployment has been accompanied by worker flows out of the labour force and contributed to growing inactivity (Boeri, 2000b); newly created firms have mainly poached workers from old state enterprises rather than hired them from the non-employment pool, so that a direct job-to-job movement prevailed (*ibid*); unemployment has become persistent and concentrated in particular categories of workers (e.g. low-skilled and younger workers have often been found at a particular disadvantage as compared to other groups - see Keune, 1997; Sorm and Terrell, 2000; Commander *et al.*, 2004). All these facts present new challenges to theorists of transition. At the same time, other empirical work focusing on the members of the Commonwealth of Independent States (CIS) has demonstrated that the former Soviet Union (FSU) republics can raise a different set of questions to be answered. In this respect Russia has received a good deal of consideration.

One of the most important differences between Russian and East European experience was the adjustment of employment and wages. Employment-to-output elasticities in Russia were negligible compared both to Western and CEE countries (Boeri, 2000b; Boeri and Terrell, 2002), while the wage decline was unprecedented and widely resulted in outright arrears. Official unemployment figures were low, although ILO-type unemployment in

Russia was more significant. At the same time, participation rates, very high by Western standards at the beginning of transition in all countries affected, have displayed in Russia a relatively modest fall compared to CEE.

Despite lower responsiveness of employment to output changes, however, the Russian labour market has been found to be very fluid. A few previous studies of labour mobility (*inter alia* Foley, 1997b; Lehmann and Wadsworth, 2000; Grogan, 2003) have demonstrated high degrees of reallocation along several dimensions (job-to-job shifts, tenure profiles, etc.). So, what are the most noticeable patterns in labour force reallocation in Russia? Do they resemble those in other countries of Eastern Europe? How well do they concur with theoretical predictions?

In this work we supplement previous empirical studies and portray the most important patterns of labour reallocation in Russia over the first decade of transition. Better understanding of determinants of labour force movement between labour market states can cast new light on processes driving the transformation and suggest modifications to theory. To this end we employ data from the Russian Longitudinal Monitoring Survey of households from years 1992-2002. Our attention is concentrated on labour mobility to and from two labour market states: unemployment and the out-of-the-labour-force state (inactivity). Such interest is motivated by, on the one hand, the predictions of early theoretical models of transition (see, for example, Aghion and Blanchard, 1994) that it is the pool of the unemployed that is the main destination for workers leaving employment, and is the main source of labour for firms setting up new businesses. On the other hand, an increase in the size of the informal economy in transitional countries (Johnson *et al.*, 1997; Schneider and Enste, 2000), including subsistence activities reportedly performed by the inactive,

calls for a careful analysis of those out of the labour force.

Before we proceed, it should be mentioned, however, that from the theoretical perspective of economic transformation another dimension of labour mobility is of great interest too, namely transitions between different types of employment. In particular, movements between state and private enterprises, or less and more productive firms are important to understand. Still, an attempt to put those labour force flows within the scope of research is not made here. The main reason for that is that the information contained in the RLMS data set simply does not allow a proper insight into determinants of labour demand governing transitions to and from employment. For that purpose a survey of enterprises would suit much better - for recent examples of such studies focusing on Russia see, for instance, Brown and Earle (2002, 2003). As regards the transitions to and from the state sector, the information on types of enterprise ownership provided in the RLMS is misleading, especially so with respect to participation of the state in the enterprise's stock (on which we briefly comment in the sequel). Thus, any results based on that information should not be treated as reliable.

The analysis of worker mobility in this essay is split into three major parts. The first one involves examination of average transition probabilities in order to identify the most significant flows across the three labour market states - employment, unemployment and inactivity. In the second part we employ a multinomial logit approach to see how these patterns vary with demographic characteristics, educational attainment and general economic conditions. This should address the question of uniformity of flows across the working age population. Finally, our investigation is completed with a survival analysis to understand more deeply the determinants of non-employment duration. That helps

to gain more insight into reallocation through a crucial, from a theoretical point of view, state in the labour market - non-employment. Wherever possible, the results are compared with similar research on East European countries. Implications are drawn for theoretical modelling of labour supply, job search and labour mobility in transition.

In part, our research confirms findings of other studies of the Russian labour market which draw upon the data from the first half of the 1990s. However, this work still makes a contribution in many ways. First, it is one of the first attempts to analyse an entire decade of the Russian transition covered by the RLMS data, including the information recently made public. Second, the treatment of the data is in many respects more scrupulous than in previous studies and allows us, among other things, to differentiate more appropriately between the ILO-unemployed and the inactive. Third, in our analysis we pay special attention to the flows between the two non-employment states and their determinants. The importance of such reallocation was surprisingly overlooked in previous research which either misclassifies the unemployed (Foley, 1997b), leaves the inactive out (Smirnova, 2003b) or merges them with the pool of unemployed (Lehmann and Wadsworth, 2000; Smirnova, 2003a). Fourth, in an attempt to clarify the most important issues of labour reallocation in Russia we draw upon numerous other studies of different aspects of country's labour market transition. Fifth, in the investigation of non-employment duration we use a set of relatively recent techniques that have not previously been applied to the analysis of the Russian labour market. In particular, we use a competing risks framework with two sources of unobserved heterogeneity (a Gamma-distributed "frailty" factor, see Jenkins, 2004, and the Heckman-Singer, 1984, correction) to test two models, recently applied to the sclerotic Portuguese labour market (Addison and Portugal, 2003). On that basis we

draw useful insights about both the long-term unemployed and the long-term inactive, and are able to reject the applicability of certain theoretical aspects of search in the labour market (e.g. Blanchard and Diamond, 1994) to the Russian case. Finally, wherever possible we think of our findings in relation to the basic theory of transition (Aghion and Blanchard, 1994; Blanchard, 1997) and suggest a set of issues that should be addressed in future theoretical work, aiming at modification of the fundamentals.

The organisation of this chapter is as follows. In the next section we present a description of the data and construction of the samples. Section 2.3 contains the empirical analysis. In order to give a provisional perspective on uniformity of labour force reallocation in Russia we begin by presenting allocator tables which provide the reader with preliminary evidence of some selection effects into labour market states. Then the analysis of average transition probabilities over yearly periods in Section 2.3.2 gives an idea of the most important worker flows. These two exercises suggest further questions of particular interest in relation to labour force reallocation that we raise in Section 2.3.2. Multinomial logit models are used in Section 2.3.3 to address the first two of those questions, in particular, concerned with non-uniformity of worker shifts across labour market states and their determinants. The findings and some resulting implications of this analysis for theory are given in Section 2.3.3. Section 2.3.4 contains duration analysis of non-employment which is meant to deal with the third, final question, of non-employment stagnancy. Sections 2.3.4 and 2.3.4 discuss related findings and draw some lessons for future research. Section 2.4 concludes and outlines the main challenges for further investigation of labour force reallocation in transition. Appendices contain a detailed discussion of the methodology used in duration analysis, as well as a description of logit variables.

2.2 Data

2.2.1 Description

The main data source is the Russian Longitudinal Monitoring Survey (RLMS). It is a series of nationally representative household-based surveys administered by the Russian Institute of Nutrition, Carolina Population Center at the University of North Carolina at Chapel Hill, and the Institute of Sociology of the Russian Academy of Sciences. The RLMS is designed to measure the effects of Russian reforms on the economic well-being of households and individuals. It contains *inter alia* a mass of information on individual and household incomes, welfare, consumption, labour market position, etc.

The data have been regularly collected since 1992. The project has been split in two phases. Phase I consists of 4 rounds (I - IV) and contains data from 1992 to 1994. The main purpose of the rounds was to develop sampling methods, in particular, to build institutions such as the first national sample frame, allowing surveys to be representative at the national level.

In the ongoing Phase II (rounds starting from Round V) covering the period since 1994 the emphasis of the project changed to providing timely, high-quality information. The sample in the second phase is generally smaller, but the number of primary sampling units, i.e. geographical districts covered by the survey, was doubled to enhance the representativeness of the data. The response rate among the initially drawn sample of households and individuals is good enough by international standards and excellent by standards applicable to both Eastern and Western Europe. On average in rounds used in our analysis the household response rate exceeded 80%, while the individual response rate hovered above 90%.

Data from the RLMS may be used in two types of analysis - a repeated cross-section analysis, and a longitudinal or "panel" analysis. As regards the possibility for the latter, the RLMS is a sample of dwelling units, or addresses: in each wave the same households or/and individuals are surveyed unless they move away, in which case they are not followed. Consequently, the RLMS is not a true panel design, as it links only those households and individuals who remain in the original dwelling unit over time. Such a "panel" may be vulnerable to selection bias when reasons for moving are correlated with the dependent variable of interest. Crude sample attrition rates between rounds are in the range of about 7 to 20%. A further detailed description of the sample can be found in RLMS (2004).

The RLMS data have been used many times in work on Russian transition, including a number of studies of the Russian labour market and labour supply, in particular (Foley, 1997a,b,c; Lehmann and Wadsworth, 1999; Sabirianova, 2002; Grogan, 2003; Smirnova, 2003a,b; to name just a few). However, a host of available information and the ongoing character of the survey still leaves room for further contributions.

For our research we employ both individual and household data available from the RLMS. We make use of 9 rounds for descriptive statistics (corresponding to years 1992-2002), construct five panels using the data from successive years¹ 1992-1993, 1994-1995, 1995-1996, 2000-2001, and 2001-2002 for multinomial logit analysis of labour market dynamics, and, finally, create a stock sample with follow-up (see, e.g. Jenkins, 2004) for analysis of non-employment duration using the data from rounds² V to XI (corresponding

¹The choice of the years is explained by construction peculiarities of the RLMS sample: the data from its Phase I (rounds I-IV) cannot be merged with the data from Phase II (rounds V-XI). Another reason comes from one of the goals pursued in this work. In particular, among other things we want to analyse the dynamics of the labour market over approximately year-long periods. We therefore do not pair up rounds VII and VIII, and VIII and IX as the calendar gaps between them are roughly 2 years each.

²The choice of the years for the duration sample is explained by the following considerations. Observations in Phase II samples cannot be matched to observations in Phase I samples, so the follow-up on non-employed individuals is possible only within rounds I to IV, and, separately, V to XI. At the same

to years 1994-2002). The initial size of samples with observations of individual and household variables is given in table 2.1. For the purposes of our study we match information available for individuals and describing their labour market position to that supplied by heads of households and pertaining to certain aspects of well-being of their families. We focus our analysis on the individuals of working age which in Russia officially comprises men aged 18 to 59 and women aged 18 to 54 (see, e.g. Brainerd, 1998). Thus, after matching the household and individual data and applying age restrictions we obtain samples of 8838 (round I), 7785 (round III), 5915 (round V), 5570 (round VI), 5492 (round VII), 5689 (round VIII), 5985 (round IX), 6692 (round X) and 6945 (round XI) individuals, respectively. Merging the data from rounds I and III, V and VI, VI and VII, IX and X, and X and XI yields five panels with 7098, 4502, 4471, 5141, and 5764 observations, respectively (age restrictions apply as of the basis year). Finally, by tracing the duration of non-employment for individuals represented in rounds V-XI, picking only their first observed non-employment spell (in the case of multiple spells of non-employment), and after correcting for the initial conditions problem (see, e.g. Lancaster, 1990), two samples are obtained of 1917 and 2147 observations for unemployed and inactive individuals, respectively (age restrictions apply as of the moment of first being observed as non-employed).

2.2.2 Variables and Categories

For each individual in our samples we construct variables reflecting their demographic characteristics (such as age, sex, marital status, the number of children in the household,

time, Phase I rounds span only two years - a time period that may not be long enough to identify causes of long-term non-employment, that we focus on in Section 2.3.4. Thus, we do not make use of rounds I to IV in the duration analysis.

Table 2.1: Initial size of the RLMS samples

Round	Number of the surveyed		Period covered
	Individuals	Households	
I	16641	6334	July - November 1992
III	15037	5836	March - November 1993
V	11284	3973	November 1994 - January 1995
VI	10648	3781	October - December 1995
VII	10465	3750	October - December 1996
VIII	8701	3831	October 1998 - January 1999
IX	9074	4006	September - December 2000
X	10098	4528	September - December 2001
XI	10499	4668	September - December 2002

the number of post-work age persons in the family, education (skill level), etc.), economic well-being of their households (such as household income, access to land, ownership of the *dacha*, i.e. a country house with a land plot), their position in the labour market (labour force participation, employment status, duration of non-employment if jobless or tenure in current job if working, involvement in individual economic activity, secondary job holding), and some characteristics of the local labour market conditions (experience of wage arrears, unpaid compulsory leaves, access to non-wage social benefits, unemployment and non-employment rates). The description of constructed variables is relegated to Appendix 2.A. Several points, however, deserve further attention at this stage as they turn out to be quite important for the following analysis.

Definition of the labour market states

All individuals in our sample are assigned to one of three possible labour market states: employment, unemployment and inactivity (or out of the labour force).

Employment The RLMS allows one to measure the involvement of an individual in the production process along several dimensions. First, for each respondent we identify

whether they work in an enterprise or an organisation, are engaged in entrepreneurial activity or are involved in individual economic activity³ (IEA) on a regular basis (these three activities may be combined and are not mutually exclusive as the reading of RLMS questionnaires goes).

Using the data from rounds I-IV we define entrepreneurs as those who are engaged in entrepreneurial activities and are continuously involved in the affairs of their enterprise. In rounds V-XI entrepreneurs are defined as either those who own more than 50% of an enterprise or those who work at an enterprise, have a stake in it, and describe their main activity as being of an entrepreneurial nature.

The regular involvement in IEA is straightforward to work out from the data in rounds V-XI (the respondents were directly asked to characterise their individual employment as incidental or regular) but less so in rounds I-IV. In Phase I we assign regular IEA to all those who either worked last month at least 120 hours (i.e. about 30 hours a week) or worked at least 9 out of last 12 months, or indicated that their main occupation is IEA.

Having done that we split all individuals between "working" (who belong to any one of the aforementioned categories) and "non-working".

Unemployment and inactivity The split of the non-employed between the unemployed and the inactive is important for the following analysis of the labour market dynamics. One way to proceed is to use the RLMS question about the main occupation of respondents, where they could indicate that they *consider* themselves unemployed and seeking a job (this avenue was followed by, e.g. Foley, 1997a,b). Another way is to use

³By the entrepreneurial activities the RLMS questionnaires understand owners or co-owners of businesses, companies, cooperatives, stores, farms, etc. The IEA is, effectively, an alias for self-employment - that is, the activities such as private cab driving, private tailoring, marketing some of the produce grown on personal land plots, etc.

other information available in the RLMS on the desire to find work and concrete actions (such as applications to employment agencies, employers, asking friends or relatives about possible job openings, etc.) taken to this end in the last 30 days prior to the interview date. The group of the unemployed conflated according to this criterion would be closer to the ILO-OECD definition of unemployment (this route was taken, e.g. by Smirnova, 2003a,b).

We have found that the method used by Foley (1997a,b) leads to the assignment of individuals to wrong labour market states. We estimate that on average about 40% of those who reported their main occupation as temporarily unemployed and looking for a job, had not, in fact looked for a job (i.e. had not taken concrete actions) in a month preceding the survey in all rounds I-XI. Thus, those people are in fact passive unemployed by the ILO-OECD classification and should be assigned to the out-of-the-labour-force state. This is done in the analysis that follows.

Definition of the skill levels

In this work skill levels are defined on the basis of information on the educational attainment of respondents.⁴ In the two phases of the RLMS questions about the level of education vary across rounds which generates some difficulties for the comparison of certain educational categories (Turunen, 2004). To avoid these comparability problems we reclassify available information into three educational groups according to ISCED-97 (UNESCO, 1998). The three categories are defined as follows. Category 1 includes

⁴Foley (1997b), Turunen (2004) and some other students of the RLMS have also used additional variables to capture skill levels of respondents and their effect on interstate mobility. For example, data on occupational codes have been employed. Here we do not follow these avenues, however, because transition in Russia has been noticeable for the significant occupational reallocation with unconventional career switches (Sabirianova, 2002).

low skilled respondents (persons with lower education, defined as ISCED-1997 levels 0-2); category 2 is for skilled respondents (people with education at ISCED-1997 levels between 2 and 5); finally, category 3 encompasses high skilled respondents (i.e. those with higher education, defined as ISCED-1997 levels 5-6 (levels 5-7 by ISCED-1976). More information on educational levels and variables further used in econometric analysis is provided in Appendix 2.A.

Land use and dacha ownership

In addition to a customary set of demographic and other characteristics we construct two dummies for land use over the past year and *dacha* ownership in the respondent's family/household. Our motivation for including these two variables comes from the fact that in many countries, successors of the former Soviet Union, agricultural and subsistence activities often play an important and dominant role within the array of informal activities⁵ (see, e.g. Yoon *et al.*, 2003). It is also widely known that *dachas* (or country houses with land plots) are an important element in subsistence farming and survival strategies of Russian households (see, e.g. Clarke *et al.* 2000; Clarke, 2000a,b). Thus, one could expect that access to land may affect the decision to participate in the formal labour market, and thus may well affect movements of individuals across labour market states. For each individual in our samples we indicate if their household has in possession a country house or/and has had access to land over the 12 month period preceding the date of the interview. These two variables along with the dummy for engagement in

⁵EBRD (2000) points out that the importance of the agricultural sector has grown over the transition in many CIS countries. Yoon *et al.* (2003) estimate that in countries such as Armenia, Azerbaijan, Georgia, Kyrgyzstan, Moldova, Tajikistan, and Uzbekistan, about three quarters of all informal activity is in agriculture. Voicu (2002) points to important links between the private agricultural sector and non-participation in Romania.

individual economic activity should provide a set of indicators of involvement with the informal market.

2.2.3 Survival Time Data

When investigating transitions from the two non-employment states, with the help of both multinomial logit and survival time analyses in Sections 2.3.3 and 2.3.4, respectively, we use RLMS information on the elapsed duration of non-employment for jobless individuals. This issue is worth a detailed explanation as treatment of data on non-employment duration may bear upon the implications of the study.

In the case of the multinomial logit analysis a set of indicator variables is introduced to control for the elapsed duration of joblessness. In that part of the research the focus, in general, is characteristics affecting an average year-to-year probability of changing current labour market position. The aim is to determine which broad categories of workers are more likely to leave a non-employment state over the year's course, but not the specific dependence of the probability of transition in the next, reasonably short, period of time on the time already spent jobless. With the former objective in mind, we include in the sample for logit analysis all the non-employed individuals of working age regardless of the length of their jobless spell - what is of primary interest is not when precisely did the spell start, but how long it is.

By contrast, in the survival analysis the moment that the spell of non-employment starts is of crucial importance. As has been mentioned above, we have constructed two stock samples with follow-up. This means that when a non-employed individual is observed she has already been in that labour market state for some time, the length of which

is reported. Then this individual is followed up until she moves out of the state or drops out of observation (i.e. her spell is censored⁶). Such stock sampling is known to suffer from the length bias problem. In particular, the distribution of survival times could not be appropriately inferred from the elapsed duration information, because: a) rates at which people enter the non-employment state in question may have been different at different moments of time in the past, and b) if the reinterview period is long enough (in our case it sometimes reaches nearly 24 months, depending on the RLMS rounds considered), the investigator is bound to deal only with relatively long durations, while short completed durations drop out of observation (as people with short spells of non-employment will have entered and left the state between the interview dates). When constructing the likelihood function, the latter problem is corrected by conditioning on the elapsed duration times - we shall return to this in Appendix 2.B that contains information on the methodology used in Section 2.3.4. The former problem is more complicated and is known as the initial conditions problem (see Lancaster, 1990, p.189). For its fully satisfactory solution one needs to know a stochastic process describing the full state biography of each individual. A partial solution to the problem is obtained if one includes in the sample only spells commencing after the first sample date. In the absence of person-specific heterogeneity such treatment leads to a sensible and correct way to proceed (*ibid*). When constructing our duration samples we, thus, include only spells starting after 1 November 1994, i.e. a month when Phase II surveys of RLMS were first conducted. Among the spells that were left out are those initiated before the start of economic transition in Russia, and those that began during the first wave of mass privatisation (October, 1992 - June, 1994), i.e.

⁶For successive and interrupting non-employment spells for the same individual only the first spell is taken to avoid person-specific serial correlation (see Foley, 1997a, for an analogous approach to sample construction).

the periods not covered by the RLMS Phase II surveys.

As the information on the precise number of days spent non-employed is not available for all individuals sampled by the RLMS, the length of non-employment is determined up to a month. Thus, a minimum duration for any individual observed without work is 1 month. In the rest of the treatment of spells and time gaps between interviews we follow Foley (1997a), Kartseva (2002), and, especially, Grogan and Van den Berg (1999). Findings of the latter work suggest that using their approach should not impair conclusions of the duration analysis of the RLMS data.

2.3 Empirical Analysis

This section presents the main findings of the empirical analysis of the data, the main points in treatment of which have been emphasised above. We start our analysis by looking into the static distribution of the Russian working age population across the three labour market states. By presenting summary statistics of the samples we try to see tentatively if selection into employment, unemployment and inactivity has been uniform during the course of transition. Having done that we move on to establishing a general pattern of transitions between employment and non-employment⁷ with the help of average

⁷It should be noted here that an analysis of transitions between sectors of different types of ownership would be another important exercise. Such reallocation is the main focus of attention in the theoretical literature. In this regard the RLMS does contain information on the type of ownership of respondents' employers. That information has already been used in research (see e.g. Foley, 1997b; EBRD, 2000; Turunen, 2000, 2004) on labour market flows - in particular, for the analysis of worker transitions into and out of state and non-state sectors. However, we are very sceptical about any attempts to use this information for the analysis of the dynamics of the Russian labour market for the following reason. Enterprise ownership is self-assessed by RLMS interviewees, and our examination of the data suggests that their answers are highly unreliable. In particular, the average of 60 to 80% (depending on a round) of individuals that did not change a place of work between the dates of the first and second interviews indicated that participation of the government in their enterprise had changed, so that effectively the type of enterprise ownership had also changed. Obviously, part of this is explained by nameplate changes accompanying privatisation. However, the pattern is also true for those individuals that reported employment in a non-state enterprise on the date of the first interview and employment in a state enterprise on the

separation and accession probabilities over year-long periods. Judging by those we find that the Russian labour market displays a remarkable fluidity compared to CEE countries. Among other things, it turns out that high degrees of job-to-job movements, as well as of shifts between unemployment and inactivity are especially noticeable. At the same time, both unemployment and inactivity have become more persistent labour market states. These observations either run counter to the theoretical predictions or are offered no explanation by theorists. Thus, we raise a set of questions to be further investigated here, and which may be helpful in better understanding of the Russian transformation. These questions are as follows. First, whether movements in and out of inactivity as well as other labour force flows are uniform or typical of particular groups of individuals? Second, what are the determinants of shifts in and out of inactivity? Finally, what can cause the stagnancy of the non-employment states? The multinomial logit modelling in Section 2.3.3 addresses the first two questions, while survival analysis in Section 2.3.4 tackles the third one.

2.3.1 Summary Statistics

Survey sample statistics are presented in table 2.2. It is easy to see that the non-employed display a set of characteristics clearly distinguishing them from the employed: the former are younger (while the average age is smaller for the non-employed, across all rounds about 30% or more of the jobless are people younger than 25, as compared to some

date of the second interview. In the absence of mass nationalisation practices these nameplate changes can only be explained by the majority of RLMS respondents being ill-informed or confused about the type of ownership of their employer.

Table 2.2: Summary statistics for RLMS samples

Variable	1992 (round I)			1993 (round III)			1994/95 (round V)		
	Total	E	U	Total	E	U	Total	E	I
Rel. sample size, %	100.0	84.0	4.7	100.0	83.2	4.1	100.0	77.1	12.7
Average age	37.2	37.7	32.7	37.5	38.2	32.2	36.3	37.1	34.5
Age < 25, %	14.7	10.7	33.5	14.4	10.9	31.7	17.7	13.1	32.2
Age > 50, %	15.7	15.3	11.3	15.8	15.3	8.4	12.6	11.8	21.0
Male, %	48.0	49.3	41.9	47.6	48.5	49.1	49.6	51.1	41.4
Married, %	71.2	74.6	51.6	74.0	77.7	54.0	73.5	77.6	55.9
Education. cat. 1, %	8.5	8.0	6.0	15.2	13.9	17.4	6.1	4.7	23.5
Education. cat. 2, %	65.6	66.2	68.4	63.2	64.4	65.8	74.1	73.4	54.4
Education. cat. 3, %	25.9	25.8	25.5	21.6	21.7	16.8	19.8	21.9	22.1
HH income < 1/3, %	16.3	14.3	26.3	19.0	16.6	37.6	21.9	19.7	28.7
HH income > 1.5, %	18.0	18.5	14.9	17.2	17.6	14.6	18.6	19.3	15.3
Second job / IEA, %	5.9	6.8	0.0	5.9	6.5	2.5	12.9	13.3	2.7
Land use, %	61.8	62.4	54.9	---	---	---	69.0	69.7	---
Have <i>dacha</i> , %	18.0	18.1	21.7	---	---	---	27.5	29.1	---

Variable	1995 (round VI)			1996 (round VII)			1998/99 (round VIII)		
	Total	E	U	Total	E	U	Total	E	I
Rel. sample size, %	100.0	76.5	8.0	100.0	74.0	9.1	100.0	70.2	16.8
Average age	36.4	37.1	33.1	36.3	37.3	33.0	36.3	37.3	33.9
Age < 25, %	18.1	13.8	30.0	18.5	13.2	28.5	19.3	13.7	36.6
Age > 50, %	11.5	10.5	9.0	10.7	10.0	7.4	11.1	10.6	15.4
Male, %	49.5	50.8	47.5	49.2	50.4	50.2	48.5	49.1	43.7
Married, %	73.8	78.0	59.5	72.8	76.6	62.2	65.2	70.7	61.7
Education. cat. 1, %	5.6	4.5	6.8	4.5	3.2	5.6	3.7	2.9	9.5
Education. cat. 2, %	69.9	70.3	75.7	70.8	70.6	75.9	71.1	70.9	68.8
Education. cat. 3, %	24.6	25.2	17.6	24.8	26.3	18.5	25.2	26.3	21.7
HH income < 1/3, %	24.9	22.7	35.1	26.6	24.4	32.9	45.4	41.2	33.1
HH income > 1.5, %	18.6	19.7	15.1	22.5	23.8	19.1	8.3	8.1	18.8
Second job / IEA, %	12.6	11.8	21.4	11.4	11.1	17.8	12.1	11.6	9.5
Land use, %	68.8	69.1	62.7	69.1	68.6	67.0	67.7	66.5	72.1
Have <i>dacha</i> , %	28.3	29.9	19.9	28.3	30.2	23.2	27.4	28.5	22.7

Notes: Total sample sizes (individuals) – 8,838 (Round I); 7,785 (Round III); 5,915 (Round V); 5,570 (Round VI); 5,492 (Round VII); 5,689(Round VIII).

Table 2.2: Summary statistics for RLMS samples (continued)

Variable	2000 (round IX)				2001 (round X)			
	Total	E	U	I	Total	E	U	I
Rel. sample size, %	100.0	72.2	8.4	19.4	100.0	72.7	8.1	19.2
Average age	36.0	37.0	32.8	34.0	35.9	36.8	32.8	33.7
Age < 25, %	20.2	14.7	32.5	35.4	20.3	14.9	30.1	36.5
Age > 50, %	11.2	10.9	6.7	14.3	11.1	10.6	6.3	15.2
Male, %	47.0	48.5	50.9	39.8	46.4	47.9	53.0	37.7
Married, %	62.8	67.4	50.5	51.1	60.2	65.0	43.4	49.5
Education. cat. 1, %	3.4	2.5	3.0	6.9	2.8	2.0	3.9	5.5
Education. cat. 2, %	70.7	70.8	72.5	69.7	68.9	68.5	71.4	69.1
Education. cat. 3, %	25.9	26.7	24.6	23.4	28.3	29.5	24.8	25.5
HH income < 1/3, %	22.2	18.9	35.8	28.5	20.5	17.4	34.5	26.4
HH income > 1.5, %	21.3	22.0	18.6	19.7	22.2	22.7	17.5	22.0
Second job / IEA, %	14.0	13.3	23.0	12.7	14.4	13.6	25.1	12.8
Land use, %	65.7	64.5	65.7	70.3	61.3	60.1	62.7	65.1
Have <i>dacha</i> , %	24.8	26.1	22.4	20.9	25.3	27.0	19.4	21.6

Variable	2002 (round XI)			
	Total	E	U	I
Rel. sample size, %	100.0	72.7	7.6	19.7
Average age	36.0	36.8	33.2	33.9
Age < 25, %	20.2	14.9	30.8	35.9
Age > 50, %	11.3	10.6	7.2	15.3
Male, %	47.0	47.9	49.6	42.4
Married, %	58.0	62.9	42.6	46.2
Education. cat. 1, %	2.8	1.6	3.4	7.3
Education. cat. 2, %	68.4	68.3	70.1	68.1
Education. cat. 3, %	28.8	30.2	26.5	24.6
HH income < 1/3, %	19.4	15.7	32.3	27.9
HH income > 1.5, %	18.8	20.5	16.2	13.5
Second job / IEA, %	13.1	11.6	23.5	14.4
Land use, %	60.7	59.3	57.6	67.0
Have <i>dacha</i> , %	23.9	25.6	20.3	18.7

Notes: Total sample sizes (individuals) – 5,985 (Round IX); 6,692 (Round X); 6,945 (Round XI).
Variables: Age < 25 – percent of people aged 24 and younger; Age > 50 – percent of people aged 51 and older; Educational category 1 – unskilled, category 2 – skilled, category 3 – high skilled (see Appendix 2.A for more details); HH income < 1/3 – percent with real household income less than one third of the sample average; HH income > 1.5 – percent with real household income more than 1.5 the sample average; Second job / IEA – percent having a secondary job or being involved in individual economic activity (IEA).

Table 2.3: Summary statistics for RLMS samples (some household characteristics)

Variable	RLMS rounds								
	I	III	V	VI	VII	VIII	IX	X	XI
Share of workless households (with no adults working), %	6.83	7.80	10.56	11.29	11.23	14.44	12.07	12.07	12.99
Share of households with all adults working, %	73.25	72.51	62.21	60.81	58.59	53.98	56.40	58.11	58.05
Working age adults in workless households, %	4.28	5.13	7.13	7.50	8.08	10.88	8.69	8.74	9.43
Working age adults in households with all adults working, %	68.90	68.66	57.53	55.98	52.29	47.76	50.16	51.32	52.10
Share of the unemployed from workless households, %	25.54	28.88	32.35	33.11	36.65	38.47	35.45	35.79	35.93
Share of the inactive from workless households, %	27.28	31.07	30.57	31.26	28.14	35.44	29.40	30.40	34.04

Notes: Total number of households represented by persons in working age – 4,687 (Round I); 4,179 (Round III); 3,001 (Round V); 2,825 (Round VI); 2,804 (Round VII); 2,866 (Round VIII); 3,000 (Round IX); 3,380 (Round X); 3,502 (Round XI).

10-15% for the employed) and less likely to be married. Both categories of the jobless are likely to come from households with lower income (the share of families with household income less than one third of a sample average yields persistently higher numbers for the non-employed when pitted against the employed, whereas a proportion of families with income higher than two thirds of a sample average is always higher for the employed).

The comparison of unemployment and inactivity reveals the following results. People in inactivity are on average older. At the same time there are more people younger than 25, and more people in pre-retirement age among non-participants than in any other category - a fact which is quite intuitive and suggests a U-shaped age effect of selection into the out-of-the-labour-force state. It is also easy to spot that inactive individuals are likely to be married females - as a rule both the share of males and the proportion of non-married individuals are lower there than among the unemployed. The inactive have more low-skilled workers in their ranks than any other labour market state. The majority of non-participants, however, are the people with skills of a medium level. The proportion of high skilled individuals is close to that among the unemployed and in some years even higher. On the one hand, this does not fully concur with the view that the inactive are mostly people with primary education (EBRD, 2000). On the other hand, as far as the educational attainment is concerned, the evidence from table 2.2 fairly accords with the view that in many transition countries the unemployed display characteristics more similar to the employed than to the inactive⁸ (*ibid*). Gauging by household income characteristics it appears that the inactive come from families relatively more well-off than the families of the unemployed - the proportion of low income families is often lower, while

⁸EBRD (2000) also points to a higher proportion of people with higher education among the unemployed in Russia compared to their counterparts in Poland.

the proportion of high income families is often higher among non-participants.

Finally, it is found that the unemployed are more often engaged in individual economic activity (in their case - incidental) than people out of the labour force. In general, table 2.2 suggests that the incidence of irregular IEA among both categories of the non-employed in Russia has grown over the 1990s, which is very distinctive of CIS as opposed to CEE countries (EBRD, 2000). Meanwhile, whereas there are no substantial differences in dacha ownership among the unemployed and the inactive, the proportion of those having access to land is in fact highest among non-participants often even compared to the people in employment.⁹ In general, land use increases and then drops over the period from 1992 to 2002. This pattern roughly prevails both for employment and non-employment.

Table 2.3 supplements table 2.2 and presents some evidence on the distribution of unemployment and inactivity among households. In particular, it suggests that both types of non-employment have become a phenomenon more entrenched within families. Both the share of households with no adults having a regular job and the proportion of working age adults in workless families have increased over the 10 years under study. At the same time, the proportion of families where all adult members work has decreased over the same period. Both the unemployed and the inactive have more often come from the workless households. However, this pattern is clearly more pronounced for the former category rather than for the latter.

To summarise the descriptive part of our samples, it seems that on average the employed are people in their thirties, mainly married and better educated, while the jobless

⁹Together with the fact that, as noted above, agriculture is gaining more weight in present day Russia, this observation is well in line with the EBRD (2000) finding on a relative standing of Polish and Russian farmers. EBRD notes that farmers in Poland tend to own their land and to be officially registered as employed in contrast to Russia, where farmers do not declare themselves as employed but merely subsist on land.

are usually younger and less likely to be married. The inactive are more likely to be married females and people close to post-work age, they also appear to be low skilled and possibly having better access to land.

These observations as well as the evidence on household concentration of non-employment in table 2.3 tentatively suggest that the transition has not had the same effect on different categories of workers in Russia. At the same time, the evidence on land use may well imply that certain sectors of the economy (subsistence farming) may actually play a role in labour force reallocation not envisaged by the theory. However, we postpone making any further conclusions about non-uniformity of the transition impact and its correspondence to theoretical predictions till Section 2.3.3, where the static evidence is supplemented by the analysis of labour market dynamics. Before that, however, we shall also establish the relative importance of certain patterns of labour force reallocation by analysing average transition probabilities between different labour market states.

2.3.2 Aggregate Worker Flows

After having looked into the static characteristics of the distribution of the working age population across labour market states we now move on to considering the issues of dynamics. During the 1990s Russia as any other country in Eastern Europe has undergone a massive structural economic change that has directly affected its labour markets. From Fig.2.1-2.3 and table 2.2 we can see that the Russian economy has experienced a decline in the general level of employment accompanied by growing and then diminishing unemployment, and increasing inactivity. Gimpelson and Lippoldt (1999) also report a decreasing share of the state and a rising share of the private sector in total employment.

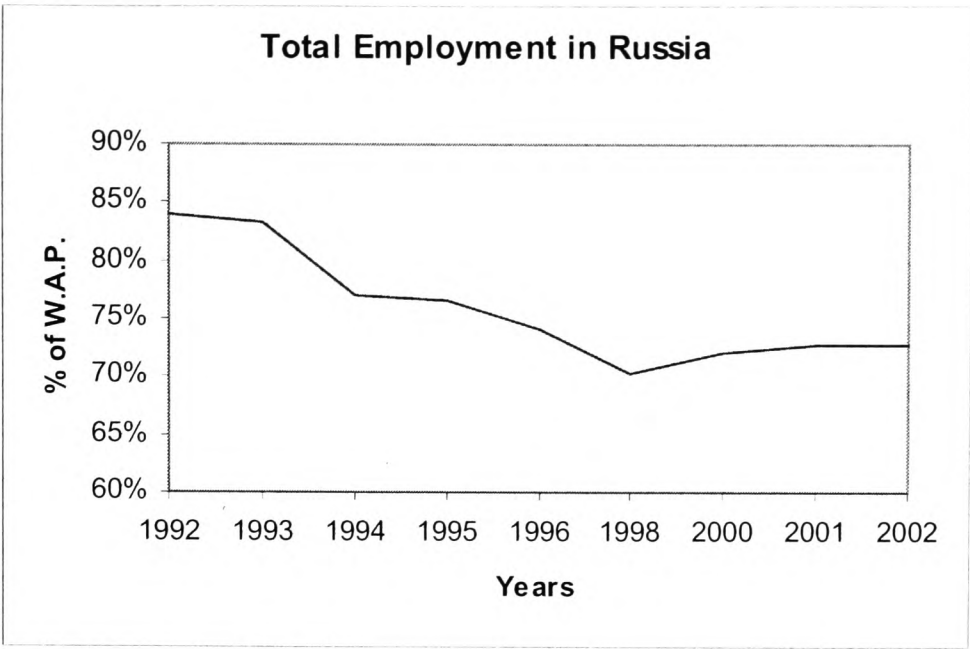


Figure 2.1: Employment rate (% of working age population; source: RLMS 1992-2002)

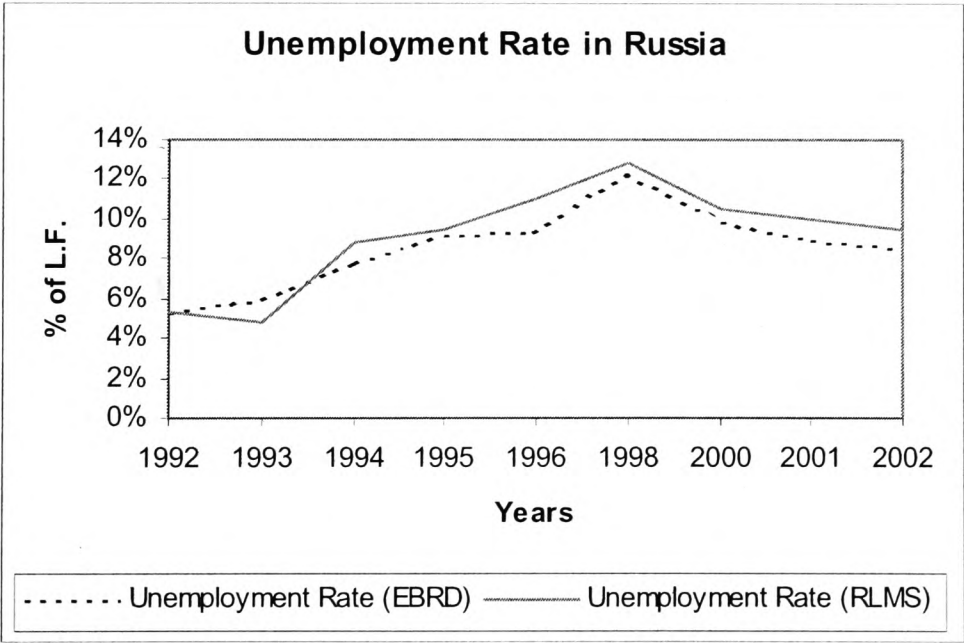


Figure 2.2: Unemployment rate (% of labour force; sources: EBRD, 2000, 2003; RLMS 1992-2002)



Figure 2.3: Inactivity (% of working age population; source: RLMS 1992-2002)

These trends accord well with evidence from other sources (see e.g. EBRD, 2000, 2003), and suggest a picture of the Russian transition comparable in its main features to the experience of CEE countries. However, certain differences are known too - for example, the hump in unemployment in Russia was less pronounced than in CEE, while the relative drop in participation was lower. Still despite the smaller swings in the size of unemployment and inactivity, the flows across the Russian labour market were strongly affected.

Worker flows and transition probabilities: *Quo vadis? Unde venis?*

In order to ascertain the most significant worker flows we analyse average transition probabilities between different labour market states. This approach has become customary in the literature on transition from plan to market (see *inter alia* Boeri and Bruno, 1997; Sorm and Terrell, 2000; Haltiwanger and Vodopivec, 2002; Vodopivec, 2002).

Assuming that shifts between labour market states follow a Markov process (that is,

given the present, the future state is conditionally independent of the past) we calculate separation and accession probabilities for each of the three labour market states.¹⁰

The average constant probability of separation from state i for state j over the time period from t to $t + 1$ in the Markov model can be estimated by:

$$\pi_{ij} = \hat{p}(j_{t+1}|i_t) = \frac{m_{ij}^{t,t+1}}{s_i^t},$$

where s_i^t - is the stock of individuals in state i at moment t , and $m_{ij}^{t,t+1}$ is the number of persons in state i at time t and in state j at moment $t + 1$. Thus, the estimators of such probabilities are defined as gross flows weighted by the initial population in each state.

By analogy, the average constant accession probability to state j from state i is estimated as

$$p_{ji} = \hat{p}(i_t|j_{t+1}) = \frac{m_{ij}^{t,t+1}}{s_j^{t+1}},$$

where s_j^{t+1} - is the stock of individuals in state j at moment $t + 1$.

Boeri and Bruno (1997) stress that such measures underestimate mobility across labour market states "insofar as they exclude "round-tripping", i.e. individuals flowing from i to j and back to i within the two survey dates".

Tables 2.4 and 2.5 present the results of estimation of flows between employment, unemployment and inactivity, while table 2.6 singles out job-to-job transitions for five

¹⁰The view that the probability of transition depends only on the state currently occupied (as modelled when it is assumed that transitions are governed by a Markov process) is vulnerable to critique. So, in her study of the labour market dynamics in Russia Grogan (2003) uses an alternative data set and stresses that it provides advantages over the RLMS by allowing her to trace labour market transitions for a few years back. However, Foley (1997b) notes that using the Markov model approach seems appropriate for an economy subject to a strong, sudden structural shock which mitigates the importance of individual work histories.

Table 2.5: Separation probabilities for employment (E), unemployment (U) and inactivity (I) in Russia, CEE and the US

Country	Reference year	π_{EE}	π_{EU}	π_{EI}	π_{UE}	π_{UI}	π_{IE}	$\pi_{II'}$	$\pi_{II''}$
Russia (for men 18-59 y.o. and women 18-54 y.o.)	1992-1993	92.6	2.2	5.2	51.8	20.3	27.9	25.9	9.0
	1994-1995	91.5	3.5	5.0	45.5	29.2	25.3	22.7	15.7
	1995-1996	90.1	4.0	5.9	42.1	31.0	26.9	22.2	16.1
	2000-2001	90.1	3.7	6.2	46.6	28.1	25.3	24.6	11.8
	2001-2002	90.3	3.6	6.2	43.6	24.9	31.5	23.4	11.8
Russia (for 15-64 y.o.)	1992-1993	75.4	19.3	5.3	35.3	41.4	23.3	10.8	24.9
	1994-1995	70.4	24.7	4.9	31.7	48.1	20.2	10.0	31.2
	1995-1996	73.0	21.2	5.8	29.9	45.8	24.3	9.6	27.9
	2000-2001	78.1	15.5	6.4	38.6	37.8	23.6	12.7	23.0
	2001-2002	77.8	15.9	6.3	34.0	37.5	28.5	13.2	19.8
Bulgaria	6/1994-3/1995	84.9	5.9	9.2	32.3	43.3	24.4	9.2	4.4
Czech Republic	1/1994-4/1994	94.8	2.0	3.2	50.7	31.8	17.6	4.6	1.8
	1/1995-4/1995	95.7	1.5	2.8	47.1	36.2	16.7	5.1	1.1
	1/1996-4/1996	95.5	1.3	3.2	42.6	41.8	14.7	4.6	1.0
	1/1997-4/1997	95.8	1.9	2.4	42.9	44.8	12.4	4.1	1.6
	1/1998-4/1998	95.0	2.3	2.8	36.6	53.5	9.9	3.6	2.2
Estonia	1992-1993	85.5	4.8	9.7	46.5	44.1	9.3	14.3	3.6
	1993-1994	87.3	5.3	7.4	44.7	43.4	12.0	12.8	3.4
	1994-1995	88.3	5.2	6.6	49.9	39.4	10.7	15.6	4.9
	1995-1996	92.7	4.2	3.0	33.6	59.8	6.6	6.2	3.0
	1996-1997	90.1	4.9	5.0	37.4	53.5	9.2	8.5	5.1
	1997-1998	91.3	4.7	4.0	37.2	56.4	6.4	7.4	3.8
Poland	1994-1995	91.0	3.5	5.5	38.0	46.4	15.6	6.0	3.2
Romania	1994-1995	91.2	3.1	5.7	43.1	34.7	22.2	22.5	8.4
Slovak Republic	1995-1996	94.4	2.9	2.7	34.8	61.0	4.2	2.0	1.6
Slovenia	1994-1995	91.8	2.4	5.8	38.7	41.9	19.4	6.3	2.1
USA	1992-1993	91.9	2.8	5.3	65.9	5.3	28.8	4.3	16.1
									79.6

Notes: π_{ij} denotes the separation probability from state i to state j ; all transition probabilities are taken on a yearly basis unless otherwise stated; transitions from employment to employment, π_{EE} for all countries apart from Estonia include job-to-job flows as well as those who remained on the same job; for Estonia and the Czech Republic separations are defined on the basis of the first transition undertaken in the base year, i.e. “round-tripping” workers beginning and ending the year in employment with unemployment spell in between will count as moved from employment to unemployment.
Sources: Russia – own calculations using RLMS (1992-2002); Bulgaria – Boeri (1998); Czech Republic, USA – Sorm and Terrell (1999); Estonia – Vodopivec (2002); Poland, Slovak Republic and Slovenia – Boeri and Bruno (1997); Romania – own calculations based on data from Voicu (2002).

Table 2.6: Churning in employment for the working wage population in Russia and selected CEE states

Country	Job-to-job mobility (% of the employed)
Russia	
1992-1993	11.1
1994-1995	12.8
1995-1996	13.9
2000-2001	17.5
2001-2002	17.8
Czech Republic	
1994	2.5
1996	8.3
Estonia	
1992-1993	15.9
1993-1994	16.9
1994-1995	16.3
1995-1996	9.0
1996-1997	12.2
1997-1998	9.9
Hungary	
1999	5.4
Poland	
1996	4.9
1998	4.9

Notes: churning is defined as the percentage of those that stayed in regular employment but moved to a new job.

Sources: Russia – own calculations using RLMS (1992-2002), the Czech Republic – Sorm and Terrell (1999), Estonia – Vodopivec (2002), Hungary, Poland – EBRD (2000).

Table 2.7: Relative stagnancy of the two non-employment states in Russia

Proportions of the unemployed for 6 months and more, %								
1992	1993	1994/95	1995	1996	1998/99	2000	2001	2002
60.48	61.80	66.06	74.32	75.10	73.73	72.08	71.96	68.63
Proportions of the inactive for 6 months and more, %								
1992	1993	1994/95	1995	1996	1998/99	2000	2001	2002
83.35	82.64	85.15	89.85	86.26	88.34	88.53	86.86	87.44
Average lapsed duration of unemployment relative to duration of inactivity								
1992	1993	1994/95	1995	1996	1998/99	2000	2001	2002
0.40	0.43	0.51	0.59	0.69	0.64	0.60	0.60	0.58

Source: Own calculations using the RLMS (1992-2002).

Table 2.8: Coverage by unemployment benefits

Proportions of the unemployed receiving unemployment insurance, %								
1992	1993	1994/95	1995	1996	1998/99	2000	2001	2002
6.27	5.90	14.12	13.06	13.94	8.47	4.95	6.64	10.27
Proportions of the unemployed not receiving any unemployment insurance, %								
1992	1993	1994/95	1995	1996	1998/99	2000	2001	2002
74.94	90.99	7.52	9.91	7.37	7.12	4.16	1.85	2.66

Notes: Proportions do not sum up to 100% across rows because of missing observations.

Source: Own calculations using the RLMS (1992-2002).

panels of the RLMS data: 1992/93, 1994/95, 1995/96, 2000/01, and 2001/02. The following patterns stand out.

First and foremost, it is obvious that hirings are persistent both from the unemployment pool and the pool of the inactive. Moreover, judging by the accession probabilities (columns 5 and 6, top panel, in table 2.4) a steadily greater proportion of the newly employed arrives from the out-of-the-labour-force state rather than from unemployment. Still, however, the intake comes mainly from other employment (compare column 4, table 2.4). Such hiring from the employment pool lends support to previous observations of significant job-to-job movement in Russia: Layard and Richter (1995) note that newly established private firms recruit their workers mainly from state enterprises rather from the unemployment pool which would have otherwise offered cheaper labour, while Grogan (2003) points to a large fraction of job movers taking the job-to-job transition route. Churning (i.e. moving jobs without an intervening spell of non-employment) is indeed significant for total employment as also reported in table 2.6. All in all, it is obvious that the degree of direct job-to-job switches in Russia is increasing over time, which may be a sign of growing flexibility and efficiency of its labour market.

Second, as regards outflow rates from employment to non-employment it seems that the pace of labour reallocation in that direction becomes more moderate. From table 2.5 (columns 4 and 5, top panel) it is seen that the combined outflow rates to unemployment and inactivity rose till the middle of the 1990s and then decelerated. Accession rates to non-employment from employment in table 2.4 (columns 7 and 10) point to the same pattern. This suggests that the massive reallocation of labour from employment elsewhere was taking place in the first half of the 1990s and became more steady by the end of decade.

This is also perfectly in line with labour market dynamics shown in Fig.2.1-2.3.

The third observation is concerned with the role of unemployment and inactivity in re-deployment of labour resources. Layard and Richter (1995) and Lehmann and Wadsworth (2000) have pointed to the unlikeliness of labour reallocation through unemployment in Russia. Tables 2.4, 2.5, and 2.6 indicate that although flows through unemployment and inactivity are still less significant than direct job-to-job shifts, both non-employment states appear to have become more important in the reallocation of labour in the second half of the 1990s. The data in columns 5 and 6 (table 2.4, top panel) posit that the proportion of newly employed coming from the non-employment pool is increasing. The change between 1992 and 2002 is highly statistically significant (z-value is 8.7).

However, the growing importance of the non-employment pool in the reallocation of labour is accompanied by an increase in average duration of joblessness. Table 2.4 (columns 8 and 12) suggests that the proportions of the non-employed abiding in the same state as a year ago have risen from 1992 to 2002. Similarly, from table 2.5 (top panel, columns 4, 6, 8, and 10) it can be seen that while the flows from employment and inactivity to unemployment on average rose over 10 years, the total flows out of unemployment to employment and inactivity shrank, albeit remaining at quite a high level. A similar pattern is observed for inactivity, although it is interesting that the state also exhibits high outflow rates to employment (Smirnova, 2003a, makes a similar conclusion). In general, inactivity obviously represents a more stagnant labour market state than unemployment (table 2.7).

Finally, we conclude by noting that there are considerable flows between inactivity and unemployment in both directions (see columns 9 and 11 in table 2.4, and columns 8

and 10 in table 2.5). We discuss these issues in more detail in the next subsection where we set Russia against other transitional countries.

How does Russia compare to other countries of CEE?

There is a number of studies providing evidence on transition probabilities drawn from labour force or household surveys for countries in CEE (*inter alia* Boeri and Bruno, 1997; Boeri, 1998; Sorm and Terrell, 1999; Vodopivec, 2002; Voicu, 2002; etc.). However, before getting down to a comparison of the peculiarities of labour market dynamics in Russia and other transitional economies we should mention one issue of practical importance.

Countries often differ in their definition of the working age, or researchers use different age restrictions applied to their samples.¹¹ This may affect the results of the analysis as can be seen from top two panels in tables 2.4 and 2.5, where we report transition probabilities for samples of Russian individuals in the official working age (men 18-59 y.o. and women 18-54 y.o.) and in the working age as often defined by OECD (i.e. 15-64 y.o.). Transition rates for movements between employment and unemployment, and between inactivity and employment or unemployment are often significantly different.¹² The following results of our comparison of Russia and Central and Eastern Europe, however, are largely unaffected by which definition of the working age we apply to the Russian sample.

The two major points highlighted by tables 2.4, 2.5, and 2.6 are interesting.

First, the overall labour turnover in Russia is higher than in other CEE countries

¹¹For example, Foley (1997b) and Brainerd (1998) use samples of Russian men 18-59 y.o. and women 18-54 y.o. (i.e. the official working age according to the Russian laws). Vodopivec (2002) in his analysis of the Estonian labour market uses a sample of individuals from 15 to 69 y.o. Others (e.g. Grogan, 2003) may consider people from 23 y.o. to the local retirement age, depending on a country in focus. The ILO statistical database, LABORSTA, provides evidence for people from 15 to 72 y.o.

¹²This can probably be explained by different intensity of shifts on the margin of labour force, i.e. young people, including students, and people close to the retirement age.

(previous observations of this fact include *inter alia* EBRD, 2000, and Lehmann and Wadsworth, 2000). On the one hand, this is evident from the degree of job-to-job transitions. The share of those who change jobs directly is very high in Russia and rises over the decade of transition from 11.1% to 17.8%. These figures are comparable to those of a highly dynamic labour market in Estonia and are much higher than those in other transition countries¹³ (see table 2.6). On the other hand, inflows from unemployment and inactivity in Russia are at least at comparable levels or even higher than in other CEE economies¹⁴ (tables 2.4 and 2.5).

Second, the observation which is striking is the flows between ILO unemployment and inactivity that are much larger in Russia than in CEE, and are comparable to those in the US labour market (table 2.5). Moreover, both Russian unemployment and inactivity clearly appear to be less stagnant states than their counterparts in Eastern Europe (see columns 8 and 12 in table 2.4, and columns 7 and 11 in table 2.5). The only possible East European exception is Romania which has displayed a similar level of mobility between the two non-employment states. Our intuition for this fact is twofold. First, the significant extent of the informal economy in Russia (for some recent estimates see, e.g. Lackó, 2000, or Feige and Urban, 2003) may well affect flows into and out of inactivity. In particular, widespread subsistence activities in farming may have a bearing on flows into inactivity, especially from unemployment. Voicu (2002) argues in the Romanian example that employment in the agricultural sector can be a labour market state similar to unemployment

¹³It appears that job-to-job transitions are countercyclical in Russia - they steadily rose during the years of the most severe output decline 1992-1998. Vodopivec (2002) reports that quits are procyclical in the West and many CEE, but countercyclical in Estonia (quits surged in the period of most intensive output reduction - 1992-1993).

¹⁴EBRD (2000) reports lower outflow rates from unemployment and inactivity in Russia (6.5%) than in Poland (8.0%) or Hungary (8.9%) in 1998. Our findings above (table 2.5) show that the outflow rates in Russia were roughly comparable to those in the CEE for the working age population defined as by the OECD, but much higher for the working age population defined according to Russian law.

because it absorbs people laid off elsewhere. Second, a system of unemployment compensation, which is nearly non-existent, may be to blame for the more mobile behaviour of the Russian unemployed in comparison to Eastern Europe where unemployment benefits have been much more generous (Boeri, 2000b). Table 2.8 provides some evidence on the coverage by unemployment benefits in Russia, according to which among the unemployed in the RLMS samples at most slightly more than 14% could confidently indicate the receipt of unemployment compensation in any of the rounds considered.

Does such a mobile labour market do any good to Russian transition? According to EBRD (2000), higher mobility is not necessarily desirable in itself, it is the reallocation from less to more productive enterprises that is of crucial importance. The existing literature studying the Russian case suggests that in this respect the Russian performance has thus far been less impressive than that of Eastern Europe. Large worker shuffling among the existing set of jobs (Clarke, 1999) does not create a precondition for growth. The substantial worker reallocation has to be accompanied by enough job creation and job reallocation from less to more efficient sectors. Brown and Earle (2003) show that although there has recently been a significant increase in most measures of job and worker reallocation in Russia, most of the increases are due to higher job destruction and worker separations. In other words, job creation - a factor widely held to be vital for growth - has yet to pick up to make the process of restructuring more balanced. At the beginning of transition, as found in another study by Brown and Earle (2002), a legacy of the Soviet economic system displayed the job flow behaviour quite different from that in market economies, with rates of job reallocation that bore little relationship to relative productivity across firms and sectors. This stands in contrast with the highly mobile Estonian

labour market where substantial worker reallocation was accompanied by job reallocation with gross job creation coming from the more productive private sector (Haltiwanger and Vodopivec, 2002). Further evidence from Brown and Earle (2002) suggests that transition in Russia is bringing substantial increases in the productivity-enhancing consequences of the reallocation process. However, there is an obvious need for more research on this subject which reaches beyond the scope and capabilities of this work.¹⁵

Summary and further questions

To conclude the discussion of the main patterns of labour force reallocation the main findings can be summarised as follows. It is evident that the Russian labour market has exhibited significant fluidity over the last decade that distinguishes it from other transitional countries. Especially, the direct job-to-job movements seem to be of centre importance. At the same time, the RLMS gives evidence pointing to the growing role of non-employment in reallocating labour resources, in particular, significant flows between unemployment and inactivity. Still, an increase in stagnancy of both non-employment states has been observed over the course of transition. Also, according to other sources, generally buoyant labour reallocation is, however, unlikely to have been substantially improving efficiency of the economy.

In this regard the following three issues appear to be important and possible to investigate further with the help of the RLMS data set.

First is the question of uniformity of movements. The significant mobility of the Russian labour force is unlikely to have taken a uniform pattern for all members of the

¹⁵Being a household survey the RLMS does not provide any information as regards productivity or a degree of restructuring of employers.

society. Tables 2.2 and 2.3 analysed above have provided some initial evidence of that. So, which groups of individuals are more susceptible to job loss or face lower chances of reemployment? Which groups face a higher probability of exiting the labour force?

Second, what are the factors conducive to movements into and out of inactivity? While the basic theory of transition has attached a crucial importance to the reallocation through the unemployment pool, the Russian example clearly shows that inactivity may play an equally significant part in redeployment of the workforce.¹⁶

Finally, given the fluidity observed in the labour market, why do unemployment and inactivity become more stagnant states? Again, little if any insights can be drawn from the theoretical studies that have in the main envisaged unemployment in transition that is not a persistent phenomenon.¹⁷

We now proceed to obtain some answers to these questions. First we shall take up the issues of movement non-uniformity and factors possibly affecting shifts into inactivity. It is done by means of a multinomial logit analysis of the individual and household characteristics and their impact on the yearly probability of transitions across the three labour market states. A special focus will be those workers who drop out of the labour force. With regard to the latter we try, among other things, to verify a hypothesis that access to land may be a determinant of the decision to non-participate.

¹⁶We avoid dealing with any in-depth analysis of job-to-job transitions on the basis of the RLMS data, as it is not particularly informative on types of ownership (see also footnote 7), industries, employer characteristics, etc. The reader, however, can be referred to Grogan (2003) for some insights into movement between Soviet-era and post-Soviet jobs, given with the help of two data bases, one of which is the RLMS.

¹⁷Commander *et al.* (2004) make the most recent attempt at understanding the fact of persistence.

2.3.3 Logit Modelling

In order to identify some of the determinants of shifts between different labour market states we pool the data from the five panels of the RLMS (containing information on labour market transitions in 1992/93, 1994/95, 1995/96, 2000/01, and 2001/02) and estimate multinomial logit models of labour market dynamics between employment, unemployment and inactivity.

Two main types of model are constructed - one for transitions from employment, and another one for transitions from non-employment states, i.e. unemployment and inactivity. In the base model for employment transitions (labeled hereafter E1) fitted to 1992-2002 data independent variables include characteristics both of the individual (such as age, age squared, dummies for being younger than 25 y.o. or older than 50 y.o., gender, marital status, indicator variables of being a female in the family with children, the skill (education) level), and of the household (household income, the number of post-work age persons, the number of children, indicator variables of presence of children younger than 7 y.o. and in between 7 y.o. and 18 y.o., access to land, dacha ownership), and indicator variables of economic conditions (such as possession of a secondary job or involvement in individual economic activity, belonging to the metropolitan labour market). As some additional data become available in years 1994-2002 we expand the base model to include a variable for tenure in the current job and indicators of wage arrears and unpaid involuntary leave (model E2). This model is in turn completed with a variable for the number of various social benefits, to which a respondent has access at her primary place of employment, and it is fitted to 2000-2002 data (model E3). In addition to the set of variables described for the base model of employment transitions the models for non-employment transitions

(U1 and I1) include indicator variables of various duration of unemployment. These two models are fitted to 1992-2002. Finally, in order to control for yearly changes in the base probabilities due to national demand conditions or other shocks coming at different periods, two variables for local unemployment and non-employment rates¹⁸ (i.e. both unemployment and inactivity in percent of the working age population) are included in all models both for transitions from employment and from non-employment.¹⁹ As each individual in our logit samples is surveyed twice over approximately a year-long period, the values of all independent variables are taken as of the date of the first interview, i.e. preceding a potential change of the labour market state in the year that follows.

Results of estimation²⁰ are presented in tables 2.9-2.11. All the models are highly significant (model χ^2 -statistics are reported in the bottom panels of the tables). The baseline category is always the origin state. For each type of transition the first column of the table reports estimates and their standard errors (in parentheses), while the second column gives the marginal effects²¹ on transition probability and their standard errors

¹⁸We have also tried to fit the models with a variable for the inactivity rate (i.e. the percentage of the inactive in the working age population) instead of the non-employment rate. However, such an exercise did not affect the conclusions qualitatively.

¹⁹The consequences of introduction into the model of yearly dummies for 1994-95, 1995-96, 2000-01, and 2001-02, with the base category 1992-93, have also been investigated. Whenever the dummies were significant the only substantial effect was a change in size and significance of coefficients on the unemployment or/and non-employment rates. This held good for all models considered. A similar effect on coefficients on the unemployment or/and non-employment rates was caused by a separate introduction of a dummy for observations coming from panels dating before the ruble devaluation crisis in August, 1998. Thus, it seems that the unemployment and non-employment rate variables catch most of the yearly effects in our pooled sample.

²⁰All multinomial logit models are estimated in Stata, version 8.2.

²¹If we have N labour market states the multinomial logit model for labour market transitions is described by $P_j = \Pr(Y_i = j) = \frac{e^{\mathbf{X}_i' \gamma_j}}{\sum_{k=1}^N e^{\mathbf{X}_i' \gamma_k}}$, $j = 1, \dots, N$. The estimated equations provide a set of probabilities for N choices for an individual with characteristics $\mathbf{X}_i$. The interpretation of coefficients γ_j in this model is difficult. It would be misleading to associate γ_j with j -th outcome. An easy way to see that is calculate the marginal effects of the characteristics $\mathbf{X}_i$ on probabilities: $\frac{\partial P_j}{\partial \mathbf{X}_i} = P_j \left[\gamma_j - \sum_{k=1}^N P_k \gamma_k \right] = P_j [\gamma_j - \bar{\gamma}]$. Thus, every subvector of γ contributes to every marginal effect both directly and via the probabilities P_k that appear in the summation term above. Greene (2003) notes that by tradition the focus is on coefficient estimates, but they should not be confused with the marginal effect on probability.

Table 2.9: Multinomial Logit models of transitions from employment

Independent variable	Model E1 (1992-2002)				Model E2 (1994-2002)			
	E to U		E to I		E to U		E to I	
	Estimate	dP/dx	Estimate	dP/dx	Estimate	dP/dx	Estimate	dP/dx
AGE	-0.0668 (0.058)	-0.0014 (0.002)	-0.3493** (0.040)	-0.0149** (0.002)	0.0593 (0.072)	0.0019 (0.002)	-0.3575** (0.052)	-0.0137** (0.002)
AGESQRD	0.0550 (0.078)	0.0009 (0.002)	0.4766** (0.053)	0.0203** (0.002)	-0.0922 (0.098)	-0.0029 (0.003)	0.5007** (0.068)	0.0192** (0.003)
AGE24	-0.0018 (0.196)	0.0001 (0.005)	-0.1421 (0.161)	-0.0058 (0.006)	0.2705 (0.238)	0.0080 (0.008)	-0.1325 (0.209)	-0.0051 (0.007)
AGE51	0.1776 (0.269)	0.0053 (0.008)	-0.0292 (0.166)	-0.0015 (0.007)	0.5620* (0.332)	0.0182 (0.013)	-0.0101 (0.214)	-0.0011 (0.008)
SINGLFEM	-0.5845** (0.137)	-0.0135** (0.003)	-0.2316* (0.119)	-0.0087* (0.004)	-0.6044** (0.174)	-0.0131** (0.003)	-0.2270 (0.156)	-0.0076 (0.005)
MARIDFEM	-0.8824** (0.140)	-0.0223** (0.003)	-0.0123 (0.113)	0.0005 (0.005)	-0.8090** (0.174)	-0.0193** (0.004)	-0.0692 (0.148)	-0.0018 (0.006)
MARIDMAL	-0.6967** (0.126)	-0.0176** (0.003)	-0.5776** (0.109)	-0.0227** (0.004)	-0.5766** (0.155)	-0.0138** (0.004)	-0.5353** (0.141)	-0.0189** (0.005)
CHLD_7	0.1106 (0.086)	0.0027 (0.002)	0.2941** (0.067)	0.0125** (0.003)	0.1294 (0.106)	0.0032 (0.003)	0.2228** (0.091)	0.0084** (0.003)
CHLD7_18	0.0567 (0.054)	0.0014 (0.002)	0.1319** (0.042)	0.0056** (0.002)	0.0361 (0.067)	0.0008 (0.002)	0.0926 (0.056)	0.0035 (0.002)
CHILDFEM	0.0919 (0.161)	0.0026 (0.005)	0.0060 (0.120)	0.0001 (0.005)	0.1601 (0.197)	0.0044 (0.006)	0.0285 (0.162)	0.0009 (0.006)
POSTWRK	0.1110 (0.072)	0.0028 (0.002)	0.1975** (0.056)	0.0083** (0.002)	0.1516* (0.085)	0.0038* (0.002)	0.1732** (0.071)	0.0064** (0.003)
DSKILL1	0.1769 (0.192)	0.0047 (0.006)	0.3413** (0.117)	0.0165** (0.007)	0.1775 (0.260)	0.0048 (0.008)	0.1832 (0.175)	0.0073 (0.008)
DSKILL3	-0.4172** (0.106)	-0.0100** (0.002)	-0.6043** (0.089)	-0.0224** (0.003)	-0.3363** (0.128)	-0.0076** (0.003)	-0.6420** (0.119)	-0.0211** (0.003)
HHINCOME / 10 ⁴	-0.0618 (0.096)	-0.0018 (0.003)	0.0389 (0.058)	0.0017 (0.002)	-0.2377 (0.153)	-0.0062 (0.004)	-0.0137 (0.091)	-0.0003 (0.003)
LANDUSE	-0.0495 (0.095)	-0.0016 (0.003)	0.1500** (0.075)	0.0063** (0.003)	0.0106 (0.120)	0.0001 (0.003)	0.1422 (0.101)	0.0053 (0.004)
OWNDACHA	-0.0881 (0.106)	-0.0018 (0.003)	-0.5002** (0.086)	-0.0192** (0.003)	-0.0995 (0.128)	-0.0022 (0.003)	-0.3821** (0.108)	-0.0135** (0.004)
METRO	0.1832 (0.144)	0.0056 (0.005)	-0.1163 (0.132)	-0.0050 (0.005)	0.2734 (0.180)	0.0082 (0.006)	-0.1939 (0.185)	-0.0071 (0.006)
SECJBIEA	0.5150** (0.109)	0.0154** (0.004)	0.8002** (0.085)	0.0447** (0.006)	-0.0871 (0.160)	-0.0022 (0.004)	-0.0313 (0.141)	-0.0011 (0.005)
TENURE	--	--	--	--	-0.0057** (0.001)	-0.0001** (0.000)	-0.0023** (0.000)	-0.0001** (0.000)
ARREARS	--	--	--	--	0.1820* (0.105)	0.0045 (0.003)	0.3019** (0.086)	0.0118** (0.004)
ULEAVE	--	--	--	--	0.4996** (0.181)	0.0154** (0.007)	0.4315** (0.145)	0.0187** (0.008)
URATE	6.8199** (3.479)	0.1949** (0.096)	-4.0162 (2.900)	-0.1807 (0.124)	3.4817 (4.166)	0.0922 (0.109)	-1.0017 (3.565)	-0.0419 (0.136)
NONRATE	0.4543 (1.538)	0.0066 (0.043)	4.7257** (1.281)	0.2018** (0.055)	-0.0668 (1.796)	-0.0054 (0.047)	3.3711** (1.523)	0.1286** (0.058)
Constant	-1.8353* (1.049)		2.2673** (0.752)		-3.7008** (1.324)		2.3135** (0.980)	
Predicted transition probs (mean sample values)	3.21%		5.57%		3.28%		4.96%	
	Log-likelihood		-6868.93		Log-likelihood		-4272.98	
	Pseudo-R ²		0.0557		Pseudo-R ²		0.0594	
	Model $\chi^2(40)$		810.69**		Model $\chi^2(46)$		539.87**	
	No of observations		20478		No of observations		13365	

Notes: Standard errors are reported in parentheses. ** and * denote significance at 5% and 10% levels, respectively. The default category is the origin labour market state (i.e. stayers in employment). Excluded categories: single male, DKILL2 – skilled individuals.

Table 2.10: Multinomial Logit model of transitions from employment

Independent variable	Model E3 (2000-2002)			
	E to U		E to I	
	Estimate	dP/dx	Estimate	dP/dx
AGE	-0.0720 (0.101)	-0.0013 (0.002)	-0.4260** (0.073)	-0.0137** (0.002)
AGESQRD	0.0848 (0.137)	0.0014 (0.003)	0.6085** (0.095)	0.0195** (0.003)
AGE24	0.0321 (0.329)	0.0009 (0.008)	-0.1830 (0.300)	-0.0056 (0.009)
AGE51	0.2904 (0.454)	0.0075 (0.013)	-0.3252 (0.286)	-0.0095 (0.007)
SINGLFEM	-0.4453* (0.228)	-0.0088** (0.004)	-0.0686 (0.211)	-0.0019 (0.007)
MARIDFEM	-0.7807** (0.250)	-0.0158** (0.005)	-0.0096 (0.213)	0.0002 (0.007)
MARIDMAL	-0.5984** (0.221)	-0.0121** (0.004)	-0.3703* (0.201)	-0.0110* (0.006)
CHLD_7	0.1905 (0.173)	0.0041 (0.004)	0.2058 (0.157)	0.0065 (0.005)
CHLD7_18	0.0526 (0.107)	0.0011 (0.002)	0.0597 (0.090)	0.0019 (0.003)
CHILDFEM	0.2086 (0.294)	0.0049 (0.008)	0.1087 (0.250)	0.0034 (0.009)
POSTWRK	0.2308* (0.126)	0.0050* (0.003)	0.1992* (0.110)	0.0062* (0.004)
DSKILL1	0.4153 (0.409)	0.0108 (0.013)	0.3731 (0.302)	0.0137 (0.013)
DSKILL3	-0.0721 (0.175)	-0.0014 (0.004)	-0.3135** (0.155)	-0.0094** (0.004)
HHINCOME / 10 ⁱ	-0.8741** (0.352)	-0.0195** (0.008)	-0.0077 (0.131)	0.0004 (0.004)
LANDUSE	0.0382 (0.174)	0.0009 (0.004)	-0.1093 (0.144)	-0.0036 (0.005)
OWNDACHA	-0.1436 (0.199)	-0.0030 (0.004)	-0.1969 (0.163)	-0.0060 (0.005)
METRO	0.4731* (0.253)	0.0128 (0.008)	0.0710 (0.244)	0.0019 (0.008)
SECJBIEA	-0.2057 (0.244)	-0.0041 (0.005)	-0.2990 (0.224)	-0.0085 (0.006)
TENURE	-0.0059** (0.001)	-0.0001** (0.000)	-0.0024** (0.001)	-0.0001** (0.000)
ARREARS	0.0289 (0.168)	0.0003 (0.004)	0.4594** (0.130)	0.0163** (0.005)
ULEAVE	0.1648 (0.399)	0.0035 (0.010)	0.4166 (0.274)	0.0160 (0.013)
BENEFIT	-0.1192** (0.038)	-0.0025** (0.001)	-0.2181** (0.034)	-0.0069** (0.001)
URATE	-5.2564 (7.583)	-0.1213 (0.169)	5.1412 (6.252)	0.1695 (0.201)
NONRATE	2.7793 (3.225)	0.0605 (0.072)	1.9832 (2.590)	0.0617 (0.083)
Constant	-0.8580 (1.844)		3.6478** (1.393)	
Predicted transition probs (mean sample values)	3.06%		4.66%	
	Log-likelihood		-1974.63	
	Pseudo-R ²		0.0856	
	Model $\chi^2(48)$		369.50**	
	No of observations		6670	

Notes: Standard errors are reported in parentheses. ** and * denote significance at 5% and 10% levels, respectively. The default category is the origin labour market state (i.e. stayers in employment). Excluded categories: single male, DKILL2 – skilled individuals.

Table 2.11: Multinomial Logit models of transitions from non-employment

Independent variable	Model U1 (1992-2002)				Model I1 (1992-2002)			
	U to E		U to I		I to E		I to U	
	Estimate	dP/dx	Estimate	dP/dx	Estimate	dP/dx	Estimate	DP/dx
AGE	0.0270 (0.084)	0.0305* (0.018)	-0.1984** (0.089)	-0.0413** (0.014)	0.2257** (0.052)	0.0340** (0.009)	0.1789** (0.066)	0.0130* (0.007)
AGESQRD	-0.0420 (0.115)	-0.0440* (0.024)	0.2794** (0.120)	0.0587** (0.019)	-0.3410** (0.070)	-0.0513** (0.012)	-0.2738** (0.089)	-0.0200** (0.009)
AGE24	0.2608 (0.305)	0.0974 (0.064)	-0.2786 (0.339)	-0.0816 (0.050)	0.6965** (0.220)	0.1175** (0.040)	0.3044 (0.282)	0.0121 (0.029)
AGE51	-0.2038 (0.412)	-0.0524 (0.083)	0.0180 (0.424)	0.0279 (0.071)	0.1250 (0.255)	0.0265 (0.045)	-0.1717 (0.323)	-0.0206 (0.031)
SINGLFEM	0.1750 (0.186)	0.0309 (0.040)	0.1035 (0.201)	-0.0015 (0.033)	0.0684 (0.128)	0.0128 (0.022)	-0.0348 (0.153)	-0.0055 (0.015)
MARIDFEM	0.4743** (0.199)	0.0695* (0.041)	0.3883* (0.214)	0.0158 (0.035)	0.0989 (0.140)	0.0222 (0.024)	-0.1904 (0.171)	-0.0223 (0.017)
MARIDMAL	0.4289** (0.194)	0.1235** (0.041)	-0.1431 (0.217)	-0.0766** (0.032)	0.6947** (0.151)	0.1216** (0.030)	0.3176* (0.183)	0.0123 (0.020)
CHLD_7	0.2531* (0.140)	0.0392 (0.029)	0.1972 (0.160)	0.0075 (0.025)	0.0508 (0.085)	0.0101 (0.014)	-0.0507 (0.113)	-0.0067 (0.012)
CHLD7_18	-0.1624** (0.074)	-0.0275* (0.016)	-0.1076 (0.081)	-0.0012 (0.014)	0.0120 (0.051)	0.0006 (0.008)	0.0547 (0.062)	0.0055 (0.006)
CHILDFEM	-0.2635 (0.245)	-0.0106 (0.051)	-0.4995* (0.281)	-0.0617 (0.040)	0.1774 (0.156)	0.0303 (0.027)	0.0353 (0.205)	-0.0011 (0.021)
POSTWRK	-0.0898 (0.106)	-0.0318 (0.022)	0.0792 (0.113)	0.0260 (0.018)	-0.1302* (0.073)	-0.0222* (0.012)	-0.0071 (0.087)	0.0027 (0.009)
DSKILL1	0.0246 (0.311)	-0.0603 (0.061)	0.5005 (0.308)	0.1023* (0.057)	-0.7283** (0.180)	-0.0953** (0.022)	-0.6375** (0.220)	-0.0447** (0.017)
DSKILL3	0.3535** (0.155)	0.0775** (0.032)	0.0870 (0.177)	-0.0267 (0.027)	-0.1055 (0.104)	-0.0171 (0.017)	-0.0272 (0.130)	-0.0001 (0.013)
HHINCOME / 10'	0.2189 (0.157)	0.0367 (0.028)	0.1473 (0.171)	0.0020 (0.023)	0.0208 (0.063)	0.0034 (0.010)	0.0075 (0.079)	0.0002 (0.008)
LANDUSE	-0.0566 (0.141)	-0.0578* (0.030)	0.3746** (0.158)	0.0764** (0.024)	-0.0812 (0.099)	-0.0102 (0.017)	-0.1406 (0.122)	-0.0130 (0.013)
OWNDACHA	0.0297 (0.161)	0.0608* (0.034)	-0.4773** (0.186)	-0.0884** (0.026)	0.1675 (0.108)	0.0311 (0.019)	-0.0575 (0.142)	-0.0105 (0.014)
METRO	0.1887 (0.225)	0.0358 (0.047)	0.0919 (0.270)	-0.0056 (0.043)	0.3456** (0.154)	0.0590** (0.029)	0.1580 (0.198)	0.0066 (0.021)
SECJBIEA	0.3723** (0.162)	0.1201** (0.034)	-0.2383 (0.193)	-0.0852** (0.027)	0.8216** (0.131)	0.1353** (0.027)	0.6833** (0.159)	0.0505** (0.020)
UDUR03	0.8900** (0.155)	0.2434** (0.030)	-0.2286 (0.188)	-0.1408** (0.023)	0.7479** (0.130)	0.1246** (0.026)	0.5786** (0.164)	0.0404** (0.020)
UDUR46	0.3479* (0.204)	0.1467** (0.043)	-0.5736** (0.256)	-0.1294** (0.029)	0.7408** (0.171)	0.1192** (0.035)	0.6706** (0.211)	0.0531* (0.028)
UDUR0712	0.4856** (0.186)	0.1224** (0.038)	-0.0134 (0.208)	-0.0606** (0.029)	0.4144** (0.151)	0.0659** (0.029)	0.3422* (0.192)	0.0255 (0.022)
URATE	-1.8339 (5.153)	0.8483 (1.095)	-10.8584* (5.926)	-1.8658* (0.971)	2.1407 (3.897)	0.1076 (0.647)	9.7873** (4.781)	0.9813** (0.489)
NONRATE	-0.6095 (2.234)	-0.6989 (0.476)	4.5571* (2.552)	0.9487** (0.420)	-2.1230 (1.709)	-0.2622 (0.284)	-3.8561* (2.080)	-0.3526* (0.213)
Constant	-0.3028 (1.534)		2.9567* (1.631)		-4.4817** (0.960)		-4.1548** (1.219)	
Predicted transition probs (mean sample values)	45.91%		27.13%		23.65%		12.53%	
	Log-likelihood		-1730.39		Log-likelihood		-3176.51	
	Pseudo-R ²		0.0578		Pseudo-R ²		0.0640	
	Model $\chi^2(46)$		212.43**		Model $\chi^2(46)$		434.31**	
	No of observations		1725		No of observations		3822	

Notes: Standard errors are reported in parentheses. ** and * denote significance at 5% and 10% levels, respectively. The default category is the origin labour market state (i.e. stayers either in unemployment or in inactivity as appropriate). Excluded categories: single male, DKILL2 – skilled individuals, duration of non-employment greater than 12 months.

(in parentheses), evaluated at the means of independent variables. Before discussing our findings, however, we should mention a potentially important drawback of the multinomial logit models particularly relevant to studying transitions between labour market states. This is the widely known independence of irrelevant alternatives property.

Testing for the IIA

One of the properties of the multinomial logit model is that the ratios of probabilities of exit to two destinations do not depend on the probabilities of exit to remaining destinations. This is called the independence of irrelevant alternatives assumption (IIA), and it follows from the assumption that model disturbances are independent and homoscedastic. The property is not necessarily indicated by any data, and can and should be tested. A simple example where the property does not hold can be found in Greene (2003). In the case of violation of the IIA the multinomial logit model produces inconsistent estimates.

It is true that whilst many researchers recognise the implications of the IIA, it often remains a maintained assumption in applications (Hausman and McFadden, 1984). Moreover, as Turunen (2000) correctly points out this assumption may not *a priori* be acceptable in the case of choice between different employment states, which is crucial to studies of labour mobility in transition economies.²² Many analysts of multiple logit models for the Russian labour market either do not test for the IIA property at all, or conduct econometrically different tests that are not aimed at verifying the validity of the IIA assumption. For example, in the widely cited paper on the labour market dynamics in Russia Foley (1997b) employs the likelihood ratio test to test for the equality of coeffi-

²²For instance, Foley (1997b), Sorm and Terrell (1999), and Grogan (2003) estimate multinomial logit models for transitions between employment, new employment, unemployment and inactivity.

cients in multiple logit equations for transitions from employment to unemployment and from employment to the out-of-the-labour-force state. However, this method does not test for the IIA, but rather allows one to conclude that the two non-employment states may be not behaviourally meaningless distinctions for the sample used (much in the spirit of Flinn and Heckman, 1983). Even if the null hypothesis of artificial distinctions in the two non-employment states is rejected, the specification of the multinomial logit model can be invalid because no provision may be made for different degrees of substitutability or complementarity among the choices. To test for the IIA property one must employ a specification test in the spirit of Hausman and McFadden (1984).

We apply a generalised Hausman specification test procedure²³ to check for the IIA in all our models, and find that at 5% significance level we cannot reject the hypothesis that the IIA assumption holds for our sample and a model with the three labour market states, i.e. employment, unemployment and inactivity.²⁴ In order to test our models for

²³If the IIA assumption holds there will be no systematic change in the coefficients if one of the outcomes is excluded from the model. The Hausman test is based on estimation of the variance of the difference of the consistent (when all choices are present) and efficient (when one or more outcomes are left out) estimators by the difference of the variances. The latter, however, is not necessarily positive definite in finite samples, so that the Hausman test could well be not identified. Thus, we need to employ a generalised Hausman test that applies to parameter estimates and associated covariance matrices combined into a single parameter-vector and a simultaneous covariance matrix of the sandwich/robust type.

²⁴It turns out, however, that the IIA assumption is likely to be violated if one tries to fit to the RLMS data a multinomial logit model for transitions from employment with four labour market states, namely current employment, new employment, unemployment and inactivity. The out-of-the-labour-force choice turns out to be not irrelevant at 5% confidence level in all samples containing data up to 1996 (the state of inactivity passes the Hausman specification test for IIA in pooled samples from 2000-2002). It is also interesting that it is not the independence of choice between inactivity and unemployment which is violated, as one could expect, but rather the independence of choice between inactivity and new employment (exclusion of these two choices from the model passes the Hausman test at 5% significance, whereas exclusion of inactivity and unemployment does not). Two implications ensue. First, it is likely that any previous attempts to estimate multinomial logit models for transitions from employment, when new employment and the out-of-the-labour-force state are jointly present, are not valid, at least if fitted to early rounds of the RLMS data (for an example of such a study see Foley, 1997b). Second, if not a peculiarity of the data set, the fact that inactivity and new job choices are interrelated for the currently employed may lend some support to the idea that being inactive is also a productive state, for example, informal work. Clarke *et al.* (2000), for instance, find that informal subsistence production is a complement rather than an alternative to paid employment.

behavioural distinctions between unemployment and inactivity we apply Wald tests to check for equality of coefficients in equations governing transitions from employment to unemployment and inactivity, and transitions from two non-employment states to employment. Test χ^2 -statistics with 21-25 degrees of freedom yield values ranging from 75.88 to 267.64, i.e. the null hypothesis of artificial distinctions in the two non-employment states is comfortably rejected.

Findings

The findings for the determinants of exits from the three labour market states are presented below and combined according to the effects on groups differed by age, gender and marital status, education, participation in irregular activities, and job as well as local labour market conditions.

Age differences For currently employed individuals age does not appear to be a significant factor in influencing the probability of becoming unemployed. However, young persons and people close to retirement may well be likely candidates for leaving the labour force - an inverse quadratic (or U-shaped) effect is clearly confirmed for transitions from employment to inactivity (tables 2.9 and 2.10). This is supported by the evidence on shifts between unemployment and the out-of-the-labour-force state, while the age effect on exits from inactivity is significantly concave (table 2.11). Age may also be viewed as a proxy for potential experience. From this perspective our results could suggest that while the market attaches some importance to potential knack, it does not value people with much of a knowledge gained in the pre-reform period - a point of view popularly mooted in the literature on economics of transition.

Also, young and post work age individuals are found to affect transition probabilities of their family members: the presence of the underaged and elderly in a household exerts a significant positive effect on the probability of movement from employment to inactivity (table 2.9, columns 4-5 and 8-9).

Gender and marital effects It seems that single employed men were hit hardest by the roaring 1990s, as both married males and females in general are found less likely to move from employment to non-employment. This is in line with the EBRD (2000) observation that in Russia as well as in large Central European countries little evidence has been found of a decline in female participation in the labour force, while many male workers left the labour force altogether after the start of transition (see also Standing, 1998; Boeri, 2000b). As regards transitions to employment the evidence does indicate that women are generally less likely to find a job than men (see also Smirnova, 2003a). Recent evidence from other sources also suggests that women are significantly less likely than men to engage in job search, they lag significantly behind men in search intensity, and significantly differ from men in their search strategies (Smirnova, 2003b). In addition, what appears to make the difference in the probability of getting employed is marital status: married individuals have better chances of landing a job if unemployed, while married males are better off in job hunting if inactive (table 2.11).

Skill differences The pattern clearly traced from tables 2.9-2.11 is that of the one favouring employability of high-skilled individuals. They are less likely to leave employment for non-employment and more likely to become employed if unemployed.

Meanwhile, it is probable that the low skilled are most prone to exiting the labour

force. The effect on probability is significantly positive in model E1, and also in its versions fitted separately to the RLMS Phase I and Phase II data (which was done as an extra exercise and is not shown in the tables). At the same time, the effect is not significant in models E2 and E3 (tables 2.9 and 2.10), where a variable for tenure in the current job, and dummies for the previous incidence of wage arrears and unpaid leave are added in. It seems that the insignificance of the coefficient on the dummy for the low-skilled is caused, on the one hand, by the fact that wage arrears in general are more widespread among low-skilled workers (see Earle and Sabirianova, 2002; Namazie, 2002) and, on the other hand, by complementarity between the practice of wage arrears and the incidence of compulsory unpaid leaves (see Namazie, 2002, for some evidence from the FSU economic space). By taking out the two introduced dummies, but leaving the tenure variable in the model, the significant coefficient on the low-skilled dummy is restored (however, it remains insignificant if only one of the dummies for the incidence of arrears or unpaid leaves is eliminated - again, all this was done as a separate exercise and is not presented in the tables). In all cases both the arrears and the unpaid leaves seem to be very significant determinants of exits to inactivity (see tables 2.9 and 2.10, and further facts below).

When comparing chances of leaving the out-of-the-labour-force state, the low-skilled are less likely to do that (table 2.11).

Taken together, these findings support well one of the stylised facts about economic transformation in Eastern Europe, namely that the transition has unequivocally brought in an increase in returns to a year of education. Sabirianova (2003) points out that such skill-bias changes in employment do not permit a uni-causal and time-invariant explanation.

Irregular activities Irregular activities, i.e. episodic work and family land use, are found to be important factors in explaining mobility in the Russian labour market.

The existence of a secondary job may well be a factor significant for leaving employment. Table 2.9 (model E1) indicates that having a secondary job/IEA has a positive and significant effect on the probability of exit from employment to both unemployment and inactivity. An intuitive explanation for this finding that may be suggested is that having a secondary employment opportunity, often more lucrative than the primary workplace (Roshchin and Razumova, 2002), could be an incentive to quit employment. However, the coefficient on the dummy for secondary work/IEA is not significant in models E2 and E3 fitted to the data from the RLMS Phase II. Moreover, the marginal effect on the probability of exit into non-employment states turns negative in comparison to model E1 (although it remains insignificant). This, as in the case of the low-skilled dummy above, is due to the introduction of the variable for tenure, and the two indicators of wage arrears and the experience of compulsory unpaid leave at respondent's primary place of work (only without all the three variables an additional experiment of fitting E1 model separately to Phase I and Phase II data produced a significant positive coefficient on the SECJBIEA dummy for exits to both non-employment states - results are not presented). The intuition behind these observations is twofold. On the one hand, it is probably linked with the effect of tenure, and the incidence of unpaid leaves or/and wage arrears on labour supply in the secondary job market. On the other hand, the explanation may also be sought in simultaneity of worker's decision to change jobs and supply efforts in the secondary market (see, e.g. Paxson and Sicherman, 1996, for insights from the US labour market). As regards the former remark, Kolev (1998) finds that holding an additional

job in the informal economy appears to be a safety valve for individuals rationed in the regular labour market, either unemployed or experiencing compulsory periods of unpaid leave. Also, Roshchin and Razumova (2002) reason that secondary employment in Russia mostly depends on worker's intent to compensate insufficient wages received from the main job, so that wage arrears and unpaid leaves increase supply in the secondary job market through an implicit impact on the income effect. None of the authors, however, investigate the effect of tenure on the labour supply in the extra employment market. As far as the interlink between mobility and dual job holding is concerned, Popov (1999) has noted that in Russia worker shifts from one job to another may be caused not by wages but by opportunities for incidental non-wage earnings. Roshchin and Razumova (2002) provide some supporting evidence in favour of the positive effect of secondary job holding on the decision to change employment, but not quitting it. In view of these results, the negative marginal effect of secondary job holding/IEA in tables 2.9 (columns 7 and 9) and 2.10 (columns 3 and 5) on the probability of leaving employment for non-employment is not surprising.

For the non-employed the involvement in irregular productive activities seems to facilitate getting a job as suggested by the positive effect on the probability of employment in models U1 and I1 (table 2.11). This may imply that irregular IEA provides those involved with network contacts useful for obtaining regular employment. In any case, it is again in line with the idea that labour supply in the irregular market and search for another (regular) job are mutually inclusive.

As regards family access to land, it is not found to be important for transition into employment and unemployment, but some evidence points to its positive influence on

transitions into inactivity. The effect is especially strong for exits from unemployment: land use raises the chance of becoming inactive by nearly as much as 7.6% (table 2.11, column 5). At the same time for the employed the marginal effect of having access to land is less than 1% in model E1 (table 2.9, column 5) and not significant in models E2 and E3 (table 2.9, column 9, and table 2.10, column 5). Already familiar caveats may still apply here as to the effect of the incidence of wage arrears and unpaid leave: it seems that these two phenomena and the decision to take advantage of available land are interconnected (elimination of the two dummies from models E2 and E3 improves significance of a coefficient on the land use variable in equations for transitions in inactivity). Additional experiments with fitting models E1 and E2 separately to, as appropriate, panels from the beginning, the middle and the end of the decade in focus, revealed that land use is only significant for shifts from employment in inactivity in samples representing the middle of the 1990s (results are not given). All in all, the data do support our hypothesis that land use is an important factor in the decision to non-participate.

As we have noted in the introductory sections, the other indicator of informal land use can be dacha ownership. However, it is found to be negatively associated with the probability of moving into inactivity nearly throughout the whole period studied. This may seem surprising given that subsistence farming on dachas has traditionally been widespread in Russia. One of possible interpretations of this result may be the endogeneity of dacha ownership to labour market position. In other words, people may sell dachas if they are in difficulties, i.e., for example, when they become non-employed. While it may as well be the case, however, we can think of an objection to this argument. As has been noted above, our logit samples include individuals who were surveyed twice -

both at the beginning and at the end of a year-long period. The information about dacha ownership is taken as of the moment of the first interview of each individual, but not as of the second one. Thus, our regressions are designed to show the effect of dacha possession on the probability of change of the labour market state in the following year, but not the reverse effect. That is, the negative effect of dacha ownership on the probability of moving into inactivity is probably not due to the endogeneity problem. Instead, another interpretation of the evidence may be that dachas, as popular week-end resorts, are not an indicator of means of subsidiary production, but rather of affluence of a respondent's family. This would accord well with Clarke *et al.* (2000) who point out that "dacha use is a leisure activity of the better-off rather than a survival strategy of the poor." Note, however, that *ad interim* we found nearly no significant impact of household income on mobility. So, the strong negative effect of having a dacha on the probability of moving out of the labour force is still difficult to explain by means of the traditional income effect on labour supply. Perhaps more research needs to be done on that aspect.

Job and local labour market characteristics From the results considered in the preceding subsections it should become obvious that such characteristics of job and/or local labour market as wage arrears and unpaid leaves have been important determinants of worker mobility. Indeed, incidence²⁵ of both has as a rule a positive effect on the probability of exit from employment to inactivity (tables 2.9, right panel, and 2.10).

By bringing all the facts mentioned thus far together one could intuitively think of the following explanation for the role of rationing in the regular labour market in increasing

²⁵Wage arrears were especially widespread in Russia before 1998 when the proportion of working age population that had experienced arrears reached 63.80% in our sample (round VIII). In RLMS round XI, corresponding to 2002, the proportion drops to 22.22% - still a substantial number. The dynamics of the incidence of unpaid leaves is somewhat similar: in samples drawn before 1998 the average of more than 8% of working age population has been sent on compulsory leave; by 2002 this figure drops to 2.71%.

worker flows into inactivity. Most likely, the reduced working schedules and wage income force workers to seek extra employment (as findings by Roshchin and Razumova, 2002, suggest), and increase labour supply in the irregular labour market. In rural areas with especially rich possibilities for subsistence agriculture it causes worker flows out of the labour force. This is consistent with observations that arrears and forced administrative leaves hit harder the low-skilled workers, especially in rural regions (Earle and Sabirianova, 2002; Namazie, 2002). Moreover, there is evidence that arrears themselves are enabled by worker survival strategies that include subsistence farming²⁶ (Earle and Sabirianova, 2002). At the same time, it is also been argued that unlike wage cuts the arrears have a theoretically ambiguous effect on worker's quit behaviour and empirically their effect varies negatively with the extent of the practice of the arrears in the local labour market (*ibid*). Finally, it should be mentioned that wage arrears are very likely to reduce *regional* mobility of workers by increasing worker's inability to meet the costs of moving (for insights see Friebe and Guriev, 2000; Andrienko and Guriev, 2003).

As regards other job characteristics and indicators of economic conditions, we find that such factors as longer job tenure and provision of non-wage benefits/perquisites are significant in lowering the probability of leaving employment. The former fact ties in with Lehmann and Wadsworth (2000) who observe that labour mobility in Russia declines as experience and tenure grow (although its level remains still quite high if compared to, e.g. Poland). The latter finding is interesting in itself as confirms previous speculation on the role of fringe benefits in reallocation process. Russian firms, especially in the state sector, have had a long history of providing a significant range of non-monetary benefits

²⁶In the other study Earle *et al.* (2004) provide strong support for significance of local community effects on firms adopting and workers tolerating the practice of wage arrears.

(kindergartens, subsidised housing, etc.) that were not only quite large, but were also an important factor in motivating workers (Commander *et al.* 1996). Such benefits comprise a good proportion of the total remuneration of workers (Standing, 1998). Lehmann and Wadsworth (2000) have pointed to the fact that as a legacy from the Soviet regime, characterised by wrong and under-utilisation of labour, while labour markets were in the state of excessive demand, fringe benefits were an important factor in competition for workers during their reallocation when wrongly placed labour started to move. They guessed that this pattern of turnover may persist in the early stages of transition. Our findings, pertaining to the brink of the new century, indicate the pattern is still strong 10 years down the transition road.

Another local labour market effect worth mentioning indicates that residence in the metropolitan areas (represented here by Moscow and St.Petersburg), while not a significant factor in affecting exits from employment, has a substantial positive impact on entering employment from the out-of-the-labour-force state (model I1).

Finally, further considerations suggest that the time pattern of transition probabilities evident from table 2.5 is well explained by changes in the local unemployment and non-employment rates (see tables 2.9-2.11 for the effects on transition probabilities, and Fig.2.2 and 2.3 for the insights into general dynamics over time). In particular, probabilities of separation into unemployment increase with the unemployment rate, while probabilities of exit out-of-the-labour force rise with an increase in the non-employment rate. Interestingly, in the models with shifts between unemployment and inactivity (U1 and I1) a rise in the unemployment rate implies a decrease in the probability of moving or staying out of the labour force. On the one hand, this observation is consistent with the idea that shocks

having an impact on local markets may be pervasive so that both regular and irregular economies are affected. In that case, if the inactive resort to informal work, a decline in both the formal and informal sectors would lead to growing unemployment and, hence, a positive observable effect of the unemployment rate on exits from inactivity. On the other hand, existence of an "unemployment culture" or an "inactivity culture" (i.e. local community norms), can explain the same patterns. We have also found, however, strong evidence pointing to the existence of the discouraged worker effect. It manifests itself in a positive marginal impact on the probability of exiting to the out-of-the-labour-force state of duration of unemployment. In general, the probability of leaving for inactivity is lower, while the chances of getting a job are higher, when unemployment duration is shorter (see table 2.11).

Comparative summary and some implications

The winners and losers of Russian transformation in many aspects appear to be among the same contingents that have fared more or less well in other transitional countries. So, for example, high skilled workers in Russia have better chances of both staying in employment and strong advantages in job hunting as is also true for the Czech Republic (Sorm and Terrell, 1999), Estonia or Slovenia (Vodopivec, 2002). Also, similarly to these studies we confirm that females seem to be less mobile than men (see also Smirnova, 2003b), although our work suggests that a lot depends on their marital status. In particular, as in the Czech Republic and Slovenia, this research does not find that women face a higher chance of moving from employment to inactivity, the fact reported to be the case in Estonia (Vodopivec, 2002). Another parallel can be drawn for persons having access to land in Russia who are more likely to exit the labour force - this finding is en rapport

with results for other countries, such as the Czech Republic (Sorm and Terrell, 1999) and Romania (Voicu, 2002). However, some differences do certainly exist. So, for instance, we do not find any age effects on reemployment probabilities of the unemployed as opposed to Sorm and Terrell (1999) for the Czech Republic and Vodopivec (2002) for Estonia, who point out that young workers have better chances of reemployment than old workers. The mobility effect of wage arrears and unpaid leaves, peculiar to the Russian labour market, is another distinct point.

What can one gather from these and other findings presented above?

First of all, labour reallocation in Russia has not been uniform. The initial indication that this is the case we have seen while looking into the static distribution of our samples across the labour market states in Section 2.3.1 (see tables 2.2 and 2.3). Now the results of our multinomial logit analysis as well as more evidence from other sources give more support to the fact that certain groups of individuals are to be disproportionately represented among the unemployed and the inactive, while others are to enjoy the state of being employed. This is particularly true of women and the low-skilled who are more likely to stay jobless than men and the better educated. On the one hand, it may imply that reallocation of labour in different sectors of the economy could well have been different from the U-shaped pattern predicted by the basic theory of transition (Aghion and Blanchard, 1994). On the other hand, it suggests either that the job search intensity or strategy varies across particular contingents, or that employers' policies towards those groups are likely to be differing as well. We are not in a position to comment on the structural issues here as the RLMS has no information on sectoral affiliation of respondents. At the same time, relevant to the guess about search behaviour, Smirnova (2003b) and Roshchin and

Markova (2004) using the same RLMS data set find that job search strategies do vary across different cohorts of individuals. Similarly, the prevalence of wage arrears and unpaid leaves among particular groups of workers, as found by Earle and Sabirianova (2002) and Namazie (2002), supports the hypothesis of non-uniformity of employers' policies.

The possibility that wage arrears and unpaid leaves affect both the labour supply in irregular markets and mobility of workers, especially out of the labour force, is a second important point to notice. At the same time, somewhat related to this is the observation that non-wage benefits are a significant factor in preserving employment. The theoretical models of transition as the labour reallocation from less to more productive firms, mentioned in the introduction, devote hardly any attention to both the wage-setting or remuneration mechanisms in the contracting sector and labour supply. Thus, our findings suggest that this is likely to be a serious drawback that should be corrected in future theoretical investigations.²⁷

Finally, findings as regards the mobility effect of land use do indicate that the inactive are likely to be involved in the irregular or informal economy, at least in its part represented by subsistence activities. However, it is probably not an alternative to regular employment: having an irregular activity is a significant factor in becoming an active job seeker and it is conducive to moving out of the non-employment states. Similarly, the possibility of moving out of the labour force may be not an irrelevant alternative when changing employment - some grounds for that are provided by logit specification tests (see, in particular, footnote 24). One way or another, this calls for more research into the role of

²⁷Other relevant evidence supporting such a conclusion is the finding that probability of arrears is negatively associated with measures of firm performance, with forms of private ownership and recent founding date of the firm (see Earle and Sabirianova, 2002). At the same time, many firms that resort to wage arrears are simultaneously hiring new employees and expand their business (Earle *et al.*, 2004).

the informal sector in the reallocation of labour.

2.3.4 Survival Analysis

The multinomial logit analysis has provided us with the insights as regards the issue of non-uniformity of labour reallocation in transition. It has also helped identify some determinants of movements in and out of non-employment. The final question that we would like to take up in this chapter is concerned with the increasing stagnancy of both unemployment and inactivity - a somewhat surprising fact given substantial fluidity of the Russian labour market (see Section 2.3.2).

As has been mentioned before in this essay, the theoretical literature on transition can explain neither the persistence of unemployment widely observed in transitional countries nor an increase in unemployment duration - a closely related issue. Table 2.7 illustrates that although the unemployment pool in Russia has become a little less stagnant recently, still it is less dynamic than at the beginning of the 1990s - the proportion of unemployed 6 months and more increased by about 13.5% between 1992 and 2002. The pattern is also observed for the inactive, although in this case the changes are less pronounced.

In this section we employ survival analysis of non-employment duration to gain some insights into possible determinants of the problem. To this end we use a competing risks framework with the original state being either unemployment or inactivity, and two destination states being employment and another non-employment state. In particular, we draw on the idea that stagnancy of the non-employment pool and, for that matter, an average increase in the duration of non-employment may be caused by the existence among the non-employed of a significant number of individuals who face the possibility of

infinite durations of non-employment. This possibility in relation to the Portuguese labour market has recently been explored by Addison and Portugal (2003). They identify two sources of infinite durations. The first is when the infinite duration of joblessness results randomly and possibly for everyone. In other words, all the non-employed face a non-negative probability of being stuck in the jobless state forever due to some peculiarities of labour demand or general processes in the economy. As a second possibility, there may exist workers who reject certain opportunities arising in the labour market, and thus continue being non-employed in spite of otherwise possible change. We elaborate on these two possibilities in the following section.

Random draws and defective risks

Long-term non-employment is likely to be provoked by many factors such as structural change in the economy, business cycle effects, leading to changes in the natural rate, changes in reservation wage, peculiarities of search behaviour of jobless workers, employers, etc. Here we explore the possibility that stagnancy of joblessness is observed in Russia either because of workers facing a specific matching irregularity or because of the presence of defective risks generated by workers' attitude to inactivity (or both). The former may be caused either by specific hiring rules, formally or informally followed by employers, or by abrupt deterioration of workers' human capital in the process of long-term search. The latter type of defective risks is due to workers rejecting an opportunity for moving out of the labour force (if they are unemployed) or becoming an active job-seeker (if they are inactive), while their personal characteristics may anyway rule them out as candidates for getting a job. Following Addison and Portugal (2003) the two possibilities can be represented by the models of "random draws" and "split population", respectively.

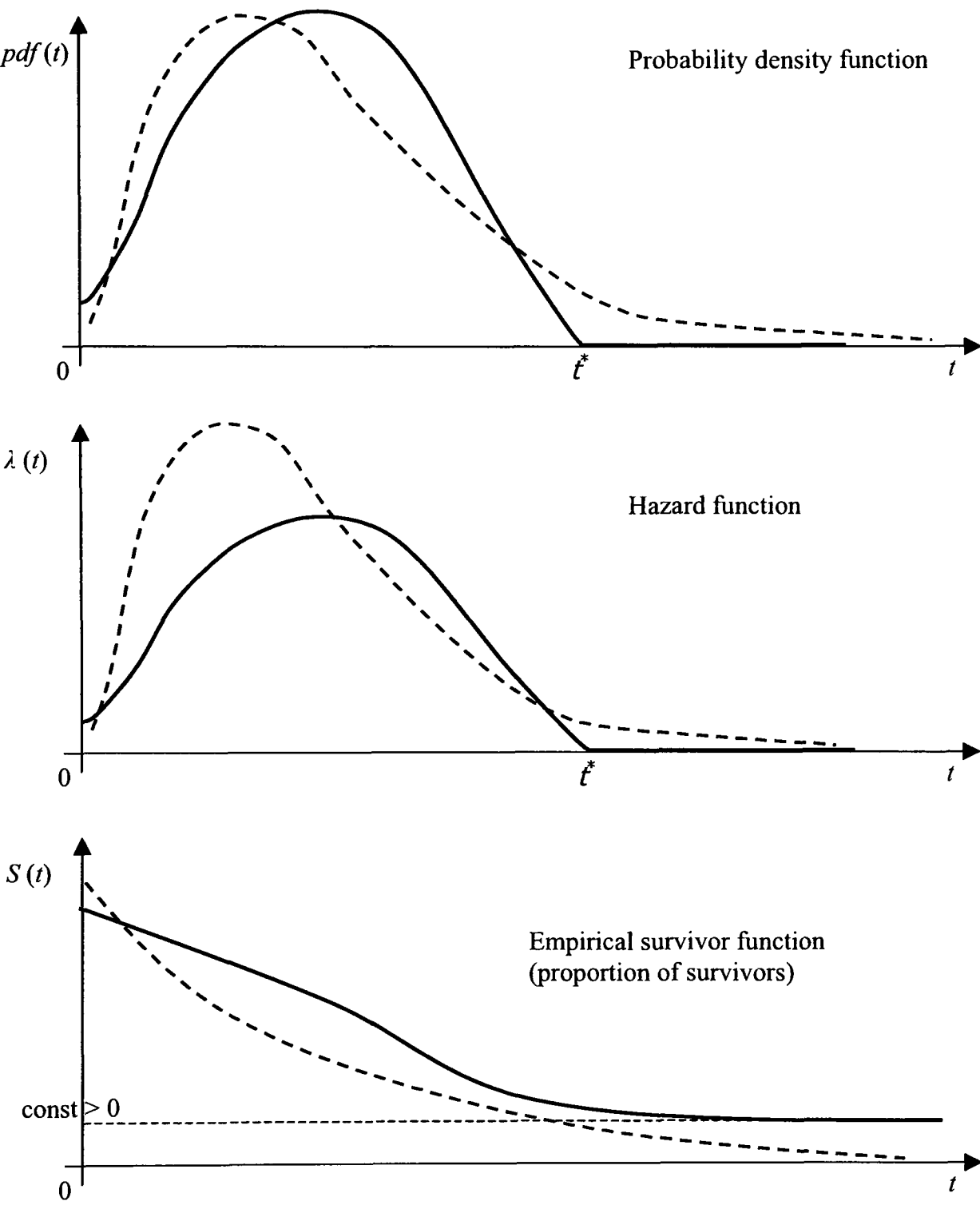


Figure 2.4: Random draws model

The random draws model The model of random draws is build around the assumption that the exit rate from a non-employment state is a rapidly decreasing function of duration (at least for some non-employment spells especially long). More specifically, consider, for example, unemployed workers. It is possible to think of a situation where the hazard of leaving unemployment may have a very specific distribution over the lengths of unemployment spells (see solid line graphs in Fig.2.4). For such distributions, the probability density function has a flat singular segment beyond some finite duration t^* (Fig.2.4, top panel). As a result, the hazard function is to cross the time axis at t^* (Fig.2.4, middle panel), so that workers with durations longer than t^* face zero probability of becoming employed. Equivalently, over the same duration interval the survivor function does not converge to zero, but instead to some constant positive value, and maintains it indefinitely thereafter (Fig.2.4, bottom panel). The intuition for the shape of the PDF or the hazard rate, for that matter, in Fig.2.4 can easily be found in economic theory. In particular, the theory suggests that an unemployed person leaves unemployment if she receives a job offer, and if she finds that offer acceptable (Jenkins, 2004). Then for a given worker the exit rate from unemployment $\lambda(t)$ can be written as the product of the job offer hazard $J(t)$ and the job acceptance hazard $A(t)$:

$$\lambda(t) = J(t) A(t). \quad (2.1)$$

If one of these functions is flat and singular beyond t^* then a situation like in Fig.2.4 results. Such distributions can emerge naturally in a variety of situations (for some discussion see Addison and Portugal, 2003). It is likely to be the case when, for example, employers apply a screening rule, as to which they never hire workers whose jobless spell exceeds

some length. Then $J(t)$ is likely to be such that $J(t) \xrightarrow[t \rightarrow t^*]{} 0$ and $J(t) = 0$ for $t \geq t^*$. This is very consistent with a dynamic model of matching in the labour market where employers rank applicants according to the time that they have been unemployed (see Blanchard and Diamond, 1994). Alternatively, laid-off workers may completely lose their skills while unemployed. If, at the same time, they do not adjust their reservation wage, which affects the job acceptance hazard $A(t)$, they will no longer find an appropriate job. Such deterioration of skills may happen in finite time, so that workers who spend in unemployment a longer period are effectively bound to be trapped in joblessness. This is especially likely during periods of substantial economic adjustment, when some sectors, jobs, occupations, etc. cease to exist. This may well be appropriate for the countries of economic transition in Eastern Europe, where old inefficient enterprises are liquidated, whereas emerging private sector firms demand a completely different set of skills from the labour they employ (for some insights see, e.g. Ljungqvist and Sargent, 1998). In both cases considered, as a result of chance any individual can fail to draw from a finite segment of the duration distribution in Fig.2.4, so that she is unlucky in her draws and has a non-zero probability of ending up in a spell that lasts forever.

The split population model The split population model presented here is effectively a generalisation of the standard mover-stayer model (see, e.g. Schmidt and Witte, 1989, for an early econometric application, or Pudney and Thomas, 1995, for a similar multi-state extension) to more than one exit state from non-employment. Our exposition follows closely Addison and Portugal (2003). Consider again the unemployed. Some of these workers will search for jobs and also not exclude the possibility of labour market withdrawal. At the same time, some other proportion of the unemployed can for some reason

		Inactivity	
		I	$\bar{I}$
Employability	E	Employable, Consider inactivity 2 Movers	Employable, Do not consider inactivity 3 Movers
	$\bar{E}$	Unemployable, Consider inactivity 1 Movers	Unemployable, Do not consider inactivity 4 Stayers

Figure 2.5: Types in the split-population model

consciously rule out the option of leaving for inactivity. Some of the workers will have characteristics that sooner or later bring them an acceptable job offer. Others, due to their skill endowments or other qualities, will never be able to secure employment. Thus, the population of the unemployed workers can conveniently be split in four parts and presented in a matrix as in Fig.2.5. The vertical axis denotes the population of employable and unemployable workers marked by E and $\bar{E}$, respectively. Employable individuals possess observed or unobserved characteristics that leave them a chance of getting a job in due course; unemployable individuals do not have such a chance and will never transition from unemployment to employment. The horizontal axis signifies the population in relation to their attitude towards inactivity. Part of the population denoted by I admits inactivity as a possible destination. These people, if get discouraged by unsuccessful job search, will sooner or later withdraw from the labour force. The people denoted by $\bar{I}$ never consider inactivity as an option and, thus, never leave unemployment for the out-

of-the-labour-force state regardless of the successfulness of their job search. People in the first, bottom-left cell of the matrix, are unemployable, but consider possibility of labour market withdrawal, so that they will ultimately move out of unemployment. People in the second, top-left cell, may find employment or instead move into inactivity. In both cases these cohorts represent *movers*. The hazard function of exits from unemployment to inactivity to be considered is correctly estimated over these two groups of jobless workers. At the same time, the third, top-right cell represents individuals who will ultimately find employment, but never consider leaving for inactivity. They are again *movers*, and the hazard function of transitions from unemployment to employment is properly estimated across the first and the third cells. Finally, the fourth cell denotes workers who are not employable and never consider inactivity. Those are the *stayers*, the long-term or rather permanently unemployed. Thus, to stay in unemployment infinitely long an individual has to be at two defective risks: one that prevents her transition into employment, and the other one that excludes her transition into inactivity. Addison and Portugal (2003) note that if the latent hazards of exiting to two destination states are independent, or, in other words, the competing risks of leaving unemployment are independent, the probability of being trapped forever in unemployment is just the product of the probability that an individual is unemployable times the probability that she never considers inactivity.

At this point it should be noted that this type of competing risks model is especially applicable to a situation where the same risks are not relevant for every individual. As emphasised by Pudney and Thomas (1995), such a case is made in a multi-sector search model, when persons never search for a job in a particular sector. It may happen if, for instance, a person's "normal" sector of employment ceases to exist because of industrial

contraction (*ibid*). Thus, this type of model is relevant to the Russian economy (also see above a reference to the work by Ljungqvist and Sargent, 1998). On the one hand, it is precisely the environment generated by the transition from plan to market. On the other hand, as we can gather from the preceding empirical analysis, many inactive workers are likely to be employed in the subsistence sector. Thus, the split-population model set up in this section and applied to the unemployed, effectively implies that some workers may not search for an opportunity in the informal economy. If, however, they do not possess characteristics that make them employable in the regular sector they risk to stay permanently jobless.

Methodology

In this part of the essay we aim to test for the presence of defective risks such as those that may result either from the process of random and unlucky draws or from representing a particular cohort of population.

In order not to overburden the exposition, all the details of the econometric models, peculiarities of their estimation, and discussions of assumptions made are relegated to Appendix 2.B. Here we just present a few essential points needed for understanding of our approach.

As we have mentioned above we use a standard competing risks approach (for reviews see, e.g. Van den Berg, 2001, and Jenkins, 2004) to estimate reduced form duration models with hazard functions of exits from each non-employment state to employment and another non-employment state. We assume that the hazard functions are proportional, i.e. they can be represented as the product of a baseline hazard, which depends solely on the duration of jobless spell, and a scaling function that includes as its arguments a

vector of personal characteristics and other factors (for a discussion of this popular type of specification for the hazard function see Van den Berg, 2001; Devine and Kiefer, 1991, provide references in labour economics).

Crucial to determining the applicability of the random draws model to our case is the shape of the baseline hazard. If we are dealing with a situation in which a specific distribution of spell durations may arise, then what we want to know is the model's behaviour in the right tail of the duration distribution. Then it seems reasonable to estimate models with the following two types of specification for the hazard.

In the first model, the baseline hazard is assumed to take a polynomial form. It is in general a flexible form for the hazard that allows various representations (i.e. constant, increasing or decreasing in duration) and does not *a priori* restrict the model by preimposing a monotonic type of time dependence in the hazard. With a negative coefficient on the highest power it may generate a function crossing the time axis just like in Fig.2.4 (middle panel). This is precisely the method that is used by Addison and Portugal (2003). However, if the model is set up in *continuous* time (as in their study) the approach may run into some serious computational difficulties (see the appendix for a discussion). At the same time, by modelling a *discrete* time baseline hazard as a polynomial we still can identify the potential for a specific distribution. As has already been mentioned, the RLMS data on duration of joblessness have been grouped into monthly intervals, therefore suggesting that modelling is to be made in discrete time. Thus, the test for the unlucky draws model is simply based on the shape of the discrete time baseline hazard.

An alternative to the polynomial specification is to assume that the discrete hazard is piecewise constant. The advantage of this specification over the polynomial one is as

follows. Suppose that in the sample there are many more observations of shorter duration times than those of longer duration times. In other words, there is some clustering for low values of duration. In such a situation coefficients of the polynomial fitted to the data, and, hence, its shape, may be influenced mainly by short duration data. If the clustering is especially strong, longer duration observations have little chance to affect the estimated coefficients. Assuming the piecewise constant hazard helps to avoid the problem caused by the potential clustering.

In order to test for the split-population scenario we use a method of correction for unobserved individual heterogeneity suggested by Heckman and Singer (1984). It is assumed that all the workers in the sample come from a number of different homogeneous subgroups (in our case, four subpopulations corresponding to cells in Fig.2.5), and for each person the probability that she belongs to a particular group is estimated. Thus, we obtain the probabilities that a person is either a mover or a stayer, and can work out if she considers or not inactivity or unemployment as an option. To avoid negative values in the process of estimation and ensure that each probability of being a mover to destination j is between 0 and 1, it is represented as $P^j = \frac{\exp \mu_j}{1 + \exp \mu_j}$, where j stands either for employment or inactivity/unemployment. So, in fact it is the parameters μ_j that are estimated. Such a specification does not affect the ability of the model to find evidence of stayers and movers, for it does not preclude P^j from being as close to one or zero as required.²⁸

Finally, to control for other sources of unobserved individual heterogeneity, omitted

²⁸Foley (1997a) estimates a somewhat similar mover-stayer model for Russia using early rounds of the RLMS data and the Heckman-Singer (1984) strategy for correcting for unobserved heterogeneity. He uses the same logit link function to specify the probability of being a motivated mover to one of the destinations.

variables, and so forth, we add a random multiplicative error term into the specification of the proportional hazard. Such a correction for unobserved individual effects is often referred to as control for "frailty", by a tradition originating in biostatistics (see Vaupel *et al.*, 1979). The random error term is, as is conventional, assumed Gamma distributed independently of time and the vector of individual characteristics and other determinants scaling the baseline hazard. The parameters of the distribution are assumed to be 1 (mean) and σ^2 (variance). The last is treated as a unknown parameter of the model to be estimated. Again, to prevent negative values of σ^2 in actual estimation we employ a logarithmic transformation of the variance.

All the models specified in the next section are estimated using the maximum likelihood method. Estimation is carried out in Ox (Doornik, 2001) using the quasi-Newton BFGS (for Broyden, 1969, 1970; Fletcher, 1970; Goldfarb, 1970; Shanno, 1970) secant method to find a maximum of the likelihood function. Starting values are obtained from fitting a simple single risk (unemployment-employment) specification. Initial values of the logarithm of the frailty parameter and proportions of movers are taken as -1 (i.e. $\sigma^2 = 0.37$ as in Jenkins, 1998) and 0.5 (i.e. $\mu = 0$). Asymptotic standard errors for estimated parameters are obtained by inverting the Fisher information matrix numerically estimated by the outer product of scores.

Specification

Six types of model are estimated for exits from both unemployment and inactivity. In the first set of three models the discrete baseline hazard is assumed to take the form of a fourth order polynomial.²⁹ Use of polynomials of higher degrees produced poor

²⁹Flinn and Heckman (1982) were the first to discuss the use of a polynomial discrete time baseline hazard. Foley (1997a) and Addison and Portugal (2003) employed precisely fourth order polynomials in

convergence results of the BFGS method. The second set of models differs from the first one by assuming the interval baseline hazard to be piecewise constant. In this case the duration time scale is divided into five intervals with knot points 12, 36, 60, and 84 months. Such a choice of rather dispersed values of non-employment duration is explained by our primary interest in the behaviour of the hazard for very long spells. Thus, we estimate the hazard for spells less than 1 year, three subsequent intervals of duration taken with a 2 year step, and the final, open-ended interval of durations greater than 7 years. The total number of parameters characterising time dependence of the discrete baseline hazard is thus the same as in the polynomial case.

As regards the person-specific non-negative function of covariates (scaling the baseline hazard), the set of regressors included in it contains most of the variables used previously in the multinomial logit analysis in Section 2.3.3. In particular, in all the models independent variables include characteristics of the individual (such as age, the gender dummy, the dummy for being a married male, the skill level³⁰), of the household (household income, access to land, dacha ownership), and of economic conditions (such as indicators of involvement in individual economic activity and belonging to the metropolitan labour market, unemployment and non-employment rates³¹). It should be noted that for all potentially time-varying covariates only the values as of the moment of first sampling as

competing risks frameworks.

³⁰We attempted to estimate models with other covariates as well. For example, separate indicator variables for single and married females were tried. However, the coefficients on these variables were often insignificantly different both from each other and from zero. Later these binary variables were substituted for the female dummy. For similar reasons we completely excluded from the models the variables for the presence of children and the elderly in the respondent's household.

³¹As in the case with the multinomial logit models, controlling for the unemployment and non-employment rates appears to do the same job as the introduction of year dummies. Thus, in all the models we have preferred not to use indicators for the rounds in which an individual was first sampled as non-employed. At the same time, in other work using the pooled data from RLMS Phase II, Smirnova (2003a) estimates a structural duration model and does not find coefficients on dummies for RLMS rounds to be significant.

being non-employed are taken³² (i.e. the value of age, for example, is taken as of the moment that a person first appears in our sample as non-employed). Also, we do not include any variables for unemployment compensation - a factor traditionally receiving special attention in duration analysis. From table 2.8 it can be seen that the coverage by unemployment benefits is very poor in Russia, while the response rate to the RLMS question about unemployment insurance is low. This justifies the omission of this otherwise crucial factor.³³

Finally, for each of the two types of time dependence in the baseline hazard the models differ in regard to control for unobserved heterogeneity. The basic model (denoted by UD1 for transitions from unemployment, and by ID1 for transitions from inactivity) does not allow for any type of unobserved individual effects. In the model of a second type (UD2 and ID2) a Gamma distributed frailty factor is introduced, while in the model of a third type (UD3 and ID3) in addition to controlling for frailty the Heckman-Singer (1984) correction for unobserved heterogeneity is used to estimate proportions of movers to each particular destination. Findings are discussed below.

³²We realise that by so doing we discard a possible host of additional information available in the RLMS. However, we believe that it does not impair the results of our estimation as the primary focus is the time dependence of the baseline hazard and the estimation of the proportions of movers.

³³There is, however, mixed evidence on the effect of unemployment insurance on the duration of unemployment. For example, Boeri (2000) argues that it is the generosity of the compensation that has contributed to the growing stagnancy of the non-employment pool and has increased flows out of the labour force in many transitional countries of Eastern Europe (including, e.g., Poland). At the same time, Puhani (1999) does not find a change in the unemployment benefit system to have a significant effect on the duration of unemployment in Poland during the first few years of economic transition.

Table 2.12: Two-destination hazard regression models of transitions from unemployment (polynomial hazard)

Independent variable	Model UD1 Transition to		Model UD2 Transition to		Model UD3 Transition to	
	Employment	Inactivity	Employment	Inactivity	Employment	Inactivity
AGE	-0.0061 (0.004)	0.0275** (0.005)	-0.0061 (0.004)	0.0346** (0.008)	-0.0061 (0.004)	0.0314** (0.007)
FEMALE	0.0237 (0.100)	0.0069 (0.137)	0.0239 (0.100)	0.0092 (0.182)	0.0238 (0.100)	-0.0201 (0.150)
MARIDMAL	0.3188** (0.119)	-0.3266* (0.172)	0.3191** (0.119)	-0.4880** (0.238)	0.3191** (0.119)	-0.4387** (0.204)
HHINCOME / 10 ⁻³	-0.0166 (0.029)	0.0411 (0.028)	-0.0166 (0.029)	0.0520 (0.051)	-0.0167 (0.029)	0.0305 (0.042)
DSKILL1	-0.2508 (0.205)	0.2845 (0.233)	-0.2506 (0.204)	0.2849 (0.337)	-0.2507 (0.205)	0.2262 (0.266)
DSKILL3	0.1164 (0.096)	0.0631 (0.151)	0.1160 (0.096)	0.1036 (0.192)	0.1161 (0.096)	0.0209 (0.158)
LANDUSE	-0.1740** (0.088)	0.1805 (0.142)	-0.1741** (0.089)	0.3192* (0.185)	-0.1744** (0.089)	0.3019* (0.156)
OWNDACHA	0.2260** (0.100)	0.0045 (0.148)	0.2262** (0.100)	-0.0922 (0.185)	0.2264** (0.100)	-0.0591 (0.157)
SECJBIEA	0.1588 (0.105)	-0.2736 (0.167)	0.1584 (0.105)	-0.3449 (0.219)	0.1586 (0.106)	-0.2457 (0.183)
METRO	0.0160 (0.158)	-0.1215 (0.258)	0.0173 (0.158)	-0.1897 (0.333)	0.0173 (0.158)	-0.0612 (0.276)
URATE	-3.7294 (2.920)	-10.4080** (4.442)	-3.8632 (2.916)	-12.6540** (5.720)	-3.8262 (2.923)	-10.2730** (4.854)
NONRATE	-0.6130 (1.253)	5.2888** (1.884)	-0.5614 (1.252)	6.7439** (2.573)	-0.5664 (1.253)	6.1188** (2.201)
Constant	-3.0543** (0.321)	-8.7834** (0.740)	-3.0511** (0.322)	-9.4345** (0.929)	-3.0569** (0.324)	-9.1697** (0.858)
t / 10	1.6419** (0.350)	5.6724** (0.989)	1.6356** (0.351)	5.9706** (1.043)	1.6432** (0.352)	5.7104** (1.009)
t ² / 10 ²	-0.7903** (0.192)	-2.5682** (0.465)	-0.7863** (0.192)	-2.6037** (0.479)	-0.7910** (0.193)	-2.5778** (0.470)
t ³ / 10 ⁵	13.7250** (3.907)	44.1290** (8.470)	13.6370** (3.909)	44.3940** (8.716)	13.7390** (3.961)	44.1910** (8.594)
t ⁴ / 10 ⁷	-7.8327** (2.601)	-25.2860** (5.190)	-7.7727** (2.603)	-25.1990** (5.326)	-7.8417** (2.669)	-25.0270** (5.286)
Frailty (log σ ²)	--	--	-11.5520 (1496.900)	-0.4400 (0.701)	-12.7230 (467.450)	-6.0759 (152.120)
μ	--	--	--	--	11.4290 (155.240)	3.0651** (0.740)
Frailty (σ ²)	--	--	0.0000 (0.014)	0.6440 (0.452)	0.0000 (0.001)	0.0023 (0.350)
Proportion of movers	--	--	--	--	1.0000** (0.002)	0.9554** (0.032)
Log-likelihood	-4616.89		-4615.31		-4614.06	
Wald χ ² (32)	155.62**		129.72**		141.61**	
No of observations	1917		1917		1917	

Notes: Asymptotic standard errors are reported in parentheses (they were received by inverting the Fisher information matrix estimated by the outer product of scores); standard errors for frailty (σ²) and the proportion of potential movers to employment and inactivity were derived by using the delta method. ** and * denote significance at 5% and 10% levels, respectively. Excluded categories: single male, DKILL2 – skilled individuals.

Table 2.13: Two-destination hazard regression models of transitions from

Independent variable	Model ID1 Transition to		Model ID2 Transition to	
	Employment	Unemployment	Employment	Unemployment
AGE	-0.0267** (0.004)	-0.0242** (0.007)	-0.0266** (0.004)	-0.0243** (0.010)
FEMALE	0.0680 (0.110)	0.1544 (0.202)	0.0679 (0.110)	0.1541 (0.207)
MARIDMAL	0.3682** (0.132)	0.2706 (0.244)	0.3674** (0.132)	0.2703 (0.259)
HHINCOME / 10 ³	0.0148 (0.020)	-0.0135 (0.043)	0.0149 (0.020)	-0.0136 (0.045)
DSKILL1	-0.4062** (0.174)	-0.4322 (0.350)	-0.4059** (0.174)	-0.4329 (0.389)
DSKILL3	-0.0432 (0.109)	-0.0157 (0.203)	-0.0424 (0.109)	-0.0156 (0.203)
LANDUSE	-0.0947 (0.096)	0.1163 (0.182)	-0.0942 (0.096)	0.1166 (0.185)
OWNDACHA	0.1839* (0.099)	-0.0593 (0.197)	0.1826* (0.099)	-0.0595 (0.197)
SECJBIEA	0.3233** (0.113)	0.3394* (0.197)	0.3235** (0.113)	0.3395 (0.219)
METRO	0.1583 (0.141)	0.2135 (0.271)	0.1554 (0.141)	0.2135 (0.273)
URATE	-3.8917 (3.079)	-9.4933* (5.157)	-3.6123 (3.077)	-9.4816 (6.153)
NONRATE	2.2231 (1.383)	3.5024 (2.326)	2.1095 (1.382)	3.4967 (2.681)
Constant	-2.6939** (0.315)	-6.5801** (0.912)	-2.6977** (0.315)	-6.6045** (0.916)
t / 10	0.0984 (0.308)	4.4728** (1.170)	0.1083 (0.308)	4.5157** (1.229)
t ² / 10 ²	-0.1304 (0.141)	-2.2040** (0.531)	-0.1359 (0.141)	-2.2253** (0.536)
t ³ / 10 ⁵	2.0870 (2.424)	38.6580** (9.379)	2.1878 (2.428)	39.0530** (9.401)
t ⁴ / 10 ⁷	-0.9166 (1.372)	-22.2280** (5.592)	-0.9749 (1.374)	-22.4690** (5.593)
Frailty (log σ ²)	--	--	-12.8780 (186.450)	-5.6716 (395.030)
μ	--	--	--	--
Frailty (σ ²)	--	--	0.0000 (0.000)	0.0034 (1.360)
Proportion of movers	--	--	--	--
Log-likelihood	-4216.16		-4216.16	
Wald χ ² (32)	196.27**		195.88**	
No of observations	2147		2147	

Notes: Asymptotic standard errors are reported in parentheses (they were received by inverse estimated by the outer product of scores); standard errors for frailty (σ²) and the proportion of movers and inactivity were derived by using the delta method. ** and * denote significance at 5% level. Excluded categories: single male, DSKILL2 – skilled individuals.

Table 2.14: Two-destination hazard regression models of transitions from unemployment (piecewise constant hazard)

Independent variable	Model UD1 Transition to		Model UD2 Transition to		Model UD3 Transition to	
	Employment	Inactivity	Employment	Inactivity	Employment	Inactivity
AGE	-0.0056 (0.004)	0.0290** (0.005)	-0.0056 (0.004)	0.0390** (0.008)	-0.0056* (0.003)	0.0390** (0.008)
FEMALE	0.0325 (0.100)	0.0206 (0.133)	0.0326 (0.101)	0.0664 (0.217)	0.0321 (0.089)	0.0664 (0.219)
MARIDMAL	0.3156** (0.119)	-0.2986* (0.169)	0.3157** (0.120)	-0.5777** (0.276)	0.3154** (0.105)	-0.5777** (0.269)
HHINCOME / 10 ⁵	-0.0162 (0.029)	0.0409 (0.029)	-0.0163 (0.029)	0.3247** (0.158)	-0.0162 (0.030)	0.3246** (0.125)
DSKILL1	-0.2333 (0.206)	0.3114 (0.221)	-0.2333 (0.204)	0.3711 (0.430)	-0.2344 (0.193)	0.3709 (0.383)
DSKILL3	0.1171 (0.097)	0.0625 (0.148)	0.1171 (0.097)	0.1422 (0.227)	0.1172 (0.084)	0.1423 (0.224)
LANDUSE	-0.1584* (0.089)	0.2013 (0.139)	-0.1584* (0.089)	0.4506** (0.211)	-0.1585* (0.082)	0.4507** (0.219)
OWNDACHA	0.2194** (0.100)	0.0110 (0.145)	0.2195** (0.100)	-0.1886 (0.217)	0.2192** (0.085)	-0.1888 (0.221)
SECJBIEA	0.1577 (0.106)	-0.2543 (0.165)	0.1577 (0.106)	-0.4083 (0.253)	0.1577* (0.090)	-0.4083* (0.243)
METRO	-0.0067 (0.160)	-0.1657 (0.258)	-0.0065 (0.161)	-0.4740 (0.397)	-0.0073 (0.119)	-0.4738 (0.376)
URATE	-3.6810 (2.923)	-10.8400** (4.332)	-3.6799 (2.924)	-14.2820** (6.451)	-3.6853 (2.272)	-14.2830** (6.297)
NONRATE	-0.5819 (1.256)	5.1572** (1.848)	-0.5759 (1.251)	6.7440** (2.886)	-0.5935 (0.882)	6.7439** (2.781)
0 < t ≤ 12	-2.2535** (0.260)	-5.9490** (0.372)	-2.2555** (0.260)	-6.4001** (0.605)	-2.2492** (0.266)	-6.3998** (0.610)
12 < t ≤ 36	-2.0684** (0.262)	-4.7061** (0.363)	-2.0704** (0.262)	-4.6855** (0.608)	-2.0639** (0.270)	-4.6852** (0.612)
36 < t ≤ 60	-2.3776** (0.287)	-4.9529** (0.398)	-2.3798** (0.292)	-4.1531** (0.657)	-2.3727** (0.296)	-4.1527** (0.685)
60 < t ≤ 84	-2.1505** (0.358)	-4.7226** (0.493)	-2.1527** (0.359)	-3.2081** (0.816)	-2.1447** (0.362)	-3.2076** (0.819)
84 ≤ t	-6.8239 (1473.900)	-7.5646 (604.300)	-6.9723 (4063.100)	-7.6287 (9978.300)	-6.9843 (16.876)	-7.6285 (29.859)
Frailty (log σ ²)	--	--	-11.5700 (1435.100)	0.2045 (0.236)	-11.5580** (2.203)	0.2047 (0.267)
μ	--	--	--	--	11.5060 (31.326)	19.0320 (130.790)
Frailty (σ ²)	--	--	0.0000 (0.014)	1.2269** (0.290)	0.0000 (0.000)	1.2271** (0.328)
Proportion of movers	--	--	--	--	1.0000** (0.000)	1.0000** (0.000)
Log-likelihood	-4629.75		-4620.57		-4620.57	
Wald χ ² (32)	10553.20**		5754.35**		7851.60**	
No of observations	1917		1917		1917	

Notes: Asymptotic standard errors are reported in parentheses (they were received by inverting the Fisher information matrix estimated by the outer product of scores for models UD1-2, and by the expectation of the Hessian for model UD3; in the latter case our numerical routine failed to produce the outer product estimator); standard errors for frailty (σ²) and the proportion of potential movers to employment and inactivity were derived by using the delta method. ** and * denote significance at 5% and 10% levels, respectively.
Excluded categories: single male, DKILL2 – skilled individuals.

Table 2.15: Two-destination hazard regression models of transitions from inactivity (piecewise constant hazard)

Independent variable	Model ID1 Transition to		Model ID2 Transition to		Model ID3 Transition to	
	Employment	Unemployment	Employment	Unemployment	Employment	Unemployment
AGE	-0.0265** (0.004)	-0.0258** (0.007)	-0.0265** (0.004)	-0.0259** (0.009)	-0.0266** (0.004)	-0.0355** (0.011)
FEMALE	0.0653 (0.110)	0.1536 (0.201)	0.0659 (0.110)	0.1534 (0.204)	0.0651 (0.110)	-0.0340 (0.259)
MARIDMAL	0.3724** (0.132)	0.3099 (0.241)	0.3732** (0.132)	0.3093 (0.246)	0.3727** (0.133)	0.2230 (0.297)
HHINCOME / 10 ⁷	0.0159 (0.019)	-0.0114 (0.042)	0.0159 (0.019)	-0.0114 (0.043)	0.0158 (0.019)	-0.0184 (0.057)
DSKILL1	-0.4146** (0.173)	-0.4411 (0.349)	-0.4150** (0.174)	-0.4412 (0.361)	-0.4134** (0.174)	-0.6051 (0.445)
DSKILL3	-0.0455 (0.109)	-0.0200 (0.201)	-0.0458 (0.109)	-0.0197 (0.201)	-0.0461 (0.110)	0.0365 (0.257)
LANDUSE	-0.0918 (0.096)	0.1132 (0.180)	-0.0916 (0.096)	0.1140 (0.180)	-0.0925 (0.097)	0.2014 (0.226)
OWNDACHA	0.1845* (0.099)	-0.0389 (0.195)	0.1846* (0.099)	-0.0406 (0.197)	0.1855* (0.100)	-0.0499 (0.246)
SECJBIEA	0.3318** (0.112)	0.3649* (0.195)	0.3317** (0.113)	0.3658* (0.204)	0.3316** (0.118)	0.3620 (0.249)
METRO	0.1477 (0.141)	0.2413 (0.266)	0.1493 (0.142)	0.2383 (0.269)	0.1495 (0.144)	0.4579 (0.337)
URATE	-3.8697 (3.069)	-11.5050** (5.144)	-4.0436 (3.072)	-11.1400* (5.708)	-4.0451 (3.061)	-11.7440* (6.642)
NONRATE	2.1515 (1.376)	3.8878* (2.314)	2.2052 (1.378)	3.7412 (2.543)	2.2282* (1.381)	2.8255 (2.995)
0 < t ≤ 12	-2.6438** (0.262)	-4.3564** (0.496)	-2.6408** (0.262)	-4.3530** (0.517)	-2.6454** (0.265)	-3.2095** (0.649)
12 < t ≤ 36	-2.9389** (0.269)	-3.7210** (0.492)	-2.9352** (0.269)	-3.7169** (0.550)	-2.9401** (0.274)	-2.3885** (0.683)
36 < t ≤ 60	-3.3865** (0.274)	-4.3964** (0.519)	-3.3827** (0.274)	-4.3909** (0.703)	-3.3879** (0.287)	-2.8307** (0.844)
60 < t ≤ 84	-3.3918** (0.296)	-4.3493** (0.539)	-3.3882** (0.297)	-4.3422** (0.811)	-3.3936** (0.325)	-2.5720** (0.971)
84 ≤ t	-3.1220** (0.472)	-10.4990 (1589.400)	-3.1196** (0.473)	-9.8641 (851.110)	-3.1232** (0.519)	-7.4048 (503.090)
Frailty (log σ ²)	--	--	-12.8830 (167.310)	-5.0173 (123.900)	-12.8440 (178.690)	-5.7350 (233.800)
μ	--	--	--	--	11.4310 (1049.000)	0.5469 (0.555)
Frailty (σ ²)	--	--	0.0000 (0.000)	0.0066 (0.820)	0.0000 (0.000)	0.0032 (0.755)
Proportion of movers	--	--	--	--	1.0000** (0.011)	0.6334** (0.129)
Log-likelihood	-4228.48		-4228.46		-4226.35	
Wald χ ² (32)	12647.00**		10088.00**		6208.64**	
No of observations	2147		2147		2147	

Notes: Asymptotic standard errors are reported in parentheses (they were received by inverting the Fisher information matrix estimated by the outer product of scores); standard errors for frailty (σ²) and the proportion of potential movers to employment and inactivity were derived by using the delta method. ** and * denote significance at 5% and 10% levels, respectively. Excluded categories: single male, DKILL2 – skilled individuals.

Findings

Results of estimation are presented in tables 2.12-2.15. Tables 2.12 (exits from unemployment) and 2.13 (exits from inactivity) show the results for specifications with the polynomial baseline hazard, while tables 2.14 and 2.15 display estimated coefficients in corresponding models with the piecewise constant baseline hazard.

All the models are highly significant (model χ^2 -statistics are given in the boxes at the bottom of the tables). For each type of transition both estimates and their asymptotic standard errors (in parentheses) are reported. Not reported are the results of the tests for artificial distinctions in equations describing transitions to different destinations. In particular, in all the models exits to specific destinations are characterised by significantly different equations: Wald tests for equality of coefficients across destinations yield values ranging between 170.00 and 267.74 for models with exits from unemployment, and between 32.58 and 233.39 for models with exits from inactivity (χ^2 -statistics in both cases have 17-19 degrees of freedom). Similarly, Wald tests for equality of coefficients in corresponding UD and ID models generate highly significant values to comfortably reject the null hypothesis of the identity of the origin states.

The following observations as regards estimated coefficients are interesting.

First and foremost, in the models with polynomial specification of the interval baseline hazard for all types of transition we find a negative coefficient on the highest order term (t^4) in the polynomial shaping the duration dependence. It is always significant apart from the models with exits from inactivity to employment (table 2.13). This indicates that the graph of the hazards will converge to minus infinity as time increases (Fig.2.6

and 2.7³⁴), which may be a sign in support of the random draws model (see Section 2.B.9 in the appendix for the logic that we use here). However, from Fig.2.6 and 2.7 it follows that for the values present *in* our samples³⁵ the baseline hazards never tend to minus infinity. This can also be inferred more convincingly from the graphs of the corresponding piecewise constant hazards (Fig.2.8 and 2.9). Thus, it allows one to argue that either the hypothesis about the presence of specific distributions is not correct and can be rejected in the case of the RLMS sample, or that the follow-up period is not long enough, i.e. the kink in the PDF (Fig.2.4) takes place at a duration longer than 90 months. In any case, the findings indicate that for reasonably long durations there is no evidence of defective risks generated by the random draws model.³⁶

Secondly, while it makes little sense to comment on the similarities in hazard's shapes in Fig.2.6-2.9, some insights can be drawn as regards differences in the position of the graphs on the relative scale.

Consider, first, the right panel of Fig.2.6. It is evident that for the unemployed the hazard into employment (top-right panel) lies on average above the hazard to inactivity (bottom-right panel). In other words, for any given duration the unemployed are more likely to move into employment than become inactive. How can one interpret this finding? Van den Berg (2001) shows that in the standard search model the exit rate out of unemployment (or hazard) into employment is equal to the product of the job arrival rate

³⁴Both Fig.2.6 and 2.7 depict interval hazards drawn using the coefficients on time variables from tables 2.12 and 2.13. Left panels show the duration dependence in the hazard, discarding the effect of the constant, whereas right panels display the same graphs with the value of the constant taken as in tables 2.12 and 2.13. The hazard implied by the basic model is always drawn by a solid line. The hazards in models with unobserved heterogeneity are shown by dashed lines.

³⁵The maximum duration of joblessness registered in our samples is 90 months for unemployed workers and 100 months for the inactive.

³⁶A more detailed examination of Fig.2.6-2.9 may actually suggest that the graphs often imply hazards increasing in low values of duration and staying constant after about 12 months or so. The drop observed in the graphs for durations longer than approximately 80 months may just be explained by the low proportion of individuals with such durations of non-employment in the sample.

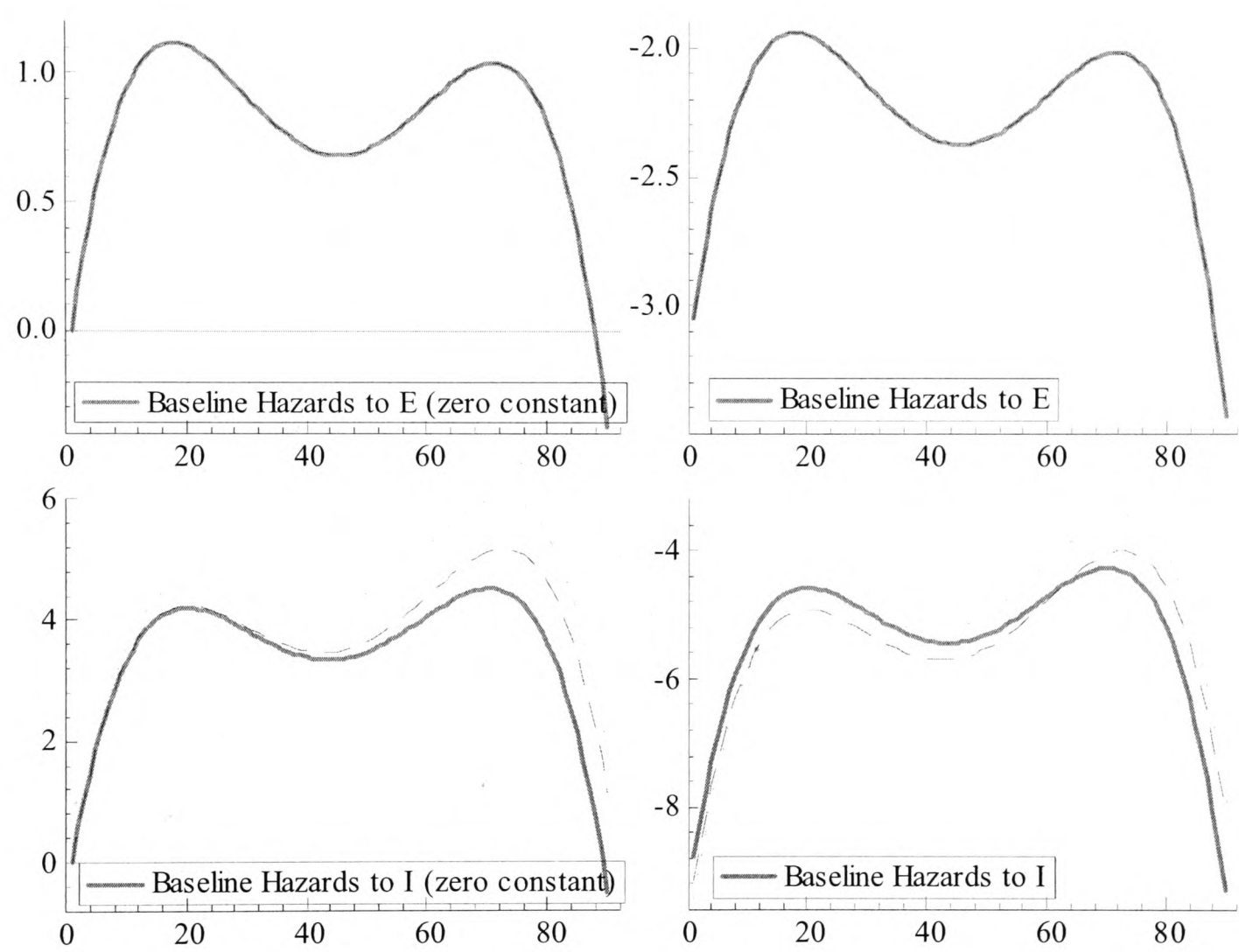


Figure 2.6: Polynomial baseline interval hazards for the unemployed transiting to employment and inactivity

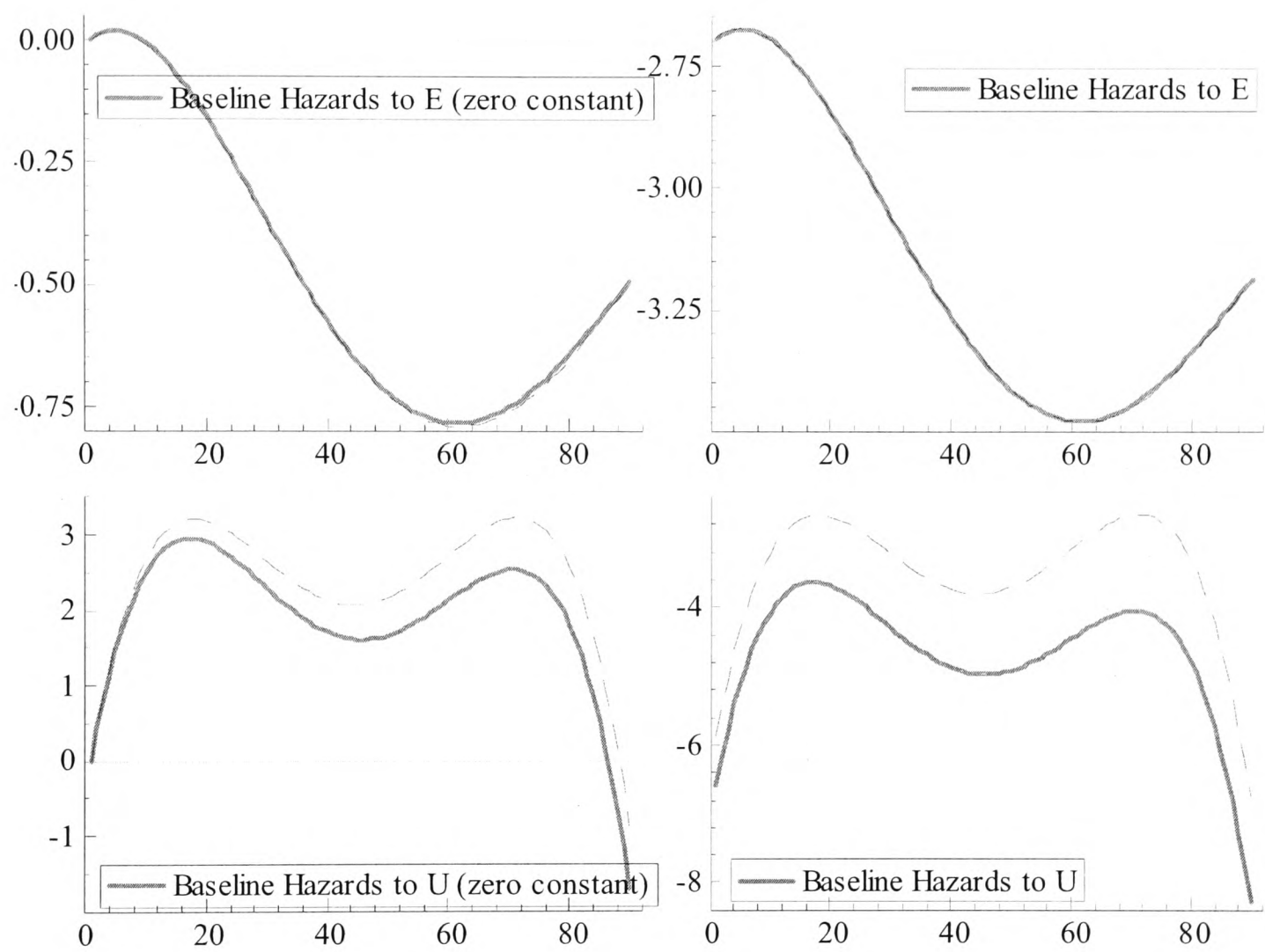


Figure 2.7: Polynomial baseline interval hazards for the inactive transiting to employment and unemployment

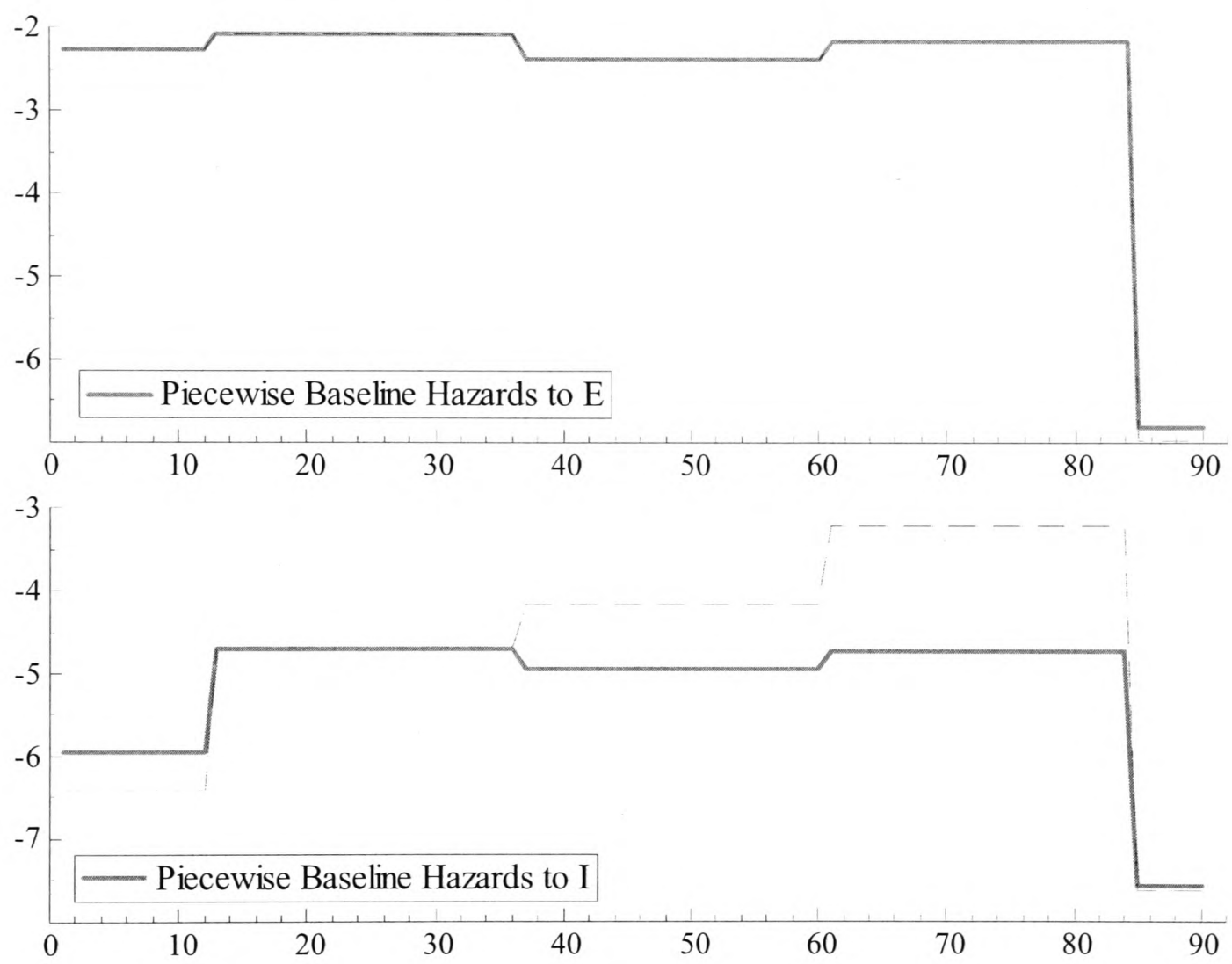


Figure 2.8: Piecewise baseline interval hazards for the unemployed transiting to employment and inactivity

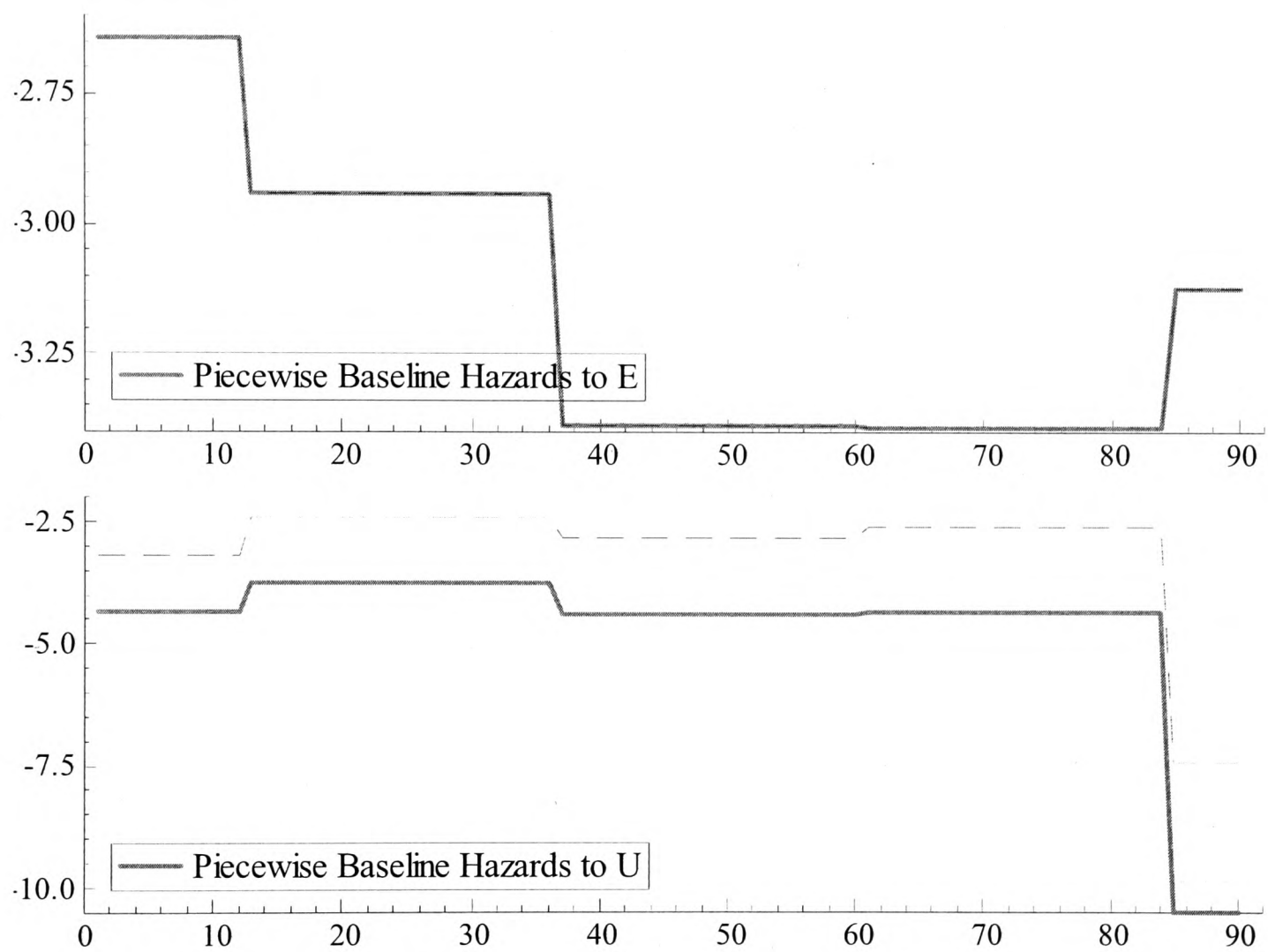


Figure 2.9: Piecewise baseline interval hazards for the inactive transiting to employment and unemployment

and the probability of accepting a job offer conditional on worker's reservation wage (see also formula (2.1) above). In the light of this, Fig.2.6 (right panel) may be a sign that inactivity is less desirable or less acceptable than employment. At the same time, if, for example, inactivity represents employment in the irregular sector, or subsistence activities, then the lower hazard into inactivity may imply that the size of the informal sector on the whole is smaller as compared to the size of the regular sector, total job creation in the former is less significant than in the latter, or worker's minimally acceptable wage is higher when she considers an irregular offer.

A different picture can be seen, however, if one draws the graphs of hazards from unemployment disregarding the effect of the constant (top-left and bottom-left panels of Fig.2.6). In that case, duration dependence of the exit rates is such that the hazard into employment lies on average below the hazard to inactivity. This may well imply that given the same average effects (i.e. disregarding the level effect of the constant) opportunities in inactivity, whatever they are, arise faster than unemployed workers come across an offer of employment. To put things differently, if inactivity is linked to the informal economy, then Fig.2.6 (left panel) clearly indicates that job opportunities arrive at a faster rate from the informal sector rather than from regular employment when the size of the sector is taken into account (i.e. assuming the same constant term in duration dependence). Similarly, by comparing the top-left panel of Fig.2.6 to the bottom-left panel of Fig.2.7, it is evident that *ceteris paribus* the inactive are faster to move into unemployment than the unemployed are to find a regular job. All this would in general be consistent with a stylised fact of a higher turnover in the informal economy (see, for example, Hoek, 2002, for evidence from a development country; Chetvernina and Lomonosova, 2001, and

Timofeev, 2001, can be referred to for some insights from the Russian labour market).

Further examining Fig.2.6 and 2.7, we can see that for the unemployed the chances of leaving their state for employment are *ceteris paribus* much greater than the chances of finding employment for the inactive. Again, by invoking theoretical implications from Van den Berg (2001), this may imply either that the job offer arrival rate is greater for the unemployed than for the inactive or that the inactive have a higher reservation wage, i.e. their job acceptance rate is lower, all other things being equal. The former guess is well in line with the idea *per se* of splitting the non-employed into the unemployed and the inactive: unemployed workers put an effort into job search so their job search effectiveness should be greater than that of the inactive. The latter guess may also be true if the inactive draw additional income from involvement in the subsistence economy. That income may be foregone by the unemployed while they are searching for regular jobs, thus leading to a lower reservation wage amongst them. Fig.2.8 and 2.9 depicting the piecewise constant hazards are also consistent with the two surmises.

The third finding concerns the effects of personal and other characteristics on the destination-specific hazards - a natural point to discuss after having looked into time dependence of the hazards implied by the coefficients on duration variables. The results of estimation are very similar for the two specifications of duration dependence in the baseline hazard. We should note, however, that one cannot apply likelihood ratio tests to differentiate between piecewise constant and polynomial specifications as the models are not nested. In general, the former appears to be more flexible, and as regards the non-duration characteristics it yields more significant estimates of the coefficients than the latter (compare tables 2.12 and 2.14, and 2.13 and 2.15). All in all, the results are

consistent with the conclusions following from the multinomial logit analysis in Section 2.3.3. In particular, it is found that being older and low-skilled reduces chances of leaving inactivity, while involvement in the irregular individual economic activity speeds up transition to regular employment (tables 2.13 and 2.15). For the unemployed land use is a factor that scales down the hazard to regular employment while increasing their chances of moving out of the labour force (tables 2.12 and 2.14). Finally, dacha ownership is a factor significantly reducing duration of joblessness both for the unemployed and the inactive (tables 2.12-2.15).

The fourth observation deserving attention is the effect of the introduction of unobserved individual heterogeneity ("frailty") in the models UD2-UD3 and ID2-ID3. The control for frailty does not worsen, as a rule, significance of other coefficients, and sometimes improves it substantially. At the same time, as indicated by the size of the variance³⁷ of the Gamma mixture distribution relative to its standard error in all four models with polynomial specification for the duration dependence in the baseline hazard, the unobserved heterogeneity is not significant (see the estimates of both $\log \sigma^2$ and σ^2 in tables 2.12 and 2.13). In the four models with piecewise baseline hazards, however, the unobserved effects are significant for the unemployed moving to inactivity (tables 2.14 and 2.15). In general, it appears that no omitted variables are likely to exert a significant impact on the results of the estimation, although the Gamma variance estimate is sensitive to the manner in which duration dependence in the baseline hazard is modelled (see also Addison and Portugal, 2003, for a similar conclusion).

Finally, the last but not the least observation relates to the second type of unobserved

³⁷Again, one should be aware that standard likelihood tests cannot be applied in order to choose, for instance, between models UD1 and UD2 and make judgements on the importance of frailty, because the former model is not a nested version of the latter model (see, e.g., Jenkins, 1998).

heterogeneity, or the proportions of movers to specific destinations estimated in models UD3 and ID3. Tables 2.12 and 2.14 indicate that on average the proportion of the unemployed not excluding inactivity as an option is close to one, and is likely to be equal to the proportion of the unemployed not excluding employment (the omitted results of the Wald tests suggest that even at 10%-significance level one cannot reject the hypothesis of equality of μ 's in the employment and inactivity equations). This effectively implies that in terms of the split population model all the unemployed are *movers*: even if they are unemployable they will eventually cease to be unemployed by exiting to inactivity. By contrast, the results given in tables 2.13 and 2.15 imply that the inactive, while do not rule out getting a regular job, are still likely to exclude unemployment at least with probability³⁸ 0.35. This indicates that there are *stayers* among the inactive: even if they are unemployable, they may never leave inactivity for unemployment because they just do not admit or do not have a possibility for such a move. All in all, the main implication of this analysis is that the unemployed workers are very unlikely to exclude inactivity as a choice. In other words, if the latter labour market state represents activities in the subsistence sector, unemployed workers in their hunt for jobs do not rule out a possibility of getting involved with an irregular opportunity.

Further implications

Thus, the main message of the findings above is that the increasing average duration of unemployment in Russia can be explained neither by the random draws model argument nor by the split-population premiss. We do not find evidence that employers may

³⁸Note, however, that once again the Wald test for equality of μ 's across equations does not allow to reject the null hypothesis.

use against the unemployed some ranking or other hiring rule based on the time spent without work or, alternatively, that there is a stigma of being the long-term unemployed. Neither do we find any indication that unemployed workers exclude inactivity as a choice, so that long-term unemployment emerges due to a significant proportion of potential stayers among job-seekers. Expanding on the surmise about involvement in the subsistence economy, this finding is likely to imply that the Russian unemployed are generally not averse to informal or irregular work. In other words, this may well be an indication that they do not differentiate job opportunities or "offers" on grounds of regularity: the workers may just use the first acceptable opportunity for employment whether it comes from the regular or the irregular sector. This way of reasoning would be consistent with the suggestion by Kolev (1998) that the involvement in the informal economy is a safety valve for the unemployed, and with the opinion recently expressed by Gërxhani (2004) that survival is the main characteristic of the informal activities in less developed countries. Perhaps, elsewhere the unemployed workers could have turned down the choice of subsistence work but not in the circumstances of the Russian transition.

By contrast, the findings as regards inactive workers may well suggest that stagnancy of the inactivity state is a product of worker's behaviour. In particular, the results in tables 2.13 and 2.15 imply that on average there may be a significant proportion of the inactive that rule out unemployment as a destination. This may be caused either by discouragement or by full-time involvement in subsistence production that does not leave time for active job-hunting in the regular sector. Notice that together with a high proportion of low-skilled workers among the inactive (see table 2.2), the application of the split-population argument here implies that discouraged workers can be helped only if they

are properly retrained and acquire skills valuable in the regular economy. Indeed, from the model it follows that the long-term jobless (the inactive in this case) are stuck in that state of the labour market both because they exclude transition to the other non-employment state, and because they do not possess characteristics wanted by employers (i.e. those people are *unemployable*). Thus, if the inactive, who are unemployable, exclude transition to unemployment and becoming participants of the labour market actively searching for jobs, it just means that they accept their estrangement from regular employment and behave rationally. They can be brought back to participation only through policies targeted at endowing them with skills demanded by the regular market.

The findings in the previous section also suggest some implications for reservation wages of unemployed and inactive workers. The shapes of the appropriate baseline hazards discussed therein may point to the fact that the reservation wage of the unemployed does not increase with the time spent without regular work (i.e. the job acceptance rate does not decrease with the time spent jobless). It would hardly be a surprising conclusion, which, however, deserves to be verified empirically. To our knowledge, there is only one study attempting to estimate reservation wages of Russian job-seekers (Smirnova, 2003a). However, it does not include a unemployment duration variable in equation for the wage offer function.

As regards the inactive, the shape of the baseline hazards to employment (Fig.2.7 and 2.9) suggests two possibilities. First, it is very likely that the regular job arrival rate is lower for the inactive than for the unemployed, and that it decreases with the time spent in inactivity. In other words, the inactive are less efficient in landing regular jobs, and are less likely to do so as their time spent in inactivity passes. Again, the discouraged

worker effect may well be responsible. The second possibility, however, is that the job acceptance rate for the inactive is lower, because they have higher reservation wages than the unemployed. Moreover, the shape of the baseline hazards in models ID1-ID3 may imply that the reservation wage of the inactive does increase with the time spent jobless. Such a situation may be realistic if, as has been mentioned above, the inactive draw substantial shadow income. In this regard it seems very interesting to estimate and compare wage offers and reservation wages both for the unemployed and the inactive. Kiefer and Neumann (1979) provide useful insights into the possible approach, but we leave this topic for future research.

All in all, the main lessons one can learn from the survival analysis in this section are as follows. First, the stagnancy of inactivity in Russia may well be explained by behaviour of workers, who realise that their skills are not demanded by the regular sector. In order to bring those people into participation a set of active labour market policies aimed at retraining of the low-skilled is needed. The second lesson is that causes of stagnancy of the unemployment pool should not be sought in peculiarities of search behaviour of workers or hiring behaviour of employers. The findings just confirm that both sides probably behave flexibly in their interaction in the labour market. Instead, an analysis of structural changes in the economy may provide us with a better understanding of causes of long-term unemployment. Also, more insights can be drawn from clarifying the factors affecting reservation wages of both the unemployed and the inactive. This points to the areas of future empirical and theoretical research that would be particularly interesting.

2.4 Concluding Remarks

In this essay we presented an analysis of worker flows between major labour market states in the Russian Federation over the 10 year period from 1992 to 2002 using the Russian Longitudinal Monitoring Survey data - a representative and ongoing survey of households. We were interested in determining the most significant trends in labour reallocation in Russia, assessing their uniformity and getting possible insights into determinants of the flows. Also, we aimed to draw some lessons for theoretical research into the labour force movements in transition that has thus far remained basic and has not been able to account for a number of phenomena both in Russia and CEE countries.

On the basis of yearly transition probabilities between three major labour market states, namely employment, unemployment and inactivity, we have confirmed previous findings that Russia is indeed notable for a highly mobile labour market, with a large degree of job-to-job transitions, which appear to be increasing as transition progresses. The extent of labour reallocation in Russia is found to be greater than that in other CEE transitional countries. Unemployment and inactivity appear to have become more important in reallocation of labour than they were at the beginning of transition, while both non-employment states are less stagnant than their counterparts in Eastern Europe.

The striking and salient observation made in this work is a substantial degree of mobility between ILO unemployment and inactivity in both directions among the Russian working age population. Previous studies have generally omitted the analysis of this aspect of labour movements. The flows between the two non-employment states in Russia comfortably outperform their CEE counterparts, while the general fall in the participation rate was smaller.

These features of labour reallocation were not predicted by the basic theory of transition, such as the seminal work by Aghion and Blanchard (1994). The theory does not envisage a crucial role for job-to-job movements and inactivity in labour force reallocation. In order to get further insights into whether these trends are uniform across various cohorts of individuals and into possible determinants of such flows we carried out a multinomial logit analysis of yearly transition probabilities.

The findings suggest that labour reallocation in Russia has not been uniform, and particular contingents of individuals have fared worse than others. As in CEE countries of economic transition, the low-skilled are the most distinct group whose labour market prospects are especially poor. We found that they are more likely to be among the inactive and face lower probabilities of finding a regular job. Women are the other group that seems to be at a disadvantage as far as successfulness of job search is concerned.

As regards determinants of various flows, several facts are worth mentioning. First, movement out of the labour force is likely to be connected with the incidence of wage arrears, forced unpaid leaves - a widespread tool of employment regulation in Russia - as well as access to land by a respondent's family. It appears that in general rationing in the regular labour market is closely interlinked with labour supply in the irregular markets and shifts into inactivity. Thus, it points to importance of the irregular economy in workforce reallocation, which presence may serve as an explanation for the marked, as noted above, circulation of labour between the two non-employment states.

If wage arrears increase the likelihood of moving out of the labour force, provision of non-wage benefits at the respondent's primary place of work is a significant factor in keeping attachment to employment. All in all, this points to the significance of effects

generated by remuneration mechanisms adopted by Russian firms. Evidence from various sources indicates that the practice of both arrears and non-monetary compensation may not be a feature of declining sectors alone. Thus, it calls for more attention to be paid to these causes in building a coherent theory of transitional labour mobility.

While the Russian labour market is indeed fluid and non-employment is less persistent than in Eastern Europe, both unemployment and inactivity have become more stagnant in Russia over the course of transition. Why did this happen? Again, theory provides no possible insights, and we attempt to test applicability of two general ideas that lend themselves to the potential for infinite duration of joblessness, and hence that may serve as an explanation for increasing stagnancy of non-employment. Both ideas, represented here by the random draws model and the split of population, suggest that particular behaviour of employers and workers in their interaction in the labour market may lead some workers to get permanently trapped in either unemployment or inactivity. However, the duration analysis conducted in this work suggested that neither model can explain the stagnancy of the unemployment pool. It is likely that both unemployed workers and employers behave flexibly in their search for appropriate matches in the labour market. As regards workers it implies that they are likely not to exclude possibilities of irregular employment and shift into inactivity if such opportunity arises. At the same time, there is no sign of employers using any screening mechanism that would make the chances for employment of the long-term unemployed be negligible. By contrast, some indication was found that long-term inactivity may indeed be a product of choice of workers, who either realise their inability to get employed in the regular sector or are involved in the irregular production and do not search for a regular placement. Still, however, other reasons for the growing stagnancy of

inactivity are possible. One such explanation may be in the dynamics of reservation wages of the inactive. We are aware of no theoretical or empirical work that investigates this issue in relation to the Russian transition. In general, the formation of reservation wages of jobless workers may be an interesting area for future research, taking into account the seeming importance of irregular employment, and hence incomes in Russia. That could provide more insights into the growing stagnancy of unemployment too.

While interesting in their own right, the findings of this essay identify challenges for theoretical modelling of labour reallocation in transition. Improvements in theory should be sought in a few directions as follows. First, the role of inactivity in redeployment of the labour force should receive a better treatment. In particular, attention should be paid to irregular employment opportunities, including home (subsistence) production. Second, and related to the previous point, labour supply of workers seems to be important and interlinked with mobility decisions. Third, job-to-job flows need to be taken into account. Fourth, wage-setting mechanisms, in general, and, in particular, wage arrears and non-wage benefits should be better attended to in a model of workforce reallocation. Fifth, formation of reservation wages of jobless workers should be studied - this may shed some light on peculiarities of reallocation through non-employment states. Sixth, heterogeneity of the labour force is an important factor that must be dealt with. The evidence clearly suggests that job matching probabilities are different across various groups of individuals. Finally, explanations for growing stagnancy of unemployment are particularly welcome. It is clear that it is not caused by specifics of search process, and reasons should probably be sought in structural adjustment of the economy.

It should be stressed in conclusion that the basic theory of labour reallocation in

transition (Aghion and Blanchard, 1994; Blanchard, 1997) hardly addresses any of the challenges above. In recent years, however, there have been a few attempts differing in the degree of success to extend the theory by incorporating job-to-job movements (Garibaldi and Brixiova, 1998; Boeri, 2000b), peculiarities of labour supply (Boeri, 2000b; Bouev, 2001), informal economy (Boeri, 2000b), the heterogeneous labour force (Boeri, 2000b; Commander *et al.*, 2004), and so forth. Still however, there is spacious room for further contributions as the topic of labour reallocation in transition remains especially hot.

2.A Description of Variables Used in the Analysis

The following variables were used either for the descriptive statistics, or for the multinomial logit and survival analyses:

AGE - age of the respondent;

AGE24 - a dummy (1 - if aged less than 25 y.o.; 0 - otherwise);

AGE51 - a dummy (1 - if aged more than 50 y.o.; 0 - otherwise);

AGESQRD - age squared divided by 100 (for scaling reasons);

ARREARS - a dummy (1 - if the respondent has experienced wage arrears (full or partial non-payment of wages) at their primary place of work; 0 - otherwise) - the necessary information is available in rounds V-XI;

CHLD_7 - a number of children 7 y.o. and younger in the respondent's household;

CHLD7_18 - a number of children from 7 to 18 y.o. in the respondent's household;

CHILDFEM - a dummy (1 - if the respondent is a female and there are children younger than 7 y.o. in their household; 0 - otherwise);

CHILDREN - a number of children 17 y.o. and younger in the respondent's household (a sum of variables CHLD_7 and CHLD7_18);

DSKILL1 - a dummy for low skilled respondents based on ISCED-97 levels (1 - for a person with lower education, defined as ISCED-1997 levels 0-2 (see e.g. UNESCO, 1998): up to lower secondary or the second stage of basic education (i.e. up to 8-9 grades of 10-11 grade school according to Russian standards³⁹); 0 - otherwise);

DSKILL2 - a dummy for skilled respondents based on ISCED-97 levels (1 - for a person with education at ISCED-1997 levels between 2 and 5; 0 - otherwise);

DSKILL3 - a dummy for high skilled respondents based on ISCED-97 levels (1 - for a person with higher education, defined as ISCED-1997 levels 5-6 (levels 5-7 by ISCED-1976): from the first stage of tertiary education to the second stage of tertiary education (i.e. incomplete higher education to post-graduate degrees in Russia); 0 - otherwise);

HHINCOME - total real income in respondent's household in 1992(I) prices (for its derivation we use data on the percentage change in CPI from the International Financial Statistics (IMF, various issues); the CPI is calculated in 1992(I) rubles; income for each household is adjusted for the inflation in the nearest quarter of a year; incomes after 1998

³⁹11-grade schooling in Russia was introduced in 1989.

are divided by 1000 to take account of the denomination of the Russian ruble on 1 January 1998 (1/1000)); missing values are substituted for the average household income in the respondent's region;

LANDUSE - a dummy for land access (1 - if the respondent's household has had access to land over the past year; 0 - otherwise);

MARIDFEM - a dummy (1 - if the respondent is a married female; 0 - otherwise);

MARIDMAL - a dummy (1 - if the respondent is a married male; 0 - otherwise);

METRO - a dummy for the metropolitan areas of Moscow and St.Petersburg (1 - if a respondent lives in this area; 0 - otherwise);

NONRATE - the non-employment rate (the working age unemployed plus the working age inactive divided by the population of working age) in the respondent's region;

OWNDACHA - a dummy for dacha ownership (1 - if the respondent's household has a dacha or a country cottage; 0 - otherwise);

POSTWRK - a number of post work age persons (55+ y.o. for women, 60+ y.o. for men) in the respondent's household;

SECJBIEA - a dummy (1 - if the respondent has a second job or is engaged in any Individual Economic Activity (either on a regular or occasional basis); 0 - otherwise);

SINGLFEM - a dummy (1 - if the respondent is a single female; 0 - otherwise);

TENURE - respondent's tenure in months at their primary place of work - the necessary information is available in rounds V-XI;

MSTATE - a labour market state/status of the respondent (1 - Employment; 2 - Unemployment; 3 - Inactivity (out of the labour force));

UDUR03 - a dummy for non-employment duration (1 - if from 0 to 3 months; 0 - otherwise);

UDUR46 - a dummy for non-employment duration (1 - if from 4 to 6 months; 0 - otherwise);

UDUR712 - a dummy for non-employment duration (1 - if from 7 to 12 months; 0 - otherwise);

UDUR12_ - a dummy for non-employment duration⁴⁰ (1 - if over 12 months; 0 - otherwise);

⁴⁰Missing duration is filled with average duration across workers from the same geographical region.

ULEAVE - a dummy (1 - if the respondent has experienced an unpaid (forced / compulsory) leave; 0 - otherwise) - the necessary information is available in rounds V-XI;

URATE - the unemployment rate (the number of the working age unemployed divided by the sum of the employed and the unemployed in working age) in the respondent's region;

BENEFIT - a number of non-wage social benefits/perquisites at primary place of employment (ranges from 0 to 10; examples of benefits include a regular paid vacation, a paid sick leave, full or partial payment for treatment in a medical institute, free childcare, transportation and educational grants, etc.) - the necessary information is available in rounds IX-XI.

2.B Methodology (Survival Analysis)

This appendix provides a detailed explanation of the methodology used to estimate reduced form duration models for unemployment and inactivity as specified in the main text. We derive log-likelihood functions for all types of model estimated in this essay.

In general, depending on the nature of an underlying phenomenon and/or data a duration model can be formulated in either continuous or discrete time. The type of timing in which a duration model is set bears on its implications and conclusions. Below we begin by explaining basic building blocks in continuous time (which is always a most straightforward way to present a model of survival) and within a single risk framework. Then we move on to considering competing risks analogues, and after making a simplifying assumption present the models in discrete time. The actual estimation is carried out for discrete time versions because of the way that the RLMS duration data are organised.

The last two sections of the appendix outline possible approaches to estimation of models whose aim is to recognise specific distributions (see the main text), discuss the issue of identifiability of defective risks, and, finally, provide a list of main assumptions that we make while modelling.

2.B.1 A Simple Hazard Model

A basic critical concept in statistical analysis of duration data is the hazard function. Applied to the investigation of non-employment duration the continuous time hazard function gives the instantaneous probability of leaving non-employment at some moment of time t given that the individual has been jobless till t :

$$\lambda(t) = \lim_{\Delta \rightarrow 0} \frac{\Pr(t \leq T \leq t + \Delta | T \geq t)}{\Delta} = \frac{f(t)}{1 - F(t)} = \frac{f(t)}{S(t)}, \quad (2.2)$$

where $f(t)$ is the probability density function of random variable T , which represents spell lengths; $F(t)$ is the cumulative distribution function, and $S(t)$ is the survivor function. The latter can be reexpressed with the help of integrated hazard function $\Lambda(t)$ as

$$S(t) = \exp(-\Lambda(t)), \text{ where } \Lambda(t) = \int_0^t \lambda(\tau) d\tau. \quad (2.3)$$

2.B.2 Proportional Hazard

We model the hazard rate by assuming that it has a proportional structure:

$$\lambda(t, \mathbf{X}) = \lambda_0(t) \exp(\mathbf{X}'\boldsymbol{\beta}), \quad (2.4)$$

where $\lambda_0(t)$ is the baseline hazard function, which depends on t , but not on vector of covariates $\mathbf{X}$. It summarises the pattern of duration dependence, assumed to be common to all individuals (Jenkins, 2004). Non-negative function $\exp(\mathbf{X}'\boldsymbol{\beta})$ is person specific and scales the common baseline hazard. The proportional hazard specification is a common way to accommodate the presence of observed individuals (see, among others, Cox, 1972; Meyer, 1990; Van den Berg, 2001). At the same time, it allows a convenient formalisation of the models in discrete time, as we shall see below.

The proportionality assumption cannot in general be justified by economic theory of search in the labour market. However, Van den Berg (2001) notes that if the optimal strategy of the individual is myopic then the proportionality can often be deduced from theory. Myopic strategies are optimal if, for example, search is repeated or if the discount rate is infinite. It is known that economic agents in East European countries of transi-

tion often resort to myopic strategies in a variety of activities⁴¹ (for some examples see Berkowitz and DeJong, 1997). Thus, it seems reasonable to believe that the proportional structure of the hazard is not too strong an assumption in our case.

2.B.3 Survival is Continuous but Spell Lengths are Interval Censored

The RLMS does not always contain information on the precise number of days spent jobless by sampled individuals. Thus, as has been noted in Section 2.2.3, we prefer to reorganise the available evidence on non-employment duration according to monthly intervals, so that for each individual i the spell is measured in the number of months, t_i , spent without regular work. Given that, the log-likelihood can be written as

$$\log L(\boldsymbol{\theta}|t, \mathbf{X}) = \sum_{i=1}^N \{ \delta_i \log (S(t_i - 1; \mathbf{X}_i) - S(t_i; \mathbf{X}_i)) + (1 - \delta_i) \log S(t_i; \mathbf{X}_i) \}, \quad (2.5)$$

where it is a function of the vector of parameters, $\boldsymbol{\theta}$, to be estimated (it may and will vary from a model to a model - we shall elaborate on its components in Section 2.B.8 of this appendix) conditional on observing covariates $\mathbf{X}$ and survival times t . Also, N is the number of observations in the sample; δ_i is the indicator function, which equals 1 if exit from non-employment is observed for an individual i (i.e. a completed spell is recorded), or 0 if the spell for that individual is censored. Notice that for uncensored spells the moment of exit is known to a particular month, $(t_i - 1; t_i]$, but not to a day.

2.B.4 Sampling Issues

The duration data drawn from the RLMS sample are left-censored, i.e. the starting date of any completed or incomplete spell is not observed. With such left-truncated (or otherwise called "delayed entry") spell data one has to take account of survival of each individual up to the time that she has been first sampled as jobless. In other words, to

⁴¹Transition brings about an increase in all kinds of uncertainty which should naturally lead to an increase in the subjective discount rates. See, for example, the seminal paper by Yaari (1965) who shows that it is optimal to discount future more heavily because of uncertain lifetime.

avoid the length bias sampling problems the expressions for completed and incomplete spells in the log-likelihood function (2.5) must be conditioned on $S(r_i; \mathbf{X}_i)$, where r_i is the *full* number of months of joblessness reported by individual i at the moment of her inclusion in the duration sample, i.e. the number of months prior to the "delayed entry". Thus, we have

$$\log L(\boldsymbol{\theta}|t, \mathbf{X}) = \sum_{i=1}^N \left\{ \delta_i \log \frac{S(t_i - 1; \mathbf{X}_i) - S(t_i; \mathbf{X}_i)}{S(r_i; \mathbf{X}_i)} + (1 - \delta_i) \log \frac{S(t_i; \mathbf{X}_i)}{S(r_i; \mathbf{X}_i)} \right\}. \quad (2.6)$$

2.B.5 Unobserved Individual Heterogeneity

We also generalise the models considered thus far to allow for unobserved individual effects. In duration literature these are often referred to as "frailty", by a tradition originating in biostatistics (see Vaupel *et al.*, 1979). The inclusion of a frailty factor may be helpful to correct for omission of relevant variables or measurement errors in observed survival times or covariates (see Lancaster, 1990, Ch.4). At the same time, if unobserved effects are important but ignored, the "no-frailty" model will over-estimate the degree of negative duration dependence in the hazard (see Jenkins, 2004, for a detailed and most recent discussion of this and other possible effects of the omission of frailty).

We consider a "mixture" hazard model, where

$$\lambda(t, \mathbf{X}|\nu) = \nu \lambda(t, \mathbf{X}) = \nu \lambda_0(t) \exp(\mathbf{X}'\boldsymbol{\beta}). \quad (2.7)$$

The random multiplicative error term, ν , is, as is conventional, assumed Gamma distributed independently of t and $\mathbf{X}$, with mean 1 and variance σ^2 . After introducing the frailty factor, the continuous time survivor function becomes (again, see Lancaster, 1990, Ch.4):

$$\bar{S}(t; \mathbf{X}) = (1 + \sigma^2 \Lambda(t; \mathbf{X}))^{-1/\sigma^2}. \quad (2.8)$$

Now the log-likelihood function is

$$\log L(\boldsymbol{\theta}|t, \mathbf{X}) = \sum_{i=1}^N \left\{ \delta_i \log \frac{(1 + \sigma^2 \Lambda(t_i - 1; \mathbf{X}_i))^{-1/\sigma^2} - (1 + \sigma^2 \Lambda(t_i; \mathbf{X}_i))^{-1/\sigma^2}}{(1 + \sigma^2 \Lambda(r_i; \mathbf{X}_i))^{-1/\sigma^2}} + \right. \\ \left. (1 - \delta_i) \log \frac{(1 + \sigma^2 \Lambda(t_i; \mathbf{X}_i))^{-1/\sigma^2}}{(1 + \sigma^2 \Lambda(r_i; \mathbf{X}_i))^{-1/\sigma^2}} \right\}. \quad (2.9)$$

2.B.6 More Heterogeneity: the Split-Population Model

Another way to account for heterogeneity among individuals is to use the mover-stayer approach by considering a split-population model (see, e.g. Schmidt and Witte, 1989; Pudney and Thomas, 1995). To illustrate the point, assume most simply, that there are two groups of individuals among the non-employed - those who are "motivated" to leave this state of the labour market, and those who are not. This motivation factor cannot apparently be captured by observable characteristics, and thus can be regarded as a factor of unmeasured heterogeneity. Assume that there is a proportion $P \in [0, 1]$ of people motivated to leave non-employment. Then the survivor function is redefined as

$$\hat{S}(t; \mathbf{X}) = 1 - P + PS(t; \mathbf{X}), \quad (2.10)$$

where $S(t; \mathbf{X})$ is the survivor function of the motivated individual in its general form given in (2.3). Thus, the survival probability for any particular individual is given by the probability of her being a unmotivated stayer, who does not exit non-employment with probability 1, plus the probability of her being a mover, who does not exit non-employment with probability $S(t; \mathbf{X})$.

The mover-stayer model set up as above is, in fact, a way of incorporating the Heckman-Singer (1984) strategy of non-parametric accounting for unobserved individual heterogeneity. This strategy is potentially vulnerable to the choice of the (*continuous*) hazard function as found by Trussell and Richards (1985). However, as shall become clear from the discussion of a discrete timing model below, the assumptions that we made about the *interval* (i.e. discrete) baseline hazard introduced later allow for flexible forms of the continuous baseline hazard, thus mitigating arguments by Trussell and Richards (1985). The important implications of introducing heterogeneous groups of individuals in our analysis is not an additional source of heterogeneity *per se*, but a possibility of detecting the

presence of defective risks inherent in individual behaviour that may lead to very long durations of non-employment. In particular, the proportions of people represented by P may give additional insights into the character of search in the labour market, especially when a model with competing risks is considered (see Section 2.3.4).

With the described split of population the log-likelihood function takes on the form

$$\log L(\boldsymbol{\theta}|t, \mathbf{X}) = \sum_{i=1}^N \left\{ \delta_i \log \frac{P(S(t_i - 1; \mathbf{X}_i) - S(t_i; \mathbf{X}_i))}{1 - P + PS(r_i; \mathbf{X}_i)} \right. \quad (2.11)$$

$$\left. + (1 - \delta_i) \log \frac{1 - P + PS(t_i; \mathbf{X}_i)}{1 - P + PS(r_i; \mathbf{X}_i)} \right\}. \quad (2.12)$$

2.B.7 The Split-Population Model with Frailty

Finally, in the most complicated version of the models that we estimate, both Gamma heterogeneity and defective risks (accounted for in the split-population model) are combined. Thus, two sources of potential unmeasured differences are incorporated to control for the effects of omitted factors. The survivor function assumes the form

$$\tilde{S}(t; \mathbf{X}) = 1 - P + P\bar{S}(t; \mathbf{X}) = 1 - P + P(1 + \sigma^2\Lambda(t; \mathbf{X}))^{-1/\sigma^2}. \quad (2.13)$$

The log-likelihood function becomes

$$\log L(\boldsymbol{\theta}|t, \mathbf{X}) = \sum_{i=1}^N \left\{ \delta_i \log \frac{P\left((1 + \sigma^2\Lambda(t_i - 1; \mathbf{X}_i))^{-1/\sigma^2} - (1 + \sigma^2\Lambda(t_i; \mathbf{X}_i))^{-1/\sigma^2}\right)}{1 - P + P(1 + \sigma^2\Lambda(r_i; \mathbf{X}_i))^{-1/\sigma^2}} + \right. \quad (2.14)$$

$$\left. (1 - \delta_i) \log \frac{1 - P + P(1 + \sigma^2\Lambda(t_i; \mathbf{X}_i))^{-1/\sigma^2}}{1 - P + P(1 + \sigma^2\Lambda(r_i; \mathbf{X}_i))^{-1/\sigma^2}} \right\}.$$

2.B.8 Dual Risks Models

Having presented the main ideas for single risk models the continuous time analogues for exits to two destinations are straightforward to derive given the assumption of independence of the competing risks. In that case the continuous hazard rate for exit to any destination can be represented as a sum of destination-specific hazard rates, while

the survivor function for exit to any destination can factor into a product of destination-specific survivor functions (see, e.g. Jenkins, 2004). Then the log-likelihood function of the model separates nicely into two parts corresponding to two single risk models of exit to the respective competing destinations. Complications, however, arise when one deals with discrete times. In particular, the contribution to the model's log-likelihood function associated with exit to one destination must be modelled as a *joint* probability that the exact non-employment spell length lay between the boundaries of the interval (representing a discrete time unit) *and* that the latent exit time to a competing destination was after the latent exit time to the first destination⁴² (*ibid*). In practice it can pose serious difficulties for analysis unless additional assumptions are made about the within-interval behaviour of destination-specific hazard rates. One of such assumptions, greatly simplifying estimation of discrete time competing risks models, is that transitions out of the primary state can occur only at the boundaries of the intervals. In general this is quite a strong assumption to make, although applications to the labour force dynamics are known (see, e.g. Narendranathan and Stewart, 1993, who study British unemployed men using duration data grouped into weekly intervals). As regards our own work, with non-employment duration data organised into monthly spells, the assumption can be justified by the fact that the reported lengths of spells without work have always been *rounded up* to the nearest month. Thus, even if the transition out of non-employment could have happen, say, in the middle of a monthly period, rounding up ensures that it can be treated as if it had happened at the end of the monthly interval.

With this assumption, the dual risks versions of log-likelihood functions considered above for a single risk case, are all represented by the expression:

$$\begin{aligned} \log L(\theta|t, \mathbf{X}) = & \sum_{i=1}^N \left\{ \delta_i^E \log \left(\frac{Z^E(t_i - 1; \mathbf{X}_i^E) - Z^E(t_i; \mathbf{X}_i^E)}{Z^E(r_i; \mathbf{X}_i^E) Z^N(r_i; \mathbf{X}_i^N)} Z^N(t_i; \mathbf{X}_i^N) \right) + \right. \\ & \delta_i^N \log \left(\frac{Z^N(t_i - 1; \mathbf{X}_i^N) - Z^N(t_i; \mathbf{X}_i^N)}{Z^E(r_i; \mathbf{X}_i^E) Z^N(r_i; \mathbf{X}_i^N)} Z^E(t_i; \mathbf{X}_i^E) \right) + \\ & \left. (1 - \delta_i^E - \delta_i^N) \log \frac{Z^E(t_i; \mathbf{X}_i^E) Z^N(t_i; \mathbf{X}_i^N)}{Z^E(r_i; \mathbf{X}_i^E) Z^N(r_i; \mathbf{X}_i^N)} \right\}, \end{aligned} \quad (2.15)$$

⁴²The problem with modelling of the joint probability does not arise if the time nature of the data is intrinsically discrete. Then an exit to one and only one destination can occur over the time interval (see, again, Jenkins, 2004).

where $Z^E(t; \mathbf{X}^E)$ and $Z^N(t; \mathbf{X}^N)$ are destination-specific survivor functions for exits to employment and another non-employment state (i.e. inactivity - for the model of exits from unemployment, and unemployment - for the model of exits from inactivity), respectively; δ_i^E and δ_i^N - are destination-specific indicator functions equal to 1 if exit to a particular destination is observed for individual i , and to 0 otherwise.

In this essay we consider three variants of the competing risk models, so that the destination-specific survivor functions $Z^E(t; \mathbf{X}^E)$ and $Z^N(t; \mathbf{X}^N)$ are defined as follows.

For the basic model they are

$$Z^j(t; \mathbf{X}^j) = \exp(-\Lambda^j(t; \mathbf{X}^j)), \quad (2.16)$$

where $\Lambda^j(t; \mathbf{X}^j) = \int_0^t \lambda^j(\tau; \mathbf{X}^j) d\tau$, and $\lambda^j(t; \mathbf{X}^j) = \lambda_0^j(t) \exp(\mathbf{X}^{j'} \boldsymbol{\beta}^j)$, $j = E, N$.

For the model with Gamma heterogeneity terms in the destination-specific hazards, we have:

$$Z^j(t; \mathbf{X}^j) = (1 + \sigma_j^2 \Lambda^j(t; \mathbf{X}^j))^{-1/\sigma_j^2}, \quad (2.17)$$

where, as before, $\Lambda^j(t; \mathbf{X}^j) = \int_0^t \lambda^j(\tau; \mathbf{X}^j) d\tau$, while $\lambda^j(t; \mathbf{X}^j) = \nu_j \lambda_0^j(t) \exp(\mathbf{X}^{j'} \boldsymbol{\beta}^j)$ and ν_j is a random state-specific multiplicative error term associated with a corresponding hazard function. It is assumed that ν_j is Gamma distributed independently of t and $\mathbf{X}^j$, with mean 1 and variance σ_j^2 ; $j = E, N$.

Finally, for the split-population model with Gamma heterogeneity, the survivor functions are given by

$$Z^j(t; \mathbf{X}^j) = 1 - P^j + P^j (1 + \sigma_j^2 \Lambda^j(t; \mathbf{X}^j))^{-1/\sigma_j^2}, \quad (2.18)$$

where all the terms are as above, while P^j represents the proportion of movers associated with destination $j = E, N$.

Discrete time models

The data on spell duration drawn from the RLMS have, as mentioned before, been grouped into the intervals of equal length (one month), so that one cannot directly apply the continuous time models outlined in the previous sections. Instead, one should set up their discrete time analogues. Under the assumptions of proportionality in hazard and of transitions occurring at the boundaries of the intervals, the reformulation becomes a straightforward matter.

In particular, the interval or the discrete hazard rate of a failure event (e.g. exit from unemployment) happening in a given month, $(t_i - 1, t_i]$, for individual i is

$$h^j(i; \mathbf{X}_i^j) = h^j((t_i - 1, t_i]; \mathbf{X}_i^j) = \frac{S^j(t_i - 1; \mathbf{X}_i^j) - S^j(t_i; \mathbf{X}_i^j)}{S^j(t_i - 1; \mathbf{X}_i^j)}, \quad j = E, N. \quad (2.19)$$

With proportionality in the continuous time hazard, formula (2.19) suggests a relationship between the continuous and discrete time hazards:

$$\begin{aligned} h^j(i; \mathbf{X}_i^j) &= 1 - \exp \left(- \exp \left(\mathbf{X}_i^{j'} \boldsymbol{\beta}^j + \log \left(\int_{t_i-1}^{t_i} \lambda_0^j(\tau) d\tau \right) \right) \right) \\ &= 1 - \exp \left(- \exp \left(\mathbf{X}_i^{j'} \boldsymbol{\beta}^j + \gamma^j(i) \right) \right), \end{aligned} \quad (2.20)$$

where

$$\begin{aligned} \gamma^j(i) &= \gamma^j((t_i - 1, t_i]) = \log \left(\int_{t_i-1}^{t_i} \lambda_0^j(\tau) d\tau \right) \\ &= \log \left(\Lambda^j(t_i) - \Lambda^j(t_i - 1) \right), \quad j = E, N, \end{aligned} \quad (2.21)$$

as

$$\begin{aligned} S^j(t; \mathbf{X}^j) &= \exp \left(- \Lambda^j(t; \mathbf{X}^j) \right) = \exp \left(- \int_0^t \lambda^j(\tau; \mathbf{X}^j) d\tau \right) \\ &= \exp \left(- \exp \left(\mathbf{X}^{j'} \boldsymbol{\beta}^j \right) \int_0^t \lambda_0^j(\tau) d\tau \right). \end{aligned}$$

Then the functions of destination-specific survival up till interval $(t_i - 1, t_i]$ are given by

$$S^j(i; \mathbf{X}_i^j) = S^j(t_i; \mathbf{X}_i^j) = \prod_{k=1}^i (1 - h^j(k; \mathbf{X}_i^j)), \quad j = E, N. \quad (2.22)$$

Therefore, we can update the expressions for the survivor functions that need to be substituted in general formula (2.15) for all the models that we estimate in this work.

Thus, for the basic model $Z^E(i; \mathbf{X}_i^E)$ and $Z^N(i; \mathbf{X}_i^N)$ are defined as

$$Z^j(i; \mathbf{X}_i^j) = \prod_{k=1}^i (1 - h^j(k; \mathbf{X}_i^j)), \quad j = E, N, \quad (2.23)$$

where $h^j(i; \mathbf{X}_i^j)$ are given by formula (2.20). Also note that by definition $Z^j(0; \mathbf{X}_i^j) = 1$, $j = E, N$.

For the model with $\Gamma(1, \sigma_j^2)$ -distributed errors to control for unobserved heterogeneity we have:⁴³

$$Z^j(i; \mathbf{X}_i^j) = \left(1 + \sigma_j^2 \sum_{k=1}^i \exp(\mathbf{X}_i^{j'} \boldsymbol{\beta}^j + \gamma^j(i)) \right)^{-1/\sigma_j^2}, \quad j = E, N, \quad (2.24)$$

and where $\gamma^j(i)$ are as in (2.21).

Finally, for the split-population model with unobserved Gamma heterogeneity the survivor functions are given by

$$Z^j(i; \mathbf{X}_i^j) = 1 - P^j + P^j \left(1 + \sigma_j^2 \sum_{k=1}^i \exp(\mathbf{X}_i^{j'} \boldsymbol{\beta}^j + \gamma^j(i)) \right)^{-1/\sigma_j^2}, \quad j = E, N, \quad (2.25)$$

where all the terms are as above, while P^j represents the proportion of movers associated with destination $j = E, N$.

The maximum likelihood function (2.15) is maximised with respect to vector of parameters $\boldsymbol{\theta}$. It includes functional form parameters of $\gamma^j(\cdot)$ and coefficients on destination-specific vectors of covariates $\mathbf{X}^j$ (both groups are present in all three types of the model),

⁴³See Meyer (1990) for the expression for the likelihood function with discrete hazards and Gamma-distributed frailty.

the variances of unobserved Gamma heterogeneity, σ_j^2 (in models with the frailty factors), and, finally, proportions of movers to the two destinations, P^j , $j = E, N$ (in the split-population model). In order to prevent negative values of σ_j^2 and ensure that $P^j \in [0, 1]$, in actual estimation we employ logarithmic and logit transformations for the two types of parameters, respectively. So, instead of σ_j^2 we estimate $\log \sigma_j^2$, and instead of P^j we estimate constants μ_j , such that $P^j = \frac{\exp \mu_j}{1 + \exp \mu_j}$, $j = E, N$.⁴⁴

2.B.9 Interval Baseline Hazard and Identification of Specific Distributions

There are two approaches to the estimation of discrete time duration models. One way is to explicitly relate a discrete time model to a underlying continuous time process. Another way is to make specific assumptions about the discrete time hazards and give up any attempt to draw the bridge between the two types of model specification. Obviously, which way is chosen is normally dictated by the objective in mind and simplicity of estimation. Both ways have their advantages and drawbacks which relate to treatment of function $\gamma(i)$ in (2.21), known as the interval or discrete baseline hazard.

As regards the first way of estimation, if one assumes a particular type of duration dependence in the continuous time baseline hazard, $\lambda_0(t)$, it is not difficult to derive a corresponding expression for $\gamma(i)$. This, however, implies that an investigator may place excessive restrictions on hazard $\lambda_0(t)$. In general, the choice of the functional form for $\lambda_0(t)$ should not be made solely on the grounds of convenient mathematical properties and tractability of the closed form expressions. A variety of potential factors (an impact on search intensity of the exhaustion of unemployment benefits, discouragement

⁴⁴A natural extension to this would be generalise further the split-population model to allow individual characteristics to affect separately both the probability of eventual failure and the timing of failure. In that case, the proportions defined in the split-population model will depend on a vector of characteristics (not necessarily coinciding with the vector of characteristics employed in the hazard function), i.e. not only a constant. Such approach is used, for example, in Schmidt and Witte (1989), who study a criminal recidivism phenomenon, while labour economics applications include, for example, Pudney and Thomas (1995) and, most recently, Addison and Portugal (2003). We have also attempted incorporation of additional covariates in the split-population equation but the numerical routine written in Ox (Doornik, 2001) and using the BFGS algorithm (see the main text) has repeatedly produced unstable results. In particular, the variance-covariance matrix for the model coefficients could not be numerically evaluated, although convergence was achieved without problems. Such a result might be explained by specificity of the data, fragility of the models with unobserved individual heterogeneity, and the presence of the regions where log-likelihood is effectively flat so that the BFGS method fails (see, e.g., Mascarenhas, 2004).

effects, dynamics of the reservation wage, mechanisms of screening of job applicants by employers, etc.) can suggest different shapes for the hazard rate in relation to survival times. Moreover, duration dependence can turn out to be negative or positive depending on the impact of the dominating factor. However, without a proper structural model of the labour market one cannot identify which factor would dominate the others and, thus, unambiguously define the hazard's functional form. Equally, it is important not to preimpose a particular shape on the hazard function, as the choice of the functional form can affect results of the analysis when the frailty factor is introduced (see, e.g. Trussell and Richards, 1985; Dolton and Von der Klaauw, 1995).

A more flexible approach is to take a somewhat agnostic position about the duration dependence in the continuous time hazard, and make assumptions about the duration dependence in $\gamma(i)$. The examples of common specifications for the discrete baseline hazard include logarithmic, piecewise constant, and p -th order polynomial functions (Jenkins, 2004). However, in that case it is impossible to identify parameters characterising the shape of $\lambda_0(t)$ by estimating $\gamma(i)$. This happens because $\gamma(i)$ summarises differences in the values of integrated hazard function $\Lambda(t)$ and is consistent with a number of different shapes of continuous hazard function $\lambda_0(t)$ within each interval.

One of the objectives of this work is try to identify the presence of defective risks that lead to specific distributions (see the main text) of survival times. We pursue this goal by estimating both polynomial and piecewise constant functions $\gamma(i)$. Below we argue that this method does not come at the cost of some additional arbitrariness, while providing information sufficient to reject a case of the specific distribution (and a specific shape of the continuous time hazard $\lambda_0(t)$) and necessary (but not sufficient) to accept it. Before explaining the details we first briefly raise the question of identifiability of defective risks, an issue of special importance for the types of model that we consider.

The principal issue that an investigator stumbles upon while dealing with individuals trapped in joblessness is identifiability of the infinite durations *per se* that she seeks to recognise in the data. Without resorting to parametric modelling, under ideal data conditions (when it is possible to observe individuals over an infinite time horizon) there should be no difficulty in identifying the presence of defective risks leading to the existence of long-term survivors. In practice, however, much depends on the length of the follow-

up period (see a discussion in Addison and Portugal, 2003). Non-parametrically the presence of defective risks would be indicated by an empirical survivor function that does not converge to zero but rather to some positive constant as time goes to infinity. This implies that after some point, corresponding to the *maximum noncensored time*, all actual survival times will always be censored (*ibid*). The length of the follow-up period is of critical importance here. If it is too short then censored observations will be generated by both movers (i.e. those whose durations are to be finite) and stayers (with infinite durations). For example, with the data used by Addison and Portugal (2003) the authors feel confident that the follow-up period of 98 months is sufficient to detect long-term survivors. In the case of the samples that we construct from the RLMS data, the maximum duration of a (incomplete) spell is 90 months for exits from unemployment and 100 months for exits from inactivity.⁴⁵

The empirical survivor functions (obtained by using the life-table estimator - a discrete version of the Kaplan-Meier product limit estimator) for both samples are given in Fig.2.10 (vertical lines represent a 95%-confidence interval). On the one hand, there is no clear indication of convergence of the survivor functions to positive constants for duration lengths corresponding to within-sample values. On the other hand, the convergence to positive values may or may not happen for duration lengths not covered by our samples. Notice that even if long-term survivors do exist among the unemployed, their proportion is likely to be very negligible, as indicated by top panel of Fig.2.10 (where the survivor function nearly converges to zero for within-sample lengths of the unemployment spell). At the same time, the existence of long-term survivors is more likely for the inactive, among whom about 20% are still in inactivity after having spent about 100 months in that state. Still the span of the 95%-confidence interval suggests that the survivor function for the inactive may just exhibit a substantially slower convergence to zero as compared to the survivor function for the unemployed. In both cases, one cannot unambiguously argue for the case of infinite durations.

As we have just seen, a non-parametric investigation can or cannot be of substantial help with the identification of infinite durations. Much depends on the quality of data. The parametric identifiability of the defective risks hinges upon the shape of the hazard.

⁴⁵For completed spells these lengths are 83 and 95 months respectively.

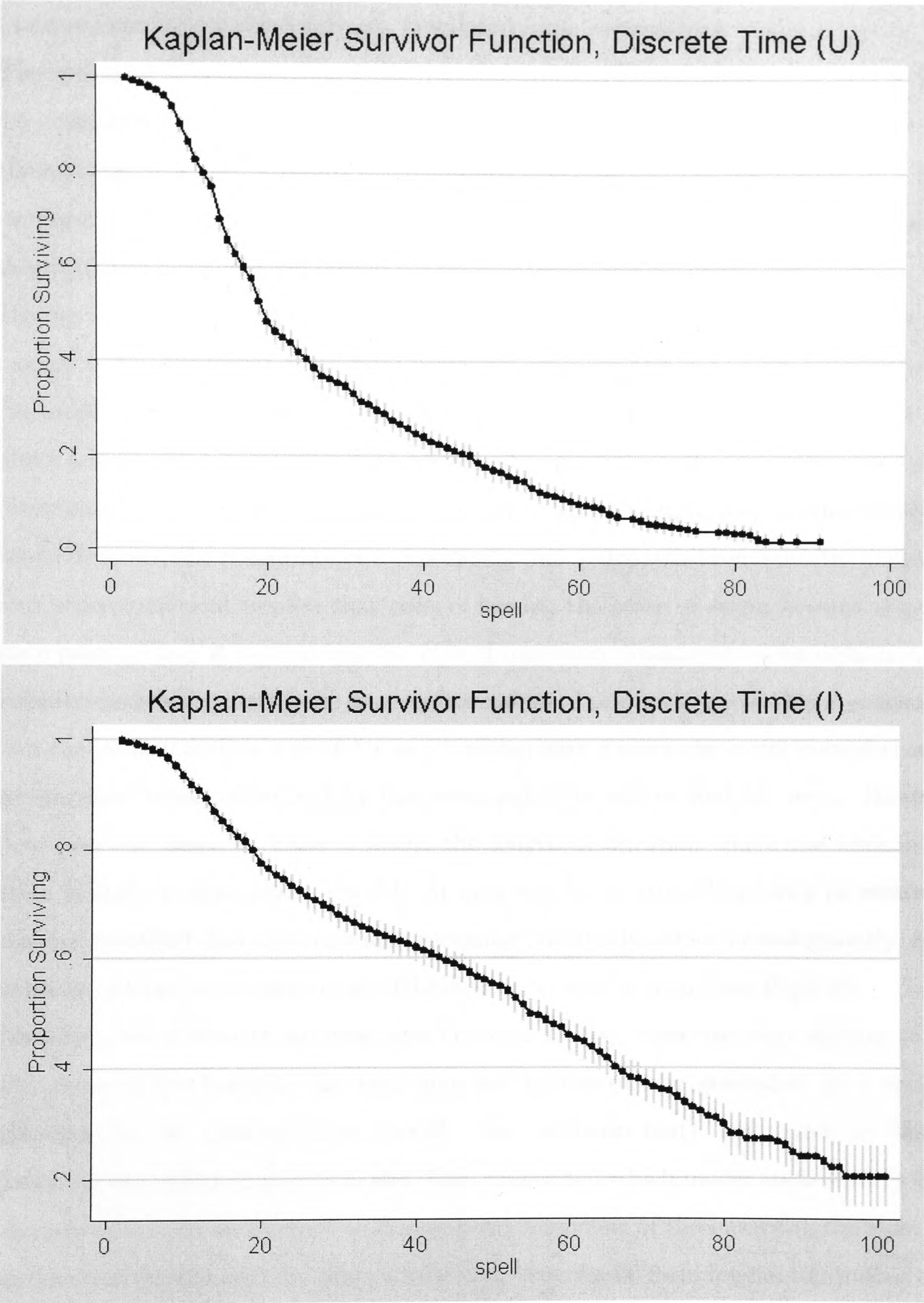


Figure 2.10: Empirical survivor functions for the unemployed and the inactive

Differently speaking, the identifiability of the defective risks (or specific distributions of that matter) can be obtained through functional form assumptions.

The specific distributions that we are interested in are characterised by a specified by of the continuous time baseline hazard. In particular, we want to ascertain if a specified continuous function with a piecewise continuous derivative as the one drawn by the solid line in Fig.2.4. The hazard of this form is likely to be the maximum of the constant c (with $c = 0$) and a polynomial function with a negative highest order coefficient. For coefficients

Having said that, one may think that in order to identify specific distributions of defective risks just needs to fit a polynomial of some prechosen order to the data, and see whether or not its coefficient on the highest order term is negative. However, the investigation of this problem runs into a series of computational problems. Whenever hazards are parametric (in this very case) duration models are normally estimated by using the maximum likelihood method. However, the possibility of a continuous time polynomial hazard with a negative highest order coefficient implies that risks of leaving the state of origin become negative at some point. Thus, it undermines the idea of maximum likelihood, as by definition a probability modelled cannot take on negative values. In order to avoid this one needs to restrict the hazard, and then model it as a function with a piecewise continuous derivative where negative values generated by the polynomial are substituted for zeros. However, in that case one needs to know *a priori* the length of duration where the kink where the function is likely to take place (Fig.2.4). It may still be no stumbling block to successful estimation, provided that data can help recognise infinite durations unambiguously, but it is clearly not the case made by the RLMS data as can be seen from Fig.2.10. In Fig

Moreover, as noticed in Addison and Portugal (2003), defective risks leading risks specific form of the survivor function may not necessarily be generated by a specific distribution (or the random draws model - see the main text), but rather by it rather derived from the split population model. One cannot tell which model applies solely from the duration data. In an attempt to find out the behaviour of the hazard in the regard to the duration distribution, by using a restricted polynomial form for the continuous time baseline hazard the researcher may impose too much structure on the data. The new data that we suggest here seems to be more flexible in that respect.

Instead of considering the continuous time baseline hazard, $\lambda_0(t)$, we suggest for we suggest

on the discrete time baseline hazard, $\gamma(i)$, and modelling it as either a polynomial or a piecewise constant function. Given the definition of $\gamma(i)$ in (2.21) above, it is above that if $\lambda_0(t)$ assumes a form corresponding to a specific distribution (see Fig.2. on (see $\gamma(i)$ must converge to $-\infty$ as t goes beyond the value of duration corresponding to the maximum length of spell with a positive probability of exit. This is so because as t approaches that value, the difference in the integrated hazards tends to zero (tends to zero) it remains zero for any t greater than the value at which the kink in the function in the function while the logarithm of the difference tends to minus infinity. Thus, assuming a polynomial or a piecewise constant form for $\gamma(i)$ does not preclude the identification of such a possibility of convergence to $-\infty$ which is possible if, for example, the highest coefficient is negative (in the case of a polynomial) or estimated constants are weakly zero (in the case of the piecewise constant specification). At the same time, it cannot impose a specific form on the continuous time baseline hazard, and is consistent with various shapes.

It should be noted, however, that convergence to $-\infty$ in $\gamma(i)$ is necessary but not sufficient for confirming the presence of a specific distribution. To illustrate this, consider another distribution free of a singularity segment and with the continuous time hazard smoothly converging to zero, as shown by a dashed line in Fig.2.4 (this may be, for instance, if, for instance, it is log-logistic). Then as $t \rightarrow \infty$ the difference between integrated hazards in (2.21) also tends to zero, while the logarithm of it converges to $-\infty$. Having seen this, we still argue that if there are no signs of convergence to $-\infty$ in the discrete time baseline hazard, $\gamma(i)$, for the duration values observed in the sample then it can be claimed that no case for a specific distribution can be made. We use this logic for interpretation of findings in Section 2.3.4 in the main text.

2.B.10 Summarising the Assumptions

We conclude this appendix by summarising the main assumptions that we use to estimate the three types of the competing risks model. They are:

- 1) Independence in competing risks;
- 2) Proportionality of the destination-specific hazards;
- 3) Parametric (Gamma) unobserved individual heterogeneity;

4) Transitions out of the origin state can be treated as if they occurred at the boundary of observed intervals (months);

5) Polynomial or piecewise constant interval (discrete time) baseline hazards.

The first three of these assumptions are common in the duration literature (see some of the references above). The fourth is made here for computational simplicity, although it is likely that the way that we organised the data does not make this assumption be too strong. Finally, the last assumption is also a common way to specify the interval hazard in order to avoid imposing restrictions on both interval and continuous time baseline hazards (Jenkins, 2004). Moreover, by allowing a flexible form for the continuous time baseline hazard (which is guaranteed by assuming either a polynomial or piecewise constant form for $\gamma(i)$), one is warranted that a possible misspecification of the unobserved heterogeneity distribution has almost no consequences to the results of estimation (see Dolton and Von der Klaauw, 1995).

Chapter 3

Diverging Paths: Transition in the Presence of the Informal Sector

3.1 Introduction

It has been about 15 years since Central and East European countries (CEEC) as well as the former Soviet Union republics (FSU) embarked on a process of structural reforms to abolish the former system of central planning and establish economies based on market principles. It has already become popular to talk about the end of transition (for some discussion see, e.g. Svejnar, 2002) as many CEE countries have joined the European Union. On the theoretical front, however, the creation of a coherent model of transition is still open to dispute.

The transition in CEE countries has led to a U-shaped response of output (Blanchard, 1997). That is, a sharp recession that all the countries started to experience right after the beginning of reforms, was followed by a recovery. However, many countries – members

of the Commonwealth of the Independent States (CIS, the successor of the FSU) – are still close to the bottom of the U-curve.

It is widely held that the principal driving force of the economies, capable of pulling them out of the downside of the U, is the private sector that widely emerged at the beginning of transition. Before that, private initiative was tolerated almost exclusively in agriculture while non-agricultural self-employment was either non-existent or confined to the shadow economy (Boeri and Terrell, 2002). The idea of a buoyant private sector proved so attractive that it was absorbed by many economists studying the transition. A lot of theoretical work was put forward in which transformation was modelled as a process of a more or less gradual fade-out of the state sector and the flowering of the private firms. This view had an important implication for studies of labour markets of transition economies.

In a number of papers (Burda, 1993; Aghion and Blanchard, 1994; Rodrik, 1995; Blanchard, 1997; etc.) processes taking place in transitional labour markets were represented as shifts of labour from the shrinking state (traditionally called "old") sector to emerging private firms (the "new" sector). However, those shifts were not usually seen as a direct movement of labour from one sector to the other: periods of employment were separated by unemployment spells. In a nutshell, the simple mechanism can be described as follows. Workers, fired from state enterprises during their restructuring, become unemployed and start to search for a job in the growing private sector. After some time more workers are hired in private firms than are laid off in the old sector, which leads to a decrease in unemployment.

These studies have become known as the OST (for Optimal Speed of Transition)

literature, since many of them try to figure out what speed of labour reallocation would be optimal for the economies. Usually, such work draws upon the Harris and Todaro (1970) two-sector migration model adapted to a transitional setting.

However, quite recently attention has been drawn (Boeri, 1999, 2000a,b) to the fact that this literature does not provide a satisfactory explanation of output trends in many transition economies nor does it explain a number of stylised facts concerning labour markets. The flowering of informal economies (Johnson *et al.*, 1997; Schneider and Enste, 2000) and a decrease in participation rates have been widely observed in Eastern Europe (Boeri *et al.*, 1998), and, particularly, in Russia (Standing, 1998), but have been ignored by the majority of OST studies. Little attention has been given to labour supply (Boeri, 1999, is among rare exceptions). On the policy prescription front, the OST models have not gone further than pointing to the speed of scrapping of the inefficient sector, the level of unemployment benefits and minimum wages, as crucial in leading to the success in transformation. However, recent empirical investigations (Jurajda and Terrell, 2002, 2003) find a striking resemblance between the patterns of old-to-new sector reallocation across countries that pursued different approaches to the reform of the old sector and established varied labour market institutions. It appears that differences in macroeconomic policies have mainly manifested themselves in the aggregate level of unemployment and wages, but not in the composition of the new sector. Other models (such as Caballero and Hammour, 1996, 2000a,b), developing the Schumpeterian idea of "creative destruction", similar in spirit to the OST models but still not directly applicable to the transitional setting, have recently been used to fill in the gaps left by the OST literature. In particular, it has been suggested that it is the importance of the right policy environment for

promoting new start-up firms that matters after all.

In this work we argue that the critique of earlier OST models (such as, e.g. Aghion and Blanchard, 1994) is correct only in pointing to the fact that those studies do not account for a number of labour market flows observed in reality. Still, the earlier models offer a general picture of transition that was hardly significantly modified in more recent and more elaborate work, allowing for better treatment of job-to-job movements and job search intensity (Brixiova, 1997), productivity shocks affecting the old sector (Garibaldi and Brixiova, 1998), labour supply (Boeri, 1999), capital relocation (Castanheira and Roland, 2000), moonlighting (Bouev, 2001), etc. However, the substantial deficiency of the OST models is still their inability to account for the informal labour market, the underground economy, whose presence may suggest significantly different dynamics of output in transition.

Further, we contend that the rent appropriability problem, embedded in the Caballero and Hammour (1996) set-up and crucial to policy implications underscoring the importance of creating vigorous incentives in the expanding sector, is effectively present in the seminal Aghion and Blanchard (1994) work. At the same time, the more recent strand of the OST studies (*inter alia* Brixiova, 1997; Garibaldi and Brixiova, 1998; Boeri, 1999, 2000b) overlooks this aspect because of slightly different assumptions regarding functioning of markets and matching between firms and workers.

The goal of this essay is to provide an example of a simple model, an extension of the work by Aghion and Blanchard (1994), that is not only able to generate the type of dynamics suggested by the previous OST studies, but also shows that qualitatively different equilibria are possible for various sets of policy parameters in the presence of

the informal sector. We suggest that while a choice of the pace of transition (a principal instrument in determining the success of transformation in preceding OST studies) is important for the total welfare of the economy, it is the policies towards the emerging new firms and informal sector that define the eventual outcome of transition. In particular, we stress the importance of certain labour market conditions (such as wage mark-ups in the emerging new and informal sectors) as well as of a broader effect of various institutions on the comparative profitability of new formal and informal firms. The numerical simulations indicate that the dynamics of a sustained L-type experienced by Russia and other former USSR countries during the 1990s may be explained by the convergence to a steady state qualitatively different from the one probably pursued by the East European transition economies.

The chapter is organised as follows. The next section is devoted to the critical review of OST literature. Section 3.3 presents a simple model of transition in the presence of the informal sector, derives main theoretical results, considers numerical simulations, and explains main findings. Section 3.4 concludes.

3.2 Background

3.2.1 Optimal Speed of Transition Literature

Theoretical models: the picture of transition

The experience of Central and Eastern Europe in the early 1990s was that of the output decline at the start of the process of economic transformation. At that time inefficient state enterprises began to restructure or close down, unemployment grew as capital and labour

were reallocated across sectors, while the new private firms emerged. A number of scholars (to name just a few - Burda, 1993; Aghion and Blanchard, 1994; Rodrik, 1995; Atkeson and Kehoe, 1996; Commander and Tolstopiatenko, 1996; Gavin, 1996; Ruggerone, 1996; etc.) took up the challenge of explaining the transitional output fall and unemployment theoretically and initiated a strand of the literature later called the Optimal Speed of Transition (OST) work. The name comes from the attempts of many authors to look into the problem of finding the speed of resource relocation that would be socially optimal. The issue was especially relevant in the light of the ongoing debate about gradualist versus shock therapy approaches to reforms in the region (e.g. Roland, 1996).

As Boeri (1999) argues, all such OST studies can be seen as developing the ideas of the Harris and Todaro (1970) migration model in the transitional setting. Typical models consider the process of search in the labour market, employing the matching function (Pissarides, 2000) and the labour market flow approach (Blanchard and Diamond, 1992). The two main exceptions, Castanheira and Roland (2000) and Castanheira (2003), use the Ramsey growth model as a framework. All the OST models can also be split into two main categories: partial equilibrium (Aghion and Blanchard, 1994; Ruggerone, 1996; Brixiova, 1997; Shimer, 1997; Garibaldi and Brixiova, 1998; Boeri, 2000b, etc.) and general equilibrium models (Atkeson and Kehoe, 1996; Castanheira and Roland, 2000; and Castanheira, 2003). The general equilibrium models as a rule devote less attention to labour markets but provide a very good set-up for the analysis of various economy-wide policies. The partial equilibrium models focus narrowly on labour market flows and policies, but their numerical simulations are still broadly consistent with those of the general equilibrium models and the main stylised facts. In particular, all the

models feature a declining old sector, a growing new sector, inverted U- or L-shaped development of unemployment. Job creation comes from the new firms, while the old sector is responsible for most of the job destruction. Reallocation of workers is driven by job destruction and creation. At the beginning of transition job destruction dominates over job creation, but in a later stage they come into balance (see Haltiwanger *et al.*, 2003, for a discussion of main stylised facts).

Speed of transition matters

In addition to being in seemingly good accordance with the early evidence, the OST models suggested their own vision on the gradualism versus big bang debate. They addressed the issue by showing that too slow or too high a speed of transition (in their context a speed of scrapping of old sector employment) can have an adverse effect on the outcome of transformation. Thus, on the one hand, Aghion and Blanchard (1994) and Chadha and Coricelli (1994) paid close attention to explicitly introduced fiscal externalities. In the economy where unemployment acts as a disciplining device (Shapiro and Stiglitz, 1984) and lowers worker's pressure on wages in the emerging sector, the government may be tempted to go for excessive scaling down of old enterprises. This, however, will lead to an increasing tax burden on new sector firms as the growing number of unemployment benefit claimants requires higher tax rates in the emerging sector while the old sector tax base is eroding. Transition may be brought to a halt because of the adverse effect of taxation. At the same time, too slow a speed of transition can be detrimental as well: low unemployment can squeeze new sector profits by driving up wages. On the other hand, Castanheira and Roland (2000) suggested another transmission mechanism, namely savings. For example, if the speed of transition is too high, a high unemployment

rate means lower savings and, hence, investment that eventually leads to depression of output. The optimal speed should level off various counterbalancing effects and maximise the economy's net output (Burda, 1993; Aghion and Blanchard, 1994; Brixiova, 1997) or life-time utility of consumers (Castanheira and Roland, 2000; Castanheira, 2003). What is the optimal speed in practice? The question was left to empirical research.

Are the OST studies any good?

More recent OST studies have confronted their predecessors with the data more thoroughly and argued that they fail to incorporate several features of the transitional labour markets that are very important in reality. In particular, it has been claimed that the incorporation of job-to-job movements and the heterogeneous labour force (Boeri, 1999, 2000a,b) may help refine the implications for the optimal speed and, in addition, explain such stylised facts as, for example, stagnant unemployment pools or a drop in labour force participation in East European countries. Brixiova (1997) introduces job-to-job movements and shows that the duration of unemployment rises, while the optimal speed lowers as the jobless need to compete for new matches with workers searching on-the-job. Boeri (1999, 2000b) goes further and points out that heterogeneous opportunities for working in the subsistence sector may explain the drop in participation rates and the stagnancy of the non-employment pool in the presence of high non-employment benefits.

Despite this success the recent OST models still turn out to be less good at explaining the behaviour of old and new sector wages as well as the behaviour of transition probabilities out of unemployment. The main stylised facts in this respect are those of average wages falling at the start of transition, but starting to recover soon after (Basu *et al.*, 2000; Jurajda and Terrell, 2002), and new-old sector wage differentials being very

significant at the start of reforms, but decreasing remarkably afterwards (see, e.g. Basu *et al.*, 2000; Jurajda and Terrell, 2003). At the same time, transition probabilities out of unemployment have also decreased over the course of transition in many countries (see table 2.5 in Chapter 2).

Why is the recent OST literature weak on this conformity to the evidence? On the one hand, in their work Castanheira and Roland (2000) make an impromptu simplifying assumption of perfectly competitive labour markets and free movement of labour between sectors, so that there is no intersector wage differential over the course of transition. Later Castanheira (2003) corroborates the absence of wage differences in a similar model by evidence on declining new-old sector wage differentials. On the other hand, Brixiova (1997), Garibaldi and Brixiova (1998) and Boeri (1999, 2000b) extended a one sector model by Pissarides (2000) to depict the transition as essentially an equilibrium phenomenon in the absence of entry barriers, when wages, as well as labour market transition probabilities, are constant throughout the process of transformation. Still, even in this set-up, Garibaldi and Brixiova (1998) show that the average real wages fall and then recover in the course of transition by means of introducing different productivity states in the old sector, each with a fixed wage. As the number of jobs with different productivity varies in the old sector the average wage in the economy changes as well.

Another weak spot of the recent OST models can be seen in the fact that although they incorporate more labour market flows into the benchmark of Aghion and Blanchard (1994) and allow for the better treatment of labour market institutions, the numerical simulations of those models still are remarkably similar to those implied by the original. This happens because what is important in the end is not the incorporation *per se* of

on-the-job search and other realistic characteristics of the labour market, but the relation between the exit rate from unemployment to employment and the wage in the new sector. This relation is essentially the same in all the models, including the most recent. The bells and whistles of the latter are undoubtedly useful policywise but they do not change qualitatively the predictions of the backbone model: in transition the economy converges at a faster or lower speed to a steady state with the dominating new sector, unemployment follows an inverted U or L-shaped (if layoffs in the new sector are allowed) trajectory, until flows in and out of the pool of job-seekers are balanced. This qualitative consistency of the OST studies with stock adjustment in CEEC is acknowledged in Boeri and Terrell (2002).

Why do CEEC and the CIS differ?

There is one more important question not successfully addressed by the OST literature, namely the diverging paths of CEE and the CIS. The two regions started the transition in broadly similar initial conditions. However, the structural change has proceeded at a faster pace in CEEC than in the CIS, the development of the private sector and of non-agricultural self-employment has been much faster in the former than in the latter, unemployment reached the peak much earlier in CEEC than in the CIS and is more stagnant in Central and Eastern Europe (see Chapter 2 and, also, Boeri and Terrell, 2002). Over the 1990s the evolution of output and employment in the CIS resembled more a L-shaped trend, while CEEC enjoyed U-shaped or even J-shaped patterns both in employment and output (for the evolution of real GDP see Fig.3.1). Boeri and Terrell (2002) argue that there are at least two different labour market adjustment trajectories in the transitional arena: the one followed by CEEC experiencing significant employment ad-

justment, fast structural change, and high unemployment (most of which long-term), and the other one pursued by the FSU (except for the Baltic states) with low responsiveness of employment to output changes, strong and persistent wage declines, slower structural change, and a more gradual build-up of unemployment which is also characterised by a relatively large turnover rate (Fig.3.2 demonstrates the development of unemployment in the two regions).

A possible explanation of the differences could simply be sought in the different speed of reforms in the two regions. In particular, it matters how quickly the old sector has been scrapped. A different pace of old sector downsizing leads the economy to different equilibria. From the Aghion and Blanchard (1994) work it follows that these equilibria differ only in the resulting level of unemployment and wages. Indeed, this can explain why unemployment in CEEC reached the peak before the same happened in the CIS. However, the divide between the two regions seems to have had other, *qualitative* differences. Boeri and Terrell (2002) point out that Russia and Ukraine adopted much more aggressive privatisation strategies than Poland or Slovenia and yet output and employment growth in the former two have been inferior compared to the latter two. So, the authors suggest that the differences have little to do with the alternative between rapid and gradual transition.

An attempt to explain the divergence in the behaviour of real wages and unemployment in CEEC and the CIS has been made in Garibaldi and Brixiova (1998). The authors argue that differences in such labour market institutions as minimum wages and unemployment benefits should be analysed together with differences in the labour market dynamics. Boeri and Terrell (2002) have supported this point too. They stress that the policy trade-offs

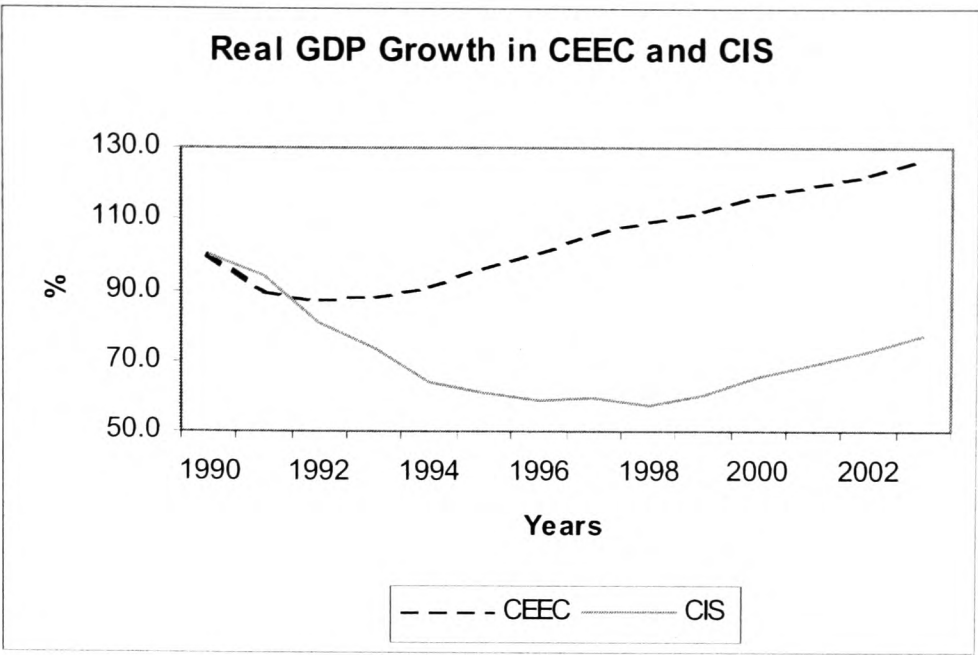


Figure 3.1: Divergence of CEEC and CIS countries: real GDP growth (CEEC including Central and Eastern Europe and the Baltic states, excluding South Eastern Europe and the Balkans; source: EBRD, 2003)

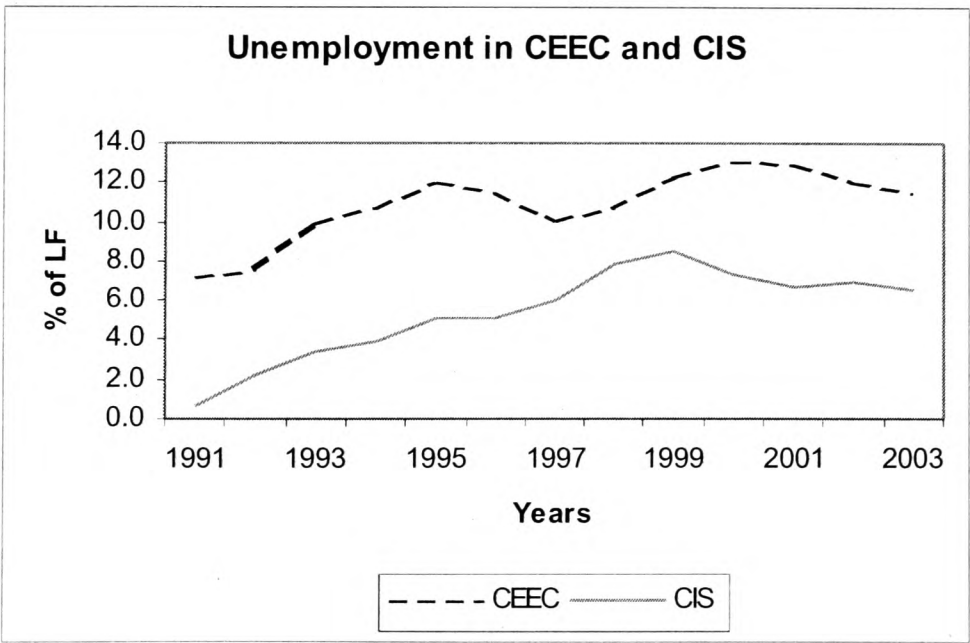


Figure 3.2: Unemployment in CEEC and CIS (excluding Azerbaijan, Kyrgyz Republic, Tajikistan, and Uzbekistan; sources: EBRD, 1999, 2000, 2003, 2004)

embedded in the OST literature have related mainly to the alternative between a big-bang strategy and a gradual transition process. At the same time, the OST theory has not been able to frame the trade-off between employment and wage adjustment in the old sector. Boeri and Terrell (2002) point to the institutional side of the labour markets, namely the design of social benefit systems, minimum wages, and adjustment of wages in general and in the old sector in particular, as at least in part explaining the CEE and CIS paths.

Shifting away from the OST

It seems that what the OST models have so far focused on is the policy towards the inefficient old sector (for the most recent example see Castanheira, 2003, who studies at more length the subsidisation of state firms), in particular, the speed of its closure as well as the difference between social models adopted in different countries.

However, the evidence suggests that this is not enough for satisfactory explanation of differences in the experience of countries within CEE, and, in particular, between CEEC and the CIS. So, EBRD (2000, p.97) writes: "It has become increasingly clear that the distinction between fast and slow speeds of reform camouflages many important similarities and provides limited guidance on the policies that need to be taken." Moreover, some recent empirical studies (Jurajda and Terrell, 2002, 2003) find a remarkable similarity in the patterns of old-to-new sector reallocation across two CEE countries that pursued different approaches to the reform of the old sector and also established varied labour market institutions. They point out that it appears as if the difference in macroeconomic policies has mainly manifested itself in the aggregate level of unemployment and wages, but not in the composition of the new sector. In other words, between the CEE countries the reallocation patterns have been broadly similar despite the use of different policies

that according to the OST theory should have produced different types of adjustment.

As regards the CIS versus CEEC experience, as opposed to what happened within CEE, the argument of Garibaldi and Brixiova (1998) and Boeri and Terrell (2002) is that the low unemployment benefits and minimum wages, observed in the CIS, may have indeed facilitated there the drop in wages and supported employment in the old sector. However, if one follows the logic of the OST theory, lower wage floors in the CIS should have reduced wage pressure and increased profit margins in the new sector, and thus lead to its faster development than in CEEC. However, it has never happened.

Because of the weaknesses of the OST theory other models expounding the transition in Eastern Europe have gained in popularity. In particular, a strand of the developing economy literature, building upon the Schumpeterian idea of "creative destruction" (in particular, Caballero and Hammour, 1996, 2000a,b), has been invoked, although some assumptions (e.g. profit maximisation in the inefficient sector) do not make it directly applicable to the transitional setting. The literature shares many similarities with the OST theory but also pays some special attention to inefficiencies in the job creation process. In particular, it stresses that if investment (in its broad sense) in the expanding sector is specific, the generated quasi-rents may be appropriable by workers, unions, or the government, increase investment costs to firms, and slow down the adjustment process. This takes the debate on gradualism to a new level. It is not only the speed of destruction that matters: boosting job creation through eliminating inefficiencies is immensely important too. Caballero and Hammour (1996) argue that policy analysis must go beyond the "gradualism versus cold turkey" debate, and examine managed adjustment policies needed in the face of unbalanced restructuring. The economy's ills are ultimately institutional in

nature. An appropriate macroeconomic policy should create a favourable climate for job creation in the expanding sectors and support employment in the contracting sectors.

Specific assumptions

This part of the argument is well understood by many scholars contributing to the development of the OST theory. In general, many agree that the superior growth performance of new private firms in transition seems to raise important policy questions regarding the development of an economic and legal environment which is conducive to start-ups (Jurajda and Terrell, 2000; Haltiwanger *et al.*, 2003). Boeri and Terrell (2002) note that the CIS, for instance, is known for significant entry barriers associated with the mafia and smuggling that slow down the take-off of the new sector. Johnson *et al.* (2000b) provide evidence supporting Johnson *et al.* (1997) that Eastern Europe and the FSU are diverging largely because of differences in the protection of property rights. They note that efforts should be made to stabilise the country's regulatory environment and to develop market-supporting infrastructure. Haltiwanger *et al.* (2003) stress that this point ties in with the widely held view (see e.g. EBRD, 2000) that it is the differences in legal infrastructure that mainly explain the diverging paths of CEE and CIS transition countries.

In order to effectively address this aspect of transition at a theoretical level one should scrutinise creation of job places in the new sector. However, it is difficult, if at all possible, to do with the help of the most recent OST studies (e.g. Boeri, 2000) because job creation is modelled there in a specific way. In particular, it is assumed that in the absence of entry barriers and with perfect capital markets, and while workers and firms have perfect foresight, the value of a created vacancy in the new sector is constant in and out of

the steady state of a system of differential equations describing dynamics of the economy. Then, given the assumption of constant returns in the matching function, vacancy creation always has to respond proportionally to the size of the pool of job-seekers, as if firms were able to close or open up vacancies instantaneously, or, in other words, enter or leave the market unimpeded. Thus, vacancy creation turns out to be a jump-variable so that it keeps the economy on the equilibrium transition path along which the market tightness is constant. This drastically simplifies the analytical complexity of such models. Then the economy always resides in a steady state of a system of Bellman differential equations, while the transition reallocation is essentially an equilibrium phenomenon. Such modelling of the job creation process, while being a fair theoretical exercise, assumes no inefficiencies and defies analysis of many possibilities that could slow down the development of the new sector. Not surprisingly the attention of students of transition has started to shift away from the OST theory.

However, it turns out that the backbone model by Aghion and Blanchard (1994) still has something to offer. In contrast to more recent studies it does not assume that transition occurs in perfect circumstances where firms can immediately inundate the market with vacancies if need be. Instead, vacancy creation becomes proportional to the pool of job seekers only in a steady state. Out of it the economy develops under the specific assumption that new sector firms do not face matching problems, i.e. they can easily fill a vacant position with a job-seeker. This is a realistic hypothesis in the presence of high transitional (often involuntary) unemployment and some entry barriers. This particular type of matching allows the authors to simplify the algebra. In such a set-up vacancy creation becomes more explicit and allows closer examination. In particular, it is assumed

that vacancy and, hence, job creation is a function of profits which are not driven to nil by firms freely flooding the market. Thus, the vacancy rate no longer jumps to keep market tightness constant, so that one can address explicitly the question of influence of various factors on the ability of existing firms in the new sector to expand. This suggests that a parallel to the Caballero and Hammour (1996, 2000a,b) approach may be drawn. In order to facilitate solution of a system of Bellman differential equations out of steady state Aghion and Blanchard (1994) make a simplifying assumption of a wage mark-up in the new sector, justifying it by efficiency wage considerations. However, it can readily be shown that it is also, in fact, one of the ways to represent the rent appropriability problem in the spirit of Caballero and Hammour (1996). We shall return to this point later in Section 3.3.7.

Summary

So, where do things stand? In this section we have seen that all the OST models provide a similar picture of transition despite a variety of their modifications put forward. On the policy front the literature emphasises the importance of proper engineering of the job destruction in the old inefficient sector, and a reasonable choice of social support programmes (unemployment compensation and minimum wages). The job creation part of the transition process has received little or no attention in the theory, apart from the benchmark model. Other literature from the development economics realm has been enlisted to fill in this gap and to provide a better guidance on policy making. The major question still to be answered is why the CEEC and CIS experience has been so different. Thus far no theoretical model has been satisfactory in giving a clue.

In the next subsection we review some more stylised facts that have been overlooked

by the OST studies. Then we go further to suggest our own modification to the theory.

3.2.2 Expanding the Seminal Model

Facts to be explained

Criticising the OST literature Boeri (1999, 2000a,b) points out that the drop in participation rates is what the theory should *inter alia* incorporate to provide a better account of reality. Boeri and Bruno (1997), Boeri (1999), and Chapter 2 of this thesis find that flows into inactivity are widely observed in both Eastern Europe and the CIS and take place mainly through unemployment. This drop in formal labour market activity is likely to be connected with a rise in the share of the informal sector (Schneider and Enste, 2000). For instance, EBRD (2000) refers to the importance of new private sector activity in the informal economy.

The striking stylised fact is that while the informal sector was expanding in many transitional economies in the beginning of the 1990s (see, e.g. the data provided by Johnson *et al.*, 1997; Lackó, 2000; Feige and Urban, 2003), its presence has especially been notable in the FSU countries such as Russia or Ukraine. Even in such a rapidly and successfully reformed economy as Estonia at the beginning of transition hiring and creation disproportionately occurred for workers with informal or temporary contracts (Haltiwanger and Vodopivec, 2002). In the CIS irregular activities probably still keep on mounting. EBRD (2000) reports that employment in the CIS has a higher informal share than in CEE.

These facts suggest that flows to inactivity and informal employment have been important in labour force reallocation in transition. The way that economies develop cannot be

completely described without taking such phenomena into account. The truth, however, is that these facts have been passed over by the majority of the OST studies.

Moving to modelling

Not all the OST studies have ignored the role of the informal sector in the adjustment of transitional labour markets. So, in their contribution to the OST thought Commander and Tolstopiatenko (1997) have modelled the new sector as the one representing the informal economy. Boeri (1999, 2000b) has focused on the role of the subsistence sector in affecting labour reallocation. Both models are essentially of a two-sector type, and either look at transitions between formal and informal parts of the economy (Commander and Tolstopiatenko, 1997) or assess the reallocation between the old and new sectors when implicit outside opportunities are present (Boeri, 1999, 2000b). Technically their results are little different from the rest of the OST literature - labour moves between two destinations in the presence of unemployment (either active or passive as in the case of Boeri's models). However, in these models the presence of the informal sector in the course of transition does not generate any adverse externalities on the rise of the new formal sector in which development the ultimate objective of transformation is contained. Meanwhile, Caballero and Hammour (1996) point to the existence of such externalities. In particular, they argue that many crisis situations are characterised by quick destruction of jobs in one sector and sluggish creation of jobs in another sector, so that an employment problem develops. Many workers who lose their jobs in the contracting sectors find themselves either in overt unemployment, if some form of unemployment compensation is available, or being forced to take on employment in the informal sector. The authors show that the presence of the informal sector can slow down the adjustment process. This happens,

for example, when its productivity rises, so that the possibility of engaging in informal sector activity strengthens the workers' threat point in wage negotiations with the formal expanding sector. Obviously, it squeezes profits and hampers job creation process, especially in the presence of specific investment that can be held up by workers.

Keeping in mind the differing experience of CEEC and the CIS as far as the development of the informal sector is concerned, one can ask: Is it possible that the diverging paths of the two regions in general are explained by such an interplay of the new formal and informal sectors in the presence of inefficiencies similar in spirit to the ones described by Caballero and Hammour (1996, 2000a,b) and, perhaps, some other factors?

In the next section we offer our own vision of what has been going on in CEEC and the CIS. We suggest a development of the Aghion and Blanchard (1994) model to incorporate an informal sector. We then show that this addition generates qualitatively different equilibria from those present in the OST literature. In particular, the economy does not necessarily always converge to a steady state where the new sector prevails in the end of transition. In some cases the informal sector wins. We believe that the existence of such equilibria may explain the divergence of the course of CEE and CIS countries in the 1990s. We illustrate how important the interplay of the formal new and informal sectors can be, and that it is the policy towards the new formal sector that matters after all for the eventual success of transformation. A direct parallel to arguments by the Caballero and Hammour models can clearly be seen.

3.3 A Dynamic Model of Transition with an Informal Sector

In this section we present an OST-type dynamic model of labour force reallocation in transition in the presence of an underground sector.

The model suggests that depending on different combinations of policy parameters as well as other factors, conducive to more or less effective job creation in the new sector relative to the informal sector, the economy ends up either in the steady state without the underground economy, or in the steady state where it is present. For a range of parameter values the informal sector completely crowds out the new formal sector in the steady state which, thus, effectively implies a complete failure of transformation. Numerical simulations of the model show that irrespective of the type of the steady state the drop in participation rates (observed in many transition economies over the 1990s) can be an inherent part of transition.

We describe the main assumptions and the idea of the model in the next subsection.

3.3.1 Main Assumptions

Consider an economy consisting of three sectors: two official sectors - old and new, and one informal sector.

By the old sector we understand all those state enterprises that are not efficient, possess large stocks of obsolete capital, and are expected to be liquidated during the transition.

Here we do not model the restructuring process of old enterprises:¹ we simply assume

¹Aghion and Blanchard (1994) consider a variant of their model where part of old enterprises is restructured rather than simply closed.

that the reforms initiated in the economy inevitably lead to the reduction in *old* sector employment, happening at some rate γ , a parameter which can in principle be affected by the government. Following Commander and Tolstopiatenko (1996) we assume that these firms do not invest. This could happen because either insiders have enough power to extract all surplus or some part of revenue is defalcated or wasted.² Old enterprises do not hire in our economy, they just shed labour.³ Workers in the old enterprises earn some fixed wage w_o equal in general to their product in the sector.

By the new sector we understand *de novo* (newly established) private enterprises and successfully restructured former state enterprises that have productivity $y > w_o$ and offer a flexible wage. These firms invest their profits causing the economy to grow. We abstract from accumulation of physical or human capital in this model. All investments in fact go into new job creation. Jobs cannot be shared and each is designated for one employee only.

The informal sector is understood as a whole range of activities including subsistence activities, home production (e.g. work on worker's own land plot), informal entrepreneurship, and unregistered activities of formal firms. We will think of firms in this sector as the entities essentially similar to their new sector counterparts, i.e. having the same productivity level, y , setting a flexible wage according to the same general bargaining

²Boeri and Terrell (2002) note that in some CEEC (e.g., Poland, Hungary and Slovenia) and in Russia the so-called worker councils or collectives had had some control over the appointment of managers, wage setting and the allocation of profits, which generally went to workers.

³Layard and Richter (1995) note that, for example, in Russia most of the unemployed are eventually rehired in old state enterprises rather than in the new private sector. However, according to Clarke (1999) this happens mainly due to churning (i.e. the worker movement between existing workplaces which is left out of our model), but not due to creation of new jobs. Clearly, in the exacerbating conditions of economic slump creation of new job places requires at least some degree of restructuring, which within the limit of our model would imply that such enterprises already belong to the new sector. Generally speaking, what is essential for our analysis it is not the fact of hirings (or their absence) in the old sector, but rather that the size of the old sector itself shrinks in the course of transition, thus providing labour resources for the new sector and informal economy.

rules, and investing their profits in job creation. The difference between the two sectors is in the government's ability to levy taxes on firms: while the new sector is subject to taxation, informal firms evade it. Also, as we shall see later, the ability of the government to monitor the informal economy implies a higher death rate of informal firms as opposed to new firms, and, as a result, a higher turnover of labour force in the underground sector.

We assume that both new sector and informal firms have no problem with hiring a worker, i.e. all the vacancies they create can immediately be closed by hiring a job-seeker. Blanchard (1998) calls such a situation "no matching problem for firms". It can be thought of as a great number of job-seekers waiting at one giant factory gate: in order to fill a vacancy firms just need to open the gate - a reasonable assumption for the transition from plan to market accompanied by a high level of unemployment. Nevertheless, this assumption does not imply that firms instantly replace matches that have broken up. Match break-up in our model implies that a job simply ceases to exist, but does not turn into a vacancy. At the same time, creation of new vacancies either to expand business or to replace dissolved matches costs firms money so they need to invest profits into it. However, they do not have to wait for a suitable candidate to fill a new position.

Workers in our economy can generally find themselves in two possible states: productive or unproductive. If workers are employed - they generate some output, whether it is in the old, new sector, or in the informal economy, but do not search for jobs. If workers are unemployed then they are "random job-seekers", i.e. they do not generate output, but search for a productive match without directing their search to a specific sector of the economy. We assume that no moonlighting (i.e. simultaneous job holding in two sectors at a time) and on-the-job search are possible. Workers involved in informal businesses drop

out of participation in the formal part of the economy and can be considered inactive.

The transition is viewed as the reallocation of labour from old sector enterprises to more efficient firms in the new sector or/and the informal economy. At the beginning of transition the economy is dominated by old sector firms. Ultimately, the basic mechanism implied by the model is as follows.

The old sector is shrinking because of job break-ups. The new sector and the informal sector both create new vacancies and start growing. Thus, workers getting unemployed in the old sector can find a job either in the new sector or in the informal economy. In the latter case they drop out of participation. While being employed in the new or informal sectors workers can be fired and become unemployed. If a worker loses her job in the informal sector, joins the ranks of the unemployed and starts searching for a new job, she effectively comes back into participation (the reallocation flows⁴ and transition probabilities are shown in Fig.3.3).

The flows of workers across our economy depend on rates of job destruction in each

⁴Note that in our model there are no direct shifts between employment in one sector and employment in another. Effectively, labour reallocation takes place through the unemployment pool. This feature relevant also to many other OST studies, is countered by Brixiova (1997) and Boeri (2000). They argue that the reallocation of labour in the transition between the old and new sectors has not only been through the unemployment pool but also often through a mechanism of direct job-to-job shifts. We have also made the same argument in Chapter 2. Still, the simplistic vision of reallocation in our model can be justified on several grounds. First and foremost, the unemployment pool here consists of *unproductive* workers. That is, it can include both unemployed workers and the so-called concealed unemployed. The latter group may consist of workers in the old enterprises compulsorily sent on an "administrative leave" (as would all too often happen in, e.g., Russia in the mid 1990s), employees put on short-time due to the absence of work for them, workers with partial payment or even without pay (but still preserving access to some non-wage benefits), etc. - see, for example, Clarke (1999). Such workers can and do in fact often spend their additional free time looking for other jobs in formal or informal sectors. When they find a job and are hired a direct job-to-job shift will be registered in reality. Within the limits of our model this is captured by the transition from unemployment to employment. Secondly, and importantly for the purpose of our study, we have argued in the background section that while the inclusion of job-to-job transitions provides a useful framework for the analysis of certain labour market policies, it still does not change the qualitative dynamics of the system. As we do not pursue the goal of taking into account as many labour flows as possible but rather the one of analysing in general the dynamics and the eventual outcome of transformation of an economy with the informal sector, the assumption of hirings solely from unemployment holds out well. Finally, as has also been mentioned above, flows into inactivity (which *inter alia* include outflows into the informal economy) take place mainly through unemployment.

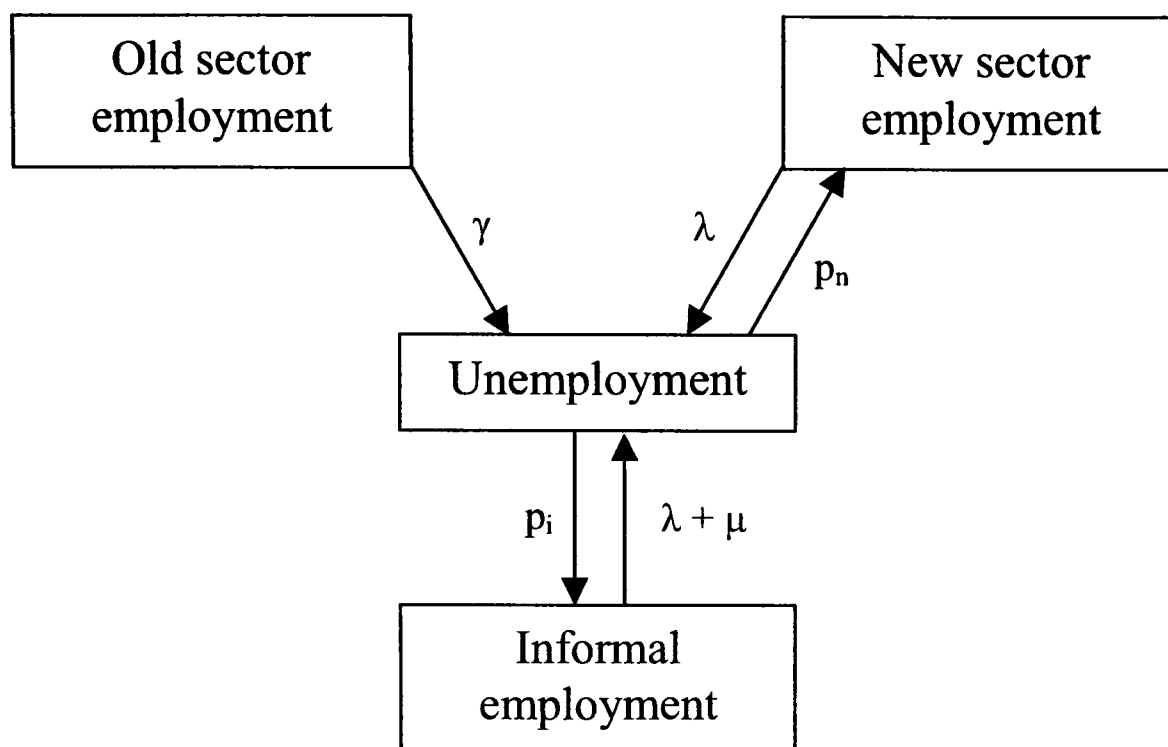


Figure 3.3: Flows of labour in the model

sector and rates of job creation in the new and informal sectors. The rates of job creation depend on profits and, therefore, wages. While the level of wages in the old sector is given, firms both in the new and the informal sectors negotiate wages with workers through a process of Nash bargaining. During this process wages are set so as to make employment in the new or informal sectors sufficiently attractive to unemployed workers. The attractiveness of employment comes from rents that employed workers appropriate either as in efficiency wage models or because of inefficiencies in the spirit of Caballero and Hammour (1996). Thus, the exit rate from unemployment depends on new and informal sector wages. They drive the dynamic process which describes how the economy evolves over time.

Now we can get round to considering the model in more detail.

3.3.2 Workers

It is assumed that the overall number of workers (working age population) available in the economy as a whole is constant and normalised to 1. This implies that the size of the labour force (labour supply in the formal economy) may vary depending on the size of the informal sector.

Workers can either be employed in only one of the three sectors at a time or search for jobs while unemployed (i.e. they can be either in the productive or nonproductive states as in, e.g. Atkeson and Kehoe, 1996). It is assumed that when employed they never leave employment voluntarily to start looking for another job. In the case of old sector employment this can be justified by workers valuing employment more than unemployment because of monetary income, w_o , as well as some additional characteristics (e.g. nonpecuniary benefits that are not accessible to the jobless). This, however, is not specified explicitly in the analysis that follows. At the same time, we shall see later that predilection for employment over unemployment in emerging sectors is guaranteed by inefficiencies in the job creation process that feeds through into wage setting.

Workers earn a fixed unemployment compensation,⁵ b_u , if they are jobless, a fixed wage, w_o , if they are employed in the old sector, and flexible wages w_n or w_i if they are in new or informal sector employment, respectively. We make the process of wage determination more explicit in Section 3.3.7.

⁵Although unemployment in this work can be interpreted as corresponding to the ILO definition, for simplicity we assume that all the unemployed receive unemployment compensation from the government. In general, what is important for our further analysis is not the existence of unemployment benefits as such but the ability of the government to affect the utility of unemployed workers.

3.3.3 The Old Sector

The wage rate in the old sector, w_o , is fixed and assumed exogenously given. We postulate that old firms do not invest and do not hire, they just shed labour at exogenously given rate γ . Also old sector insiders are presumed to have enough power to extract all the rents so that, by and large, w_o represents the worker's product in the sector.

3.3.4 The New Sector

Wages in the new sector are set flexibly at any level below the worker's product in the sector, y . This product is greater than its counterpart in the old sector, w_o , reflecting higher productivity of new firms compared to old enterprises. New sector profits are taxed at some rate τ by the government and the remaining proceeds are invested into job creation. In Section 3.3.7 we discuss wage determination and investment decisions in this sector in more detail. Finally, new sector firms are assumed to have no problem filling a *new* vacancy, i.e. there is no matching problem for the firms (Blanchard, 1998). In other words, each period no created vacancy remains idle. At the same time, filled jobs that break up do not turn into vacancies but simply cease to exist.

The Bellman equation for a worker in the new sector can be written as:

$$rV_n = w_n + \lambda(V_u - V_n) + \frac{\partial V_n}{\partial t}, \quad (3.1)$$

where r is the interest rate, λ is an exogenous probability⁶ of job loss in the sector, V_n

⁶Hereafter we deal only with flow probabilities as customary in the theory of matching functions. Such probabilities can take any value between zero and infinity, as according to a Poisson process they represent, for example, an average number of workers hired or laid off depending on the context. These probabilities do not have to be below one in the general case (for more details see, e.g., Saint-Paul, 1996; Pissarides, 2000). However, under the assumptions made in our model (labour supply normalised to 1, no matching problem for firms, and one job is one employee) it is not difficult to see that all the transition probabilities must be below 1.

and V_u are the worker values of being employed in the new sector and being unemployed, respectively. The arbitrage equation states that the return on being employed in the new sector is equal to the utility drawn from wage w_n in the new sector less the expected loss from losing the job plus the change in the value of being employed in the new sector over time.

We assume that λ in the equation above represents some shock leading to job break-ups. The magnitude of this shock may be greater than the rate of job destruction in the old sector, γ , if government accommodates part of the effect of the shock (for example, by subsidising the old sector) on old sector enterprises. Alternatively, it may be lower than γ if government speeds up the restructuring of old enterprises and goes on with reforms more vigorously.

Workers do not leave the new sector to become employed elsewhere because by assumption they can either produce or search for new jobs, i.e. perform only one activity at a time.

3.3.5 The Informal Sector

Characteristics of the informal sector are very similar to those of the new sector. It is assumed that worker's productivity, y , is equal to her productivity in the new sector,⁷ firms do not face matching problems, set flexible wages, and invest their profits (see Section 3.3.7 for more detail). These assumptions are justified by the fact that in reality it is often the same new sector firms that run part of their businesses underground by hiring informally to escape taxes and high social security contributions (De Soto, 1989; Boeri and Garibaldi,

⁷We shall see later that this is just a simplifying assumption which does not affect qualitatively the conclusions of our analysis.

2001). Also the rich literature on the informal economy (for some discussion see Johnson *et al.*, 2000a; Schneider and Enste, 2000) observes that tight regulation of the formal sectors, exposure to corruption and high barriers to entry force many entrepreneurs, who would otherwise have run their enterprises formally, start business in the shadow.

One of the differences between informal and new sector firms in our model is that profits of the former are not taxed by the government as taxes are effectively evaded in the informal sector. Another difference is that job destruction rate, and hence, labour force turnover are higher in the informal economy (for some corroborating evidence from developing countries see, e.g. Hoek, 2002). This can be due to monitoring or auditing activities of such bodies as the tax police. Alternatively, there may exist some sector specific risks (e.g. exposure to criminal racket) leading to more job destruction in the absence of formal job security regulations.

Thus, the asset value of informal employment for a worker is given by

$$rV_i = w_i + (\lambda + \mu)(V_u - V_i) + \frac{\partial V_i}{\partial t}, \quad (3.2)$$

where w_i is the informal wage; λ is, as before, an economy-wide shock leading to job loss, while μ is an additional idiosyncratic (sector specific) shock. Henceforth, we will assume that μ reflects the effect of monitoring activities carried out by the government which, thus, makes μ a possible policy parameter.

The arbitrage equation (3.2) states that the return on being in the informal sector is equal to the rate of informal income plus a change in the value of informal employment over time.

Again, direct shifts between informal and new sector employment are not considered as according to the assumption full-time engagement in these firms does not leave workers enough time to perform job search.

3.3.6 Unemployment

Workers are unemployed when they are not working in any of the three sectors. While unemployed a worker faces the following possibilities: she can find a job either in the new or the informal sector, or she can remain jobless.

Let us assume that unemployed workers perform undirected and random search for jobs, i.e. the probability of moving into the new sector is independent of the probability of moving in the informal sector. Then the value of being unemployed is

$$rV_u = b_u + p_n (V_n - V_u) + p_i (V_i - V_u) + \frac{\partial V_u}{\partial t}, \quad (3.3)$$

where b_u is unemployment income (which *inter alia* depends on unemployment compensation); p_n is a probability of moving into the new sector; p_i is a probability of moving into informal jobs. Thus, the return on being unemployed includes utility from unemployment benefits plus the expected gain from changing into new sector or informal employment plus the change in the value of being unemployed over time.

3.3.7 Wages and Job Creation in the New and Informal Sectors

Collecting together all the arbitrage equations we get the following system of Bellman equations:

$$\begin{aligned}
rV_n &= w_n + \lambda (V_u - V_n) + \frac{\partial V_n}{\partial t}, \\
rV_i &= w_i + (\lambda + \mu) (V_u - V_i) + \frac{\partial V_i}{\partial t}, \\
rV_u &= b_u + p_n (V_n - V_u) + p_i (V_i - V_u) + \frac{\partial V_u}{\partial t}.
\end{aligned} \tag{3.4}$$

Wage determination

In our economy wages w_n and w_i are continuously set so as to make employment in the two emerging sectors more attractive than unemployment. This is expressed by the two wage mark-up conditions:

$$V_n - V_u = c, \tag{3.5}$$

$$V_i - V_u = m, \tag{3.6}$$

for the new and informal sector, respectively, where c and m are some constants.

One can think of a number of reasons underpinning conditions (3.5) and (3.6).

Firstly, even if firms do not face matching problems and have all the bargaining power to hold down workers to their reservation value (i.e. V_u), they still may have to pay mark-ups over the value of unemployment to ensure, for example, that workers do not shirk. Aghion and Blanchard (1994) and Blanchard (1998) demonstrate that conditions (3.5) and (3.6) summarise the efficiency wage considerations such as those in Shapiro and Stiglitz (1984).

Secondly, the value of mark-ups, c and m , may reflect differences in preferences of workers over formal and informal work. For example, c may be viewed as disutility of working hard in the new sector, while m as disutility of working in the informal sector. This latter disutility can be derived from the value of formal pensions that informal workers forego, or from the feeling of guilt informal workers undergo while in the shadow sector.

Thirdly, Kehoe (1994) suggests that mark-ups similar to the ones above arise from introducing moral hazard into the job search. He illustrates that by supposing that workers can put an unobserved effort into searching which allows them to find matches with positive probability as opposed to other workers who do not make any effort and do not find matches. Putting effort into search is costly and workers will do so if and only if the future flow of consumption compensates for their actions. Thus, there must be a gap between the expected utility of workers who find new matches and those who do not.

Fourthly, in the spirit of Caballero and Hammour (1996) the mark-ups can reflect a different degree of appropriability of quasi-rents in the new and informal sectors. In particular, appropriability or hold-up (in the terminology of Williamson, 1985) of investment is a problem of incomplete contracts that can afflict the efficiency of job-creating transaction between a firm and a worker. When part of the investment is "specific" to the firm and a binding and complete contract cannot be written, the firm fails to hold sway over the quasi-rents attributable to the investment. Then the ability of a worker to renegotiate and hold up the investment insures that the value of employment within a sector is greater than her unemployment utility.

Finally and directly related to the previous point, the values of the two mark-ups may be affected by the existence of separation costs, redundancy pay, or relative bargaining

power of workers and firms (Grout, 1984, and Malcomson, 1997, provide good insights). Again it is often the ability of workers to exert enough bargaining power and renegotiate the contract that generates conditions (3.5) and (3.6). We illustrate this in more detail in Appendix 3.A.

The mark-up conditions imply that $\frac{\partial V_n}{\partial t} = \frac{\partial V_i}{\partial t} = \frac{\partial V_u}{\partial t}$. Then by subtracting the third equation in (3.4) from the first and the second and using the mark-up conditions we obtain expressions for wages in the new and informal sectors, respectively:

$$w_n = b_u + (r + \lambda) c + p_n c + p_i m, \quad (3.7)$$

$$w_i = b_u + (r + \lambda + \mu) m + p_n c + p_i m. \quad (3.8)$$

Thus, the wages of a worker in the new and informal sectors depend on the level of unemployment income (which effectively provides a wage floor in our economy), job break-up rates, wage mark-ups, and probabilities of finding a job in both sectors considered. The wages are not constants, but rather are functions of the market tightness: job finding probabilities p_n and p_i are expressed as ratios of job creation rates in a particular sector to the pool of the unemployed.

The wage differential⁸ between the new and informal sectors is given by

$$w_n - w_i = (r + \lambda) c - (r + \lambda + \mu) m. \quad (3.9)$$

⁸If we explain the mark-ups c and m by hold-ups of firm-specific investment the resulting wage differential echoes the Caballero and Hammour (1996), who write: "When the degree of asset specificity and appropriable quasi-rents differ across sectors - as is normally the case, for example, between the formal and informal sectors - it is very natural that wages also differ in equilibrium. Appropriability can thus very naturally account for labour market segmentation, which is preponderant feature of labour markets in the developing world."

Job creation

As we have gathered from the background section, the job creation process is the crucial part and, at the same time, a weak spot of the OST literature that bears upon many of its implications. Here we follow the approach of Aghion and Blanchard (1994), assuming that firms in the new sector and the informal economy invest out of retained profits.^{9,10} In the simple version of the model borrowing is not possible due to underdeveloped financial markets (in Section 3.3.11 we relax this assumption). The flow of investment transfers into the job creation in the absence of capital accumulation which is left out of the scope of this work.¹¹

In the most general specification we define the amount per worker of investment into new job creation by

$$J_n = \alpha(\tau, \zeta)(y - w_n) \quad (3.10)$$

and

$$J_i = \beta(\mu, \xi)(y - w_i), \quad (3.11)$$

⁹The results of the survey by Bratkowski *et al.* (2000) show that imperfections in capital markets in Central European economies do not actually seem to inhibit the growth of new private firms. Johnson *et al.* (2000b) argue that the reason for that is the existence of an alternative to external finance, namely reinvestment of enterprise's own profits. Lizal and Svejnar (2000) find that in the Czech Republic retained profit is a major determinant of new investment. Their results indicate that Czech firms cannot easily borrow investment funds externally and that net investment varies with retained profits. Pissarides *et al.* (2003) provide evidence that Bulgarian and Russian small and medium firms use internal finance to fund investment projects, but that constraints on external financing limit in important ways their ability to expand production.

¹⁰Another explanation for this specification of job creation can be costs of adjustment (learning-by-doing, accumulation of information, etc.).

¹¹Capital accumulation issues are taken up in Castanheira and Roland (2000) and Castanheira (2003).

depending on whether she belongs to the new or informal sector, respectively.¹² Equations (3.10) and (3.11) define job creation rates per single worker employed. If N and I are the numbers of workers in the two sectors then total rates of job creation in the new formal and informal parts of the economy will correspondingly be given by $J_n N$ and $J_i I$. Effectively, they are proportional to profits.

In the two equations above y is the product/output per worker in the sectors; w_n is the wage in the new sector; w_i is the informal wage; $\alpha(\cdot)$ and $\beta(\cdot)$ are two functions that capture the effect on job creation/investment of various institutional factors and government policies towards either the new or the informal sector. We assume that both $\alpha(\cdot)$ and $\beta(\cdot)$ can take on values in between 0 and 1.

In the case of the new sector $\alpha(\tau, \zeta)$ may depend on the tax rate levied on new firms, τ , as well as less tangible characteristics such as a level of corruption that makes life of formal firms more difficult. We use a catch-all variable ζ to denote such factors. Ruggerone (1996) calls α the reactivity of new job creation to profitability. Blanchard (1997) notes that α is in fact very important for the development of transition: the more various constraints and adjustment costs facing the new sector the lower is α . The logic here is as follows. If an economic environment is favourable then the probable value of α is close to 1, which implies that firms can use all their profits (the output less the wage rate in our case) for investment, i.e. new job creation. Otherwise, if the economy is highly regulated or even corrupt the profits of new firms may be taken away through excessive regulation, taxation, corruption (e.g. bribing), etc. This has been the case of a number of transition countries, and Russia in particular (see Johnson *et al.*, 1997). In such a situation the value of α is

¹²Equations (3.10) and (3.11) imply, in fact, that firms are myopic in their decision to create jobs. This, however, does not conflict with the evidence from transitional countries - see, e.g., Berkowitz and DeJong (1997).

much lower than 1, and opportunities for job creation are scanty.

We assume that $\frac{\partial \alpha(\tau, \zeta)}{\partial \tau} < 0$ and $\frac{\partial \alpha(\tau, \zeta)}{\partial \zeta} < 0$, where ζ captures other (apart from taxes) factors "unfriendly" to new sector job creation.

In a similar vein, $\beta(\cdot)$ is the factor that defines the reactivity of informal job creation to profitability. It can depend on the level of monitoring of informal activities by the tax police, μ , the level of fines charged for running informal businesses, and other factors (e.g. such as the existence of criminal rackets), represented by another catch-all variable ξ . Again we assume that $\frac{\partial \beta(\mu, \xi)}{\partial \mu} < 0$, and $\frac{\partial \beta(\mu, \xi)}{\partial \xi} < 0$.

Importantly, in our model the definition of factors $\alpha(\cdot)$ and $\beta(\cdot)$ implies that the government can directly affect their level, thus impacting on the development of a particular sector. In general, at least one of the factors $\alpha(\cdot)$ and $\beta(\cdot)$ can be made endogenous to the system, or they can be linked together. We do not develop these opportunities in this work, rather assuming that $\alpha(\cdot)$ and $\beta(\cdot)$ depend on exogenous parameters, but we touch on possible departures in Section 3.3.11.

Open form solutions

The rates of job creation in the new and informal firms help define the transition probabilities p_n and p_i from unemployment to new sector or informal employment.

So, if J_n is the flow of created vacancies per worker in the new sector then the rate of exit from unemployment to new sector jobs can be written as $p_n = \frac{J_n N}{U}$, where N is the number of new sector workers and U is the number of unemployed.

By the same token, transition probability p_i is equal to $\frac{J_i I}{U}$, where I is the stock of informal workers.¹³

¹³The way that we have written the transition probabilities with job creation rates in the numerator reflects our assumption of no matching problem for firms. In a more general case, if firms face a problem

Thus, on the one hand, wages in the new (3.7) and informal (3.8) sectors depend on transition probabilities p_n and p_i , which in turn depend on investment rates J_n and J_i . On the other hand, the investment rates in the new (3.10) and informal (3.11) sectors depend on wages w_n and w_i . The system of these four equations (3.7), (3.8), (3.10), and (3.11) can readily be solved to obtain "open form" expressions for wages and job creation rates:

$$w_n = \left(y - \frac{US_n - \beta m I (S_i - S_n)}{(U + \alpha c N + \beta m I)} \right), \quad (3.12)$$

$$w_i = \left(y - \frac{US_i - \alpha c N (S_n - S_i)}{(U + \alpha c N + \beta m I)} \right), \quad (3.13)$$

$$J_n = \alpha \left(\frac{US_n - \beta m I (S_i - S_n)}{(U + \alpha c N + \beta m I)} \right), \quad (3.14)$$

$$J_i = \beta \left(\frac{US_i - \alpha c N (S_n - S_i)}{(U + \alpha c N + \beta m I)} \right), \quad (3.15)$$

where $S_n = (y - b_u - (r + \lambda) c)$, $S_i = (y - b_u - (r + \lambda + \mu) m)$, $\alpha \equiv \alpha(\cdot)$ and $\beta \equiv \beta(\cdot)$ are introduced for notational convenience.

It can be shown (henceforth see Appendix 3.B for all technical details) that wages w_n and w_i are decreasing in the level of unemployment, U , but increasing in the size of employment in the new sector, N , and the informal economy, I . At the same time both J_n and J_i are increasing in the level of unemployment, U , and decreasing in both N and I .

filling vacancies, the flow of hirings does not coincide with the number of vacancies posted. So, in that situation a matching function (of the number of vacancies and job seekers) will be put in the numerator instead of the job creation rates.

From (3.12)-(3.15) it follows that without further assumptions nothing can prevent wages from rising above y , or, equivalently job creation rates from becoming negative. Following Aghion and Blanchard (1994) we shall assume that firms in the emerging sectors immediately close if they are losing money, i.e. they never run negative profits. As J_n and J_i are proportional to profits, the restriction just introduced implies that J_n and J_i can only be non-negative, while firms pay their workers wages below y .

3.3.8 Dynamics

Above we considered Bellman equations for individual workers and defined rules for wage determination and job creation in the new and informal sectors. Now we can derive differential equations describing development of each sector in our model. In what follows we define by O the number of workers in the old sector, by N - the number of workers in the new sector, by I - the number of informal workers, and by U - the number of the unemployed.

At the beginning of transition the bulk of the labour force belongs to the old sector. Then this employment begins to decrease at rate γ :

$$\frac{dO}{dt} = -\gamma O.^{14} \quad (3.16)$$

The equation (3.16) shows that on average γO old sector workers become unemployed each period.

¹⁴In their modelling of the dynamics Aghion and Blanchard (1994) use "change in variables" relations (i.e., e.g., $\frac{\partial O}{\partial t} = \text{const}$) rather than "rate of change" ones (i.e., e.g., $\frac{\partial O}{\partial t}/O = \text{const}$), although the latter formulation is more realistic given the interpretation of job creation through retained earnings constraints. They explain their choice by the need of better tractability of the model. However, this in fact shapes the essence of the steady states in their model, leads to discontinuities in dynamics, and makes steady state transition an inherent feature of their economy. Here we use the "rate of change" specification which generates smooth dynamics throughout. Then, transition is a process that occupies our economy before it reaches a steady state.

At the same time, in the new sector the dynamics of the total number of workers is described by the equation:

$$\frac{dN}{dt} = p_n U - \lambda N. \quad (3.17)$$

This equation reflects the fact that each period $p_n U$ workers are hired in the new sector from the unemployed ranks, while λN new sector workers lose their jobs as a result of economy-wide shock λ .

Similarly, the informal employment follows:

$$\frac{dI}{dt} = p_i U - (\lambda + \mu) I. \quad (3.18)$$

Again, as in the case of the new sector a few workers are hired from the unemployment pool, $p_i U$, while $(\lambda + \mu) I$ informal workers become jobless due to economy-wide shock λ and sector specific shock μ .

Finally, the flows in and out of unemployment determine its dynamics as

$$\frac{dU}{dt} = \gamma O + \lambda N + (\lambda + \mu) I - (p_n + p_i) U, \quad (3.19)$$

where the first two terms on the right hand side are inflows from the formal sectors, the third term describes inflows from the informal economy and the fourth term represents outflows to formal and informal employment.

Combining all the dynamic equations and substituting for $p_n = \frac{J_n N}{U}$ and $p_i = \frac{J_i I}{U}$ we obtain the following nonlinear system describing behaviour of the economy:

$$\begin{aligned}
\frac{dO}{dt} &= -\gamma O, \\
\frac{dN}{dt} &= (J_n - \lambda) N, \\
\frac{dI}{dt} &= (J_i - (\lambda + \mu)) I, \\
\frac{dU}{dt} &= \gamma O + (\lambda - J_n) N + ((\lambda + \mu) - J_i) I,
\end{aligned} \tag{3.20}$$

where w_n , w_i , J_n , and J_i , are determined by (3.12), (3.13), (3.14), (3.15), respectively.

Equations (3.20) satisfy the consistency condition $\dot{O} + \dot{N} + \dot{I} + \dot{U} = 0$ since $O + N + I + U = 1$. In general, as functions on the right hand side of (3.20) are single-valued, continuous, and continuously differentiable in O , N , U , I , on the domain defined by $O \in (0, 1]$, $N \in (0, 1]$, $U \in (0, 1]$, and $I \in (0, 1]$, and in the absence of restrictions on J_n , and J_i , it can be proved (see, e.g. Kamien and Schwartz, 2001, p.351) that this system has a unique set of continuous solutions for any set of parameters. However, by assumption rates J_n , and J_i can only be non-negative, which implies that certain restrictions apply to the functions on the right hand side of (3.20). Those restrictions can make the functions not continuously differentiable. In order to guarantee the existence of a unique set of continuous solutions we, thus, need to restrict some parameters of our model. In particular, given that O , N , U , I are considered on the domain $(0, 1] \times (0, 1] \times (0, 1] \times (0, 1]$, and that $\alpha(\cdot) \in (0, 1]$ and $\beta(\cdot) \in (0, 1]$, from (3.14) and (3.15) it follows that if either $\frac{(r+\lambda+\mu)}{(r+\lambda)}m < c < \frac{y-b_u}{(r+\lambda)}$ or $c < \frac{(r+\lambda+\mu)}{(r+\lambda)}m < \frac{y-b_u}{(r+\lambda)}$ then J_n and J_i are always positive. Then the existence and uniqueness of solution is guaranteed for any γ . We shall assume in the sequel that c and m satisfy one of the restrictions.

Steady states

The system (3.20) is in the steady state equilibrium when $\dot{O} = \dot{N} = \dot{I} = \dot{U} = 0$ in addition to satisfying (3.12), (3.13), (3.14) and (3.15). There are one trivial steady state ($N^* = 0, O^* = 0, U^* = 1, I^* = 0$) and three types of non-trivial one.

Initial conditions: whence the transition? Given the way that we defined job creation in the new and informal sectors (i.e. as functions of profits per worker), if the economy embarks on transition having all the labour force concentrated in the old sector, the only possible outcome is the trivial steady state: the new and informal sector simply cannot take off because there are no funds to invest into job creation. Also, as expressions in (3.14) and (3.15) indicate, one of the job creation rates J_n and J_i is necessarily negative if unemployment is nil at the start of transition. So, to ensure that dynamic system (3.20) converges to one of non-trivial steady states analysed below we need to assume that at the beginning of transition the labour force is distributed between all four labour market states: old, new and informal sector employment and unemployment. This assumption would not run counter to the existing evidence from transition countries (see, e.g. EBRD, 1999, for the share of the private sector in GDP at the beginning of transition; Feige and Urban, 2003, provide one of the most recent estimates of unrecorded economic activity in 1989-2001; Gregory and Collier, 1988, report on unemployment statistics for the Soviet Union¹⁵).

Thus, although we shall not put the condition $O(0) = 1$, and $N(0) = U(0) = I(0) = 0$ as the one preceding the transition, as was customary in many previous OST models, we

¹⁵In particular, Gregory and Collier (1988) report that, although the Soviet government claimed to have "liquidated" the unemployment in the early 1930s, many Western observers have agreed that the Soviet labour market shared many of the characteristics of its Western counterparts, including unemployment.

still shall assume that the vast majority of the labour force is affiliated with the old sector, and some small fractions of workforce are involved in the new sector and the informal economy, and are unemployed. This will guarantee that the economy escapes the trivial steady state and begins its journey to a non-trivial one.

Now we can move on to analysing the types of non-trivial steady state.

A steady state with the new sector alone In the steady state of the first type there exists only the new sector. From dynamic system (3.20) it immediately follows that this takes place whenever $O^* = 0$, $J_n = \lambda$ and $J_i < \lambda + \mu$, while $I^* = 0$. That is, in the steady state the job creation rate in the new sector is equal to the job destruction rate, while in the informal sector the job creation rate is less than the job destruction rate, $\lambda + \mu$. In such a situation the informal sector is not present in the steady state, neither is the old sector.

It is straightforward to show that in this steady state the number of new sector workers is equal to

$$N^* = \frac{(\alpha S_n - \lambda)}{(\alpha S_n - \lambda + \lambda \alpha c)}, \quad (3.21)$$

while unemployment is

$$U^* = \frac{\lambda \alpha c}{(\alpha S_n - \lambda + \lambda \alpha c)}. \quad (3.22)$$

This steady state occurs whenever the parameters of the model satisfy the condition

$$\alpha > \frac{\lambda\beta}{(\lambda + \mu) + \beta(S_n - S_i)}. \quad (3.23)$$

It can readily be shown that the steady state unemployment level U^* decreases in $\alpha(\cdot)$ and new sector surplus S_n , and increases in new sector mark-up c , and job destruction rate λ ; while the level of new sector employment, N^* , increases in $\alpha(\cdot)$ and S_n , and decreases in c and λ accordingly.

A steady state with the informal sector alone By analogy to the previous case, there exists a steady state with the informal sector alone, where $O^* = 0$, $J_i = \lambda + \mu$ and $J_n < \lambda$, while $N^* = 0$. In this situation the new sector is not present in the steady state because job destruction rate λ exceeds the job creation rate in the sector. At the same time flows into and out of the informal employment are balanced.

The stocks of informal workers and unemployed in this steady state are, respectively, given by

$$I^* = \frac{(\beta S_i - (\lambda + \mu))}{(\beta S_i - (\lambda + \mu) + (\lambda + \mu)\beta m)}, \quad (3.24)$$

and

$$U^* = \frac{(\lambda + \mu)\beta m}{(\beta S_i - (\lambda + \mu) + (\lambda + \mu)\beta m)}. \quad (3.25)$$

The steady state turns out whenever

$$\alpha < \frac{\lambda\beta}{(\lambda + \mu) + \beta(S_n - S_i)}. \quad (3.26)$$

In this steady state the level of unemployment, U^* , decreases in $\beta(\cdot)$ and S_i , and increases in informal mark-up m and job destruction rates λ and μ ; while the level of employment in the informal economy, I^* , increases in $\beta(\cdot)$ and S_i , and decreases in m , λ and μ , accordingly.

A steady state with both sectors present Finally, the third type of steady state has both sectors present. This happens whenever $O^* = 0$, $J_n = \lambda$ and $J_i = \lambda + \mu$. Flows both into and out of the new and informal sectors are balanced given their sizes and the level of unemployment.

As shown in Appendix 3.B, conditions $J_n = \lambda$ and $J_i = \lambda + \mu$ are in fact identical. This takes place whenever the following equality holds:

$$\alpha = \frac{\lambda\beta}{(\lambda + \mu) + \beta(S_n - S_i)}. \quad (3.27)$$

Changes in parameters and their effects on steady states Knife-edge condition (3.27) defines a whole range of various combinations of values of r , c , m , λ , μ , $\alpha(\cdot)$, and $\beta(\cdot)$ that produce a steady state with various shares of both new and informal sectors (proved in Appendix 3.B). However, in reality it is highly unlikely that the effects of a number of parameters on the development of the new and informal sectors are balanced to generate the equilibrium with both sectors. Thus, in what follows we do not focus on this type of steady state.

However, it is still interesting to examine how changes in the parameters of the model impact on condition (3.27), shifting the balance of forces in the economy towards one or another sector, so that the steady states with the new or informal sectors result.

Reactivity coefficients $\alpha(\cdot)$ and $\beta(\cdot)$ Firstly, from (3.27) it is straightforward to see that *ceteris paribus* an increase in $\alpha(\cdot)$ (or, equivalently, a decrease in $\beta(\cdot)$) would upset the balance, so that conditions to develop successfully become more favourable to the new sector. Then it prevails in the end of transition, as described in Section 3.3.8 above.

Sector mark-ups Secondly, without making an additional assumption regarding the relative value of mark-ups c and m , it can readily be seen that the right hand side of (3.27) is decreasing in m and increasing in c . If we assume that initially condition (3.27) obtains, an increase in m , for example, would imply a higher burden on informal firms, so that the economy would eventually converge to the steady state where the informal sector is not present. Similarly, the steady state with the informal economy results if c rises. Thus, the higher is the mark-up that new/informal firms have to pay its workers over their reservation value, the more likely for the economy to end up in the steady state with informal/new firms only.

Monitoring of the informal sector Similarly to the effect of an increase in m , an increase in μ leads to a decrease in the right hand side of (3.27), so that the economy is likely to end up in the steady state with the new sector alone.

Sector surpluses While monitoring has its both direct and indirect (i.e. through sector surpluses S_n and S_i) effects on the knife-edge condition, the mark-ups act solely through the surpluses. In general, the relative size of sector surpluses S_n and S_i affects the resulting steady state. As the difference $S_n - S_i$ becomes larger the economy gets more chances for the new sector to crown the end of transition. Thus, as we can see here,

our assumption about the equal productivity in the new and informal sectors, y , does not affect the qualitative outcome. It is not the productivity *per se* but the relative size of profit margins that makes all the difference.

An increase in general intensity of shocks An increase in the rate of job destruction equally affecting the new and informal sector, λ , has an ambiguous effect on the right hand side of (3.27). However, in Appendix 3.B we show that it is probable that such an increase would hurt new sector firms more than their informal counterparts, so that the economy is more likely to end up in the state where the informal sector prevails.

Unemployment benefits The higher is the level of unemployment benefits, b_u , the lower both the surplus S_n in the new sector and the surplus S_i in the informal sector. So there is no any effect of the unemployment compensation on the type of the resulting steady state. However, b_u still affects the level of unemployment (and hence, new and informal employment) in the equilibrium. This last result has a predicted effect common to the models of search in the labour market.

A rate of decrease in the size of the old sector Finally, condition (3.27) does not depend on γ , i.e. the rate at which the old sector declines in our model. Hence the choice of γ does not affect the steady state where economy converges in the end of transition.

Destination matters: whither the transition? Having established the characteristics of the steady states it is time to discuss some differences between our simple model and the preceding OST literature.

In the backbone model by Aghion and Blanchard (1994), as we have mentioned in the background section, the steady states, when they exist, differ only in the resulting level of unemployment and wages. Also, because of the fiscal externality effect on job creation there may exist both stable and unstable steady states which, however, are qualitatively identical. So, in that model the development of the new sector always crowns the transition. In more recent OST work the question of multiple steady states has hardly ever been discussed. However, the stability issue has manifested itself in the equilibrium path along which the transition has solely been considered. This path leading to the steady state is a unique opportunity, consistent with rational expectations of the agents, for the successful transition as modelled in those studies. Shimer (1997) or Boeri (2000b) may be referred to for the proof. Other possible trajectories are exploding and would derail transition, as is shown by Aghion and Blanchard (1994) for an extension of their model where firms are forward-looking.

In our model, formal and informal firms have so far been assumed to be not forward-looking (or capital markets not perfect), so that the economy always converges to a stationary point. Also, for a single set of parameters only one non-trivial steady state exists. The multiplicity does not arise because we do not introduce feedback mechanisms such as, for example, endogenous government budget constraints that create externality effects in such models as Aghion and Blanchard (1994), Boeri (2000b), etc. Such complications would make functions $\alpha(\cdot)$ or/and $\beta(\cdot)$ depend on the level of unemployment (or numbers of the employed in either sector), and thus make linear solutions to conditions $J_n = \lambda$ or/and $J_i = \lambda + \mu$ impossible. We reflect on the consequences of introducing the government budget constraint or other feedback mechanisms in our model in Section 3.3.11.

However, leaving aside for the time being the questions of equilibrium uniqueness/stability and formation of expectations, we would like to stress that the three types of steady state discussed in the previous section are qualitatively different. That is, different combinations of parameters define steady states either with or without the new or informal sector. So, the qualitative result of transition can be very different depending on the government policies even when the derailment of transformation (i.e. non-existence of steady states) is excluded as a possibility.

Reaching the apex: transition trajectories

Now that we know where our economy is likely to end up, we address two more questions, namely: 1) What does the development of the new and informal sectors look like before a steady state is reached? and 2) How does the presence of the informal sector affect the timing of transition? The insights are provided below.

Assume for convenience that the set of parameters is chosen so that the economy converges to the steady state where the new sector prevails in the end, i.e. condition (3.23) is satisfied. Other cases are easy to analyse by analogy.

The shape of trajectories From (3.20) it follows that the size of the old sector exponentially decreases till its complete disappearance in the steady state. At the same time, the shape of trajectories followed by the new and informal sectors depends, on the one hand, on the initial conditions and, on the other hand, on the relative effects on job creation rates (3.10) and (3.11) both of the growth of unemployment and of a change in size of the new and informal sectors.

The first consideration follows from the fact that initial conditions at the onset of

transition define the relation between job creation functions J_n and J_i and job destruction rates λ and $\lambda + \mu$. If, for example, at the beginning $J_n > \lambda$ and/or $J_i > \lambda + \mu$ either one or the other (or both) sectors can be increasing in size for a while, regardless of the type of the steady state that the economy eventually converges to.

The second remark implies that the job creation rates vary as unemployment grows and levels of new and informal employment change. The full time derivatives of the two job creation functions are given by

$$\frac{dJ_n(U, N, I)}{dt} = \frac{\partial J_n}{\partial U} \frac{dU}{dt} + \frac{\partial J_n}{\partial N} \frac{dN}{dt} + \frac{\partial J_n}{\partial I} \frac{dI}{dt} \quad (3.28)$$

and

$$\frac{dJ_i(U, N, I)}{dt} = \frac{\partial J_i}{\partial U} \frac{dU}{dt} + \frac{\partial J_i}{\partial N} \frac{dN}{dt} + \frac{\partial J_i}{\partial I} \frac{dI}{dt}. \quad (3.29)$$

Given the signs of the derivatives of J_n and J_i with respect to U , N , and, I , in both equations above the first effect is positive if $\frac{dU}{dt}$ is positive, while the second and the third effects are positive only if N or I are decreasing. Thus, in principle, if the economy starts with, say $J_n > \lambda$ and $J_i < \lambda + \mu$, i.e. the new sector grows, while the informal economy shrinks, the positive effect of an increase in unemployment coupled with a positive effect of a decrease in the informal sector (that drives wages both in the new and the informal sector down) can outweigh the negative effect of the growth in the size of the new sector. Then J_i begins to rise and as soon as it becomes greater than $\lambda + \mu$ the informal economy can be growing too. This emphasises two important aspects.

First, as unemployment is increasing in the rate of old sector closure, the rate of

reduction in old sector employment is a factor that has a bearing not only on the size of the new sector in transition as postulated by the previous OST literature, but also on the size of its informal counterpart. In particular, the higher that rate is, the more likely for the informal economy to grow in the process of reallocation of labour. Hence, it is also more likely for the participation in the formal economy to decrease.

Second, (3.28) and (3.29) indicate that it is the interplay between the new and informal sector development that shapes the transition paths, i.e. they are not the product of the development of unemployment alone. Appendix 3.B examines in more detail the behaviour of the job creation rates in the neighbourhood of no-growth levels λ and $\lambda + \mu$. In the end it turns out that a set of parameters defining the resulting steady state is equally responsible for the shape of trajectories.

Thus, to sum up our short discussion, the transitional development paths result from two factors. On the one hand, the initial conditions do or do not give the impetus to the development of both new and informal sectors. On the other hand, even if the initial conditions are not favourable it is the set of policy parameters defined by the knife-edge condition and the rate of reduction in old sector employment that finally configures the path of the economy.

The effect of the informal sector on the timing of transition Whatever the initial conditions in our economy transition always takes infinite time, i.e. this is how long it takes to reach any steady state in the model. However, even in this setting we can draw some interesting implications regarding the timing of transition. Suppose that the model parameters do not change so that steady state levels of new sector employment and unemployment are predefined. Then the timing of transition clearly depends on the

presence of the informal sector. In Appendix 3.D we prove that at any moment of time the economy is closer to the steady state if the informal sector is absent, i.e. if the economy starts with $I(0) = 0$. In other words, transition is faster if there is no shadow sector.

If the economy starts off with some positive level of informal sector employment then changes in parameters affecting its size, but not influencing the resulting steady state, would also have an unambiguous impact on the timing of transition. For example, an increase in the level of monitoring, μ , does not affect the steady state level of new sector employment, N^* , and unemployment, U^* . However, it is easy to verify that $\frac{\partial J_n(U, N, I)}{\partial \mu} \geq 0$, while $\frac{\partial J_i(U, N, I)}{\partial \mu} \leq 0$. Thus, by using an approach similar to the one employed to prove the negative effect of the presence of the informal sector, it is easy to show that a rise in μ speeds up the transition by eliminating the shadow economy faster.

All this clearly provides another illustration to the Caballero and Hammour (1996) argument that the existence of informality slows down the adjustment process. The reasons for that in our model are the same as the ones mentioned in the background Section 3.2.2.

Notes on dynamic efficiency

Thus far we have been silent on the welfare implications of parameter choices, pointing only to the possibility of two qualitatively different outcomes. However, changes in parameters also impact on the welfare of the economy through their effect on the job creation functions that shape employment trajectories. Different paths taken by old, new, and informal sector employment and unemployment affect the distribution of gains during transition. For example, the choice of a higher rate of closure of the old sector, γ , does not have any effect on the eventual outcome, but leads to a faster increase in unemployment. At the same time, as mentioned above, this may also lead to an increase in the informal

sector activity, especially at the beginning of transition when the negative externality from the new sector is small. Meanwhile, an increase in the rate of monitoring, μ , eliminates the shadow sector faster (if the economy converges to a steady state without informality). Either an increase in γ (or μ) may imply both gains and losses to the government. Thus, on the one hand, the higher is the value of γ the higher is the maximum level of unemployment which, apparently, creates higher pressures on the government budget. On the other hand, more unemployment implies lower wages, lower pressures on the new sector, and hence a higher level of job creation. There should exist a value of γ that balances these two effects and makes the reallocation of labour dynamically optimal. Burda (1993), Aghion and Blanchard (1994), Chadha and Coricelli (1994), Castanheira and Roland (2000) all study the effects on the economy of too slow or too high a speed of transition. In the most recent work on this topic Castanheira (2003) points to the differences in the optimal speed from the private and social viewpoints.

In our setting the issue of finding an optimal speed of scale-down of old enterprises presents a serious difficulty because even within this simple model with three sectors the algebra becomes really involved. However, on the basis of our results for the effect of the informal sector on the timing of transition we can argue that the optimal speed is not generally equal for systems with and without the informal sector. This implies that if the government tries to reform the economy without taking into account the presence of the shadow sector, it is likely to pursue sub-optimal policies.

In the next section we provide numerical simulations of our model to illustrate the issues that we have been discussing and to see how well it corresponds to the stylised facts.

3.3.9 Simulations

As we have seen in the preceding sections, the choice of policy parameters determines the steady state where the economy eventually converges in the end of transition. However, it is the development of the economy before the steady state is reached which is of great interest. It is the dynamics that are normally compared to the stylised facts and paths exhibited by economies in reality. Although we have provided some insights into the shapes of trajectories, now we illustrate the dynamics by numerically simulating our simple model for two sets of parameters. Simulations were run in GiveWin 2.10, using Ox Professional (Doornik, 2001) for programming a fourth order Runge-Kutta algorithm for a numerical solution of the dynamic system of ordinary differential equations (3.20).

Choosing parameters

We simplify by letting α and β be constant, and putting $\beta = 1 - \alpha$. In Section 3.3.7 $\alpha(\cdot)$ and $\beta(\cdot)$ are defined as the functions that capture the effect of institutions and regulations on job creation in the new and informal sectors. Here we assume that these effects are constant, and that by functioning in the shadow informal firms lose access to some boons available to new formal firms and reflected in the level of α . Then, if the total burden of regulations is measured from 0 to 1, and is equal to $1 - \alpha$, new formal firms can enjoy only share α of their profits. At the same time, informal firms save $\beta = 1 - \alpha$, but give up share α of their profits.

Having done that, the basic set of parameters is chosen as $w_o = 1$, $y = 2$, $b_u = 0.5$, $c = 1$, $m = 1$, $r = 0.1$, and $\lambda = 0.1$.

The wage/product per worker in the old sector, w_o , is normalised to 1 for convenience.

Workers in new/informal firms are seen as twice as productive, $y = 2$, which is the level of the differential between private and state marginal products of labour chosen by Brixiova (1997). The level of unemployment benefits is taken as 50% of the old sector wage. This could seem to be a high value. For example, Boeri (1999) in the context of his model treats the value of benefit replacing 35% of the wage earned in the old sector as already very high. This value does, however, look plausible for, e.g. the Russian experience with state enterprises notable for a very low level of wages. Also, as labour in our model is reallocated in the process of transition to new or informal firms paying a higher wage than the old sector, the ratio of benefits to the average wage becomes smaller. For example, if we assume that workers in the new sector receive their product of labour, equal to 2, and assign equal weights to wages in the old and new sectors then the average wage will be 1.5. In such a case unemployment benefits replace the third of the average official wage - a figure consistent with the ratio of the average unemployment benefit to the average wage reported by Rutkowski (1999) for Russia.

The mark-up values c and m are taken to be equal to 1 across the new and informal sectors. These values reflect the degree of inefficiency in market relationships between a firm and a worker (Caballero and Hammour, 1996) and are simply chosen as 50% of the maximum potential investment, y , that can be held up in the absence of any other costs. This choice satisfies one of the sufficient conditions of the existence and uniqueness of solution to system (3.20) given in Section 3.3.8. It also serves well our purpose in giving an illustrative example.

Finally, we put the interest rate, r , at 10%, and the rate of economy-wide job destruction, λ , at 10%. Davis and Haltiwanger (1999) report that in most western economies

roughly 1 in 10 jobs is destroyed every year.

What is left is to define the values of parameters γ , μ , and α . Here we consider two scenarios. The first is broadly defined as a "lagging reformer", intended to reflect the experience of such countries as Russia or Ukraine, with the rate of old sector job destruction, γ , not greater than the shock λ (i.e. it is implicitly assumed that the jobs in the old sector may be maintained due to subsidies or soft budget constraints), slack monitoring of the informal sector, and the legal and economic environment not friendly to the development of the new sector (α is low). The second scenario is called an "advanced reformer" to mirror the Polish or Estonian cases, with rate γ greater than the economy-wide shock λ , tighter monitoring of the informal economy and improved conditions on the legal, institutional, and financial front (α is high).

For the lagging reformer we put $\gamma = 0.06$ (a figure obtained by Richter and Schaffer, 1996, for Russian state-owned enterprises), $\mu = 0$ (i.e. informal firms are destroyed as often as firms in the rest of the economy), and $\alpha = 0.35$ (i.e. $\beta = 0.65$). Assuming that $\alpha = 1$ implies an economy free of corruption and where regulations are not much of a burden to the new sector, the value of 0.35 seems reasonable for such countries as Russia or Ukraine. Similar figures can be obtained by rescaling to $[0, 1]$ various indices of the legal and regulatory environment reported in EBRD (1999) and Johnson *et al.* (2000b) for the two countries.

The advanced reformer in our exercise has $\gamma = 0.15$, $\mu = 0.05$ and $\alpha = 0.7$ ($\beta = 0.3$). The figure for the rate of job destruction is consistent with the one reported for the early 1990s in Estonia by Haltiwanger and Vodopivec (2002). With $\mu = 0.05$ the cumulative rate of job destruction in the informal sector is 0.15 which is one and a half times greater

than the similar rate for the lagging reformer scenario. If μ is seen as a direct result of monitoring by the tax police then its value reflects tax police efficiency. EBRD (1999) reports that the efficiency of tax collection was roughly 1.5 times greater in Poland and Estonia than in Russia in the mid 1990s. Finally, according to the value of α for this scenario it is twice as high as the one for the lagging reformer. This is again broadly consistent with indices reported by EBRD (1999) and Johnson *et al.* (2000b)¹⁶ for, on the one hand, Poland or Estonia, and, on the other hand, Russia or Ukraine.

Before seeing the results of simulations we need to assign initial values to the stocks of workers in each of the labour market states. For both scenarios we put $O(0) = 0.9$, $N(0) = I(0) = 0.045$, and $U(0) = 0.01$. This implies that at the beginning of transition 90% of workers are employed in the old sector, new and informal sector employment is small and comparable in size, and unemployment is about 1% of the labour force (this is taken from Gregory and Collier, 1988, who estimate unemployment in the Soviet Union).

The output of simulations for the two set of parameters is presented in Fig.3.4 and 3.5. The vertical axes measure the share of sector employment or unemployment, the wage in the new sector relative to the old sector wage (in arbitrary units), output (in arbitrary units), and the share of informal output in total output, as appropriate. Time in arbitrary units (corresponding to the number of iterations) is measured on the horizontal axes.

Correspondence to stylised facts

Fig.3.4 displays the case of the lagging reformer, with the new sector starting off well but then being crowded out by the informal sector. In the meantime, the growth of the

¹⁶In particular, see Johnson *et al.* (2000b) for the data on time that managers of start-ups spend on government and regulatory matters in Poland, Russia and Ukraine.

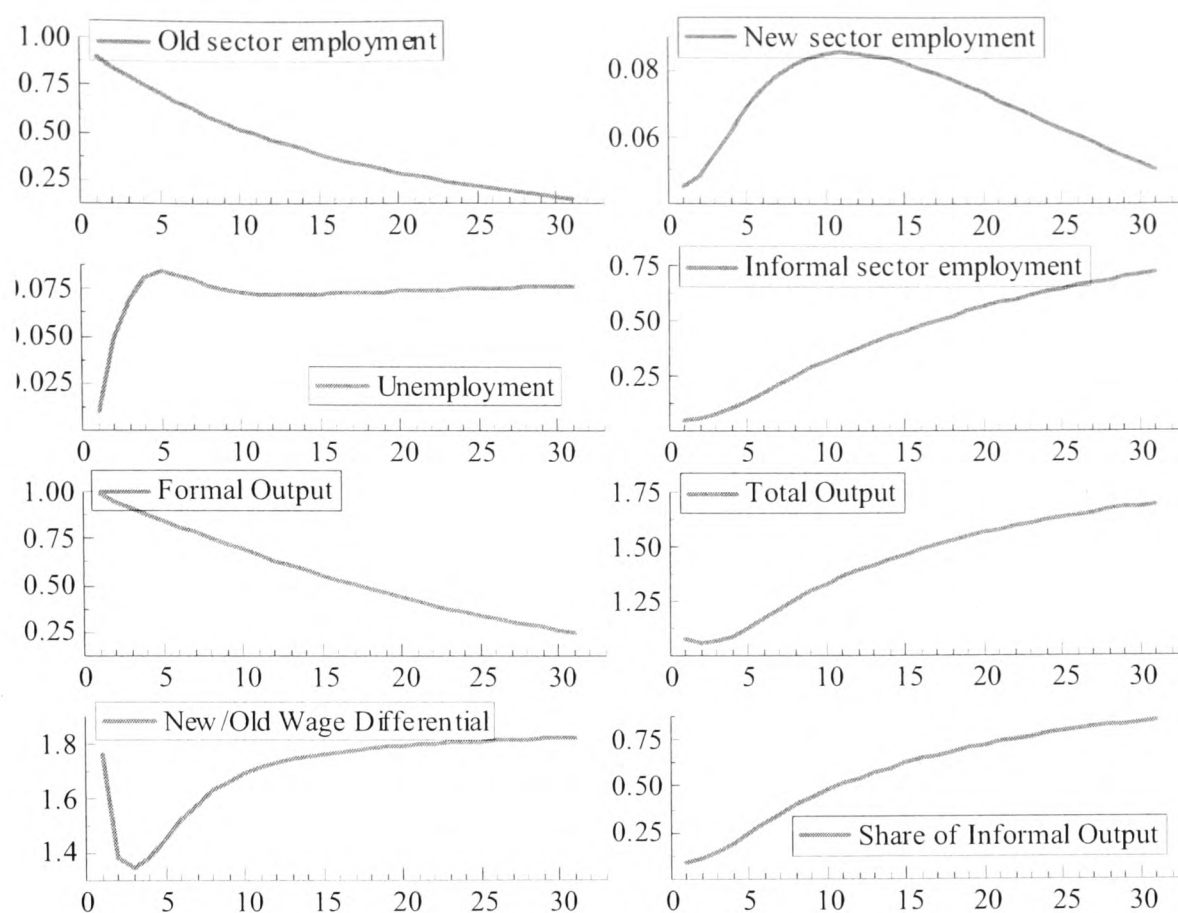


Figure 3.4: Simulations: a case of the lagging reformer - the economy converges to the steady state without the new sector

former is sustainable. The formal output exhibits L-type dynamics, while the total output follows a J-type trajectory. Such a pattern of dynamics characterises the situation when condition (3.26) applies, i.e. the burden of regulations and other factors on informal firms is lighter than that on new sector firms.

The case of the advanced reformer is given in Fig.3.5. The contrast is obvious: the new sector quickly gains speed, while the informal sector is declining after a short period of prominence. Both formal output and total output in the economy follow a J-type trajectory: a decline is followed by growth. These trajectories result from a combination of parameters satisfying condition (3.23).

It is interesting that in both cases the informal sector grows at the beginning of transition. This, as we suggested in Section 3.3.8 above, is explained by the favourable initial

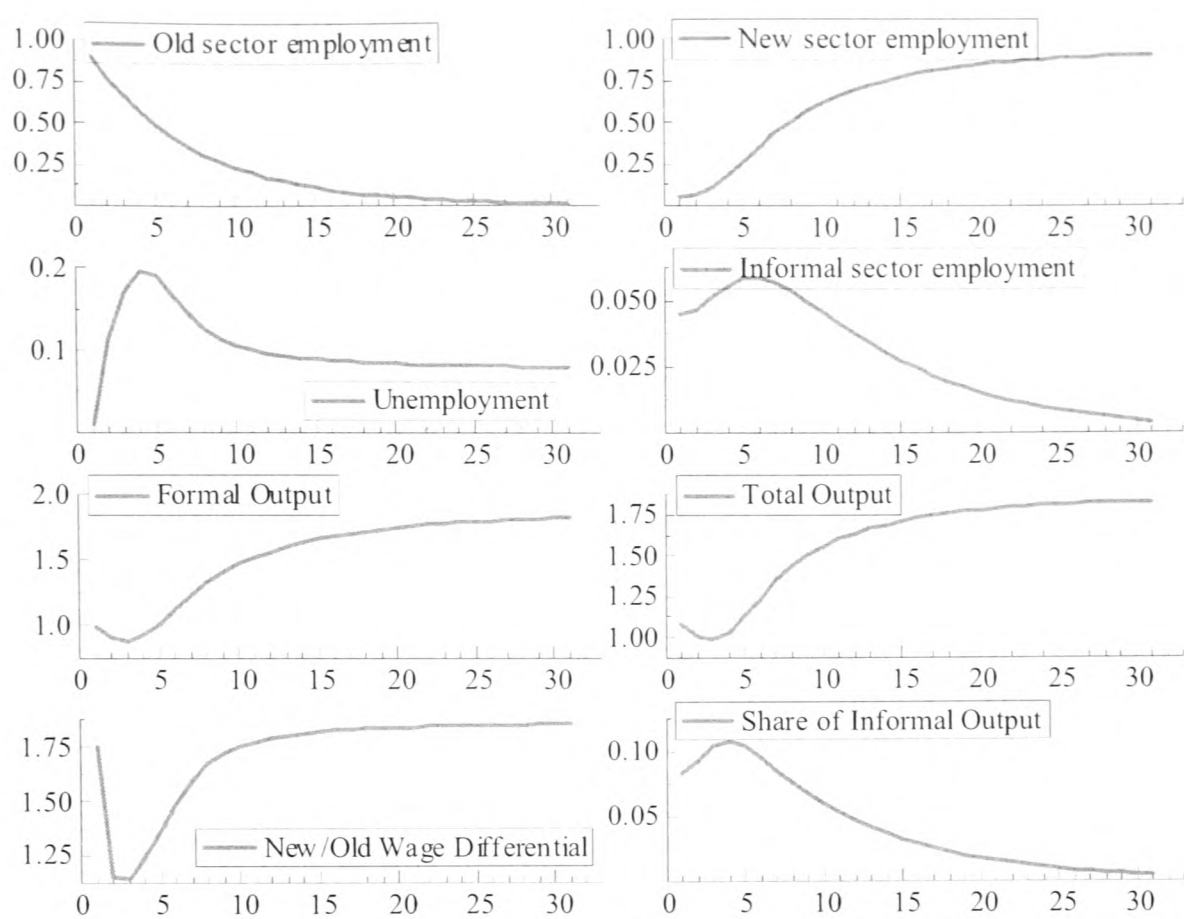


Figure 3.5: Simulations: a case of the advanced reformer - the economy converges to the steady state without the informal sector

conditions and the rise in unemployment, whose effect for some time outweighs the negative externality from the growing new sector which is small at the start. However, in the first case the rise in informality is sustained, while in the second case worker participation in the formal economy recovers after a while. In both cases unemployment follows an inverted L-type trajectory, reaching some peak in the midst of transition, decreasing somewhat afterwards, and eventually stabilising at a non-zero level.

In general, Fig.3.4 and 3.5 correspond to all the main stylised facts about employment and output dynamics in transitional countries: old sector employment declines, new sector employment grows, the unemployment path resembles an inverted L-curve. An increase in informal activity at the beginning of transition in both cases may be seen as consistent with the evidence that participation in the labour force has lowered not only in Russia

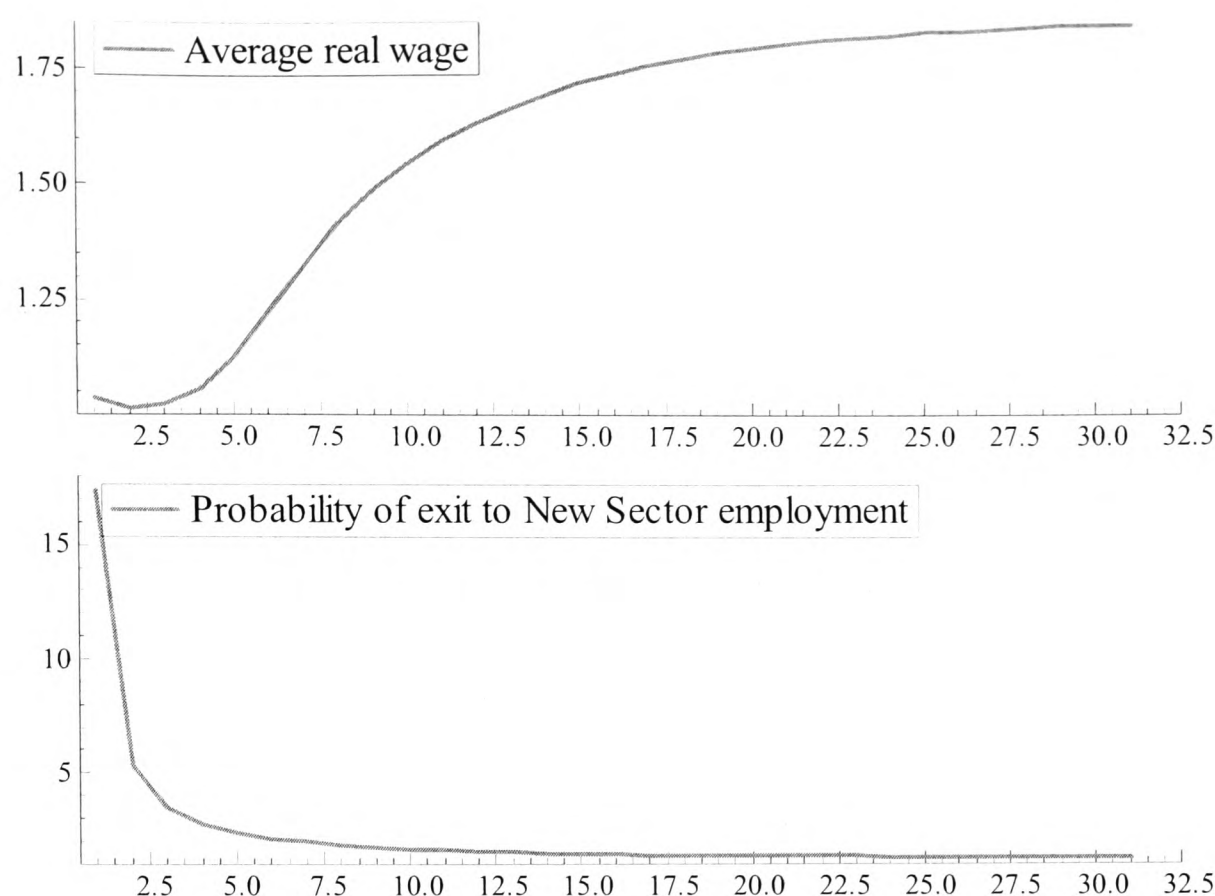


Figure 3.6: Simulations: behaviour of average wages and transition probabilities

or Ukraine, where the informal sector is believed to be notoriously big, but also in East European transitional countries.

In the background section we have mentioned that as a rule the OST studies cannot explain the falling new/old sector wage differentials and declining transition probabilities from unemployment. Fig.3.6 illustrates that our model can generate dynamics in accordance with these stylised facts too (the set of parameters is the one for the advanced reformer). The top panel shows that even for the advanced reformer average wages slightly decline at the start of transition, but quickly regain their pre-transition level afterwards. At the same time, the probability of movement out of unemployment to the new sector declines as the economy converges to the steady state. As this happens, the pool of unemployed becomes more stagnant for the duration of unemployment is decreasing in

probabilities of exit. Thus, the growing stagnancy of unemployment pools across all the transition countries can be a natural feature as their economies get closer to the steady state.

As regards the stylised facts on the difference in CEEC and CIS transitional experience, in our view the two cases presented above are telling. Broadly speaking, the first type of dynamics considered is consistent with the fortunes of such countries as Russia or Ukraine, especially in first few years of transition. The second type of dynamics is more fitting for Czech, Slovak, Hungarian, Polish, or Estonian transition. Fig.3.7 and 3.8 illustrate the development of informal sectors in the two categories of countries in the first half of the 1990s. More recent data are available in Feige and Urban (2003, Fig.A1-A4, which are partially reproduced here with permission of the authors in Fig.3.9), which pertain to the share of unrecorded income in transitional countries. Fig.3.4 and 3.5 (bottom right panel) give the dynamics of the share of informal output in our model which is generally consistent with dynamic tendencies estimated by Feige and Urban (2003) for the two groups of countries.¹⁷

Finally, Fig.3.4 clearly shows the difference in the dynamics of formal and total output in a country, the lagging reformer. This may be seen as supporting the point often made in the literature: the decline of output in Russia may have in fact been overestimated - if one takes into account the growing informal economy, the degree of the decline is much

¹⁷Notice that in Fig.3.9 some estimates of the *share* of unrecorded income in Estonia and Poland become inconceivably negative at some point. This is a result of an exercise implemented by Feige and Urban (2003) to show that some methods used for estimation of the size of the informal economy are very sensitive to certain specifying assumptions, in particular initial conditions - the degree of informalisation at the start of transition etc. Those methods may in general produce implausible values. However, the authors stress that while some estimates of the *size* of the informal economy are unreliable, "it is still possible to use the growth rate of the unrecorded sector to make important inferences about the transition process by examining the dynamic relationship between recorded and unrecorded sectors" (p.1).

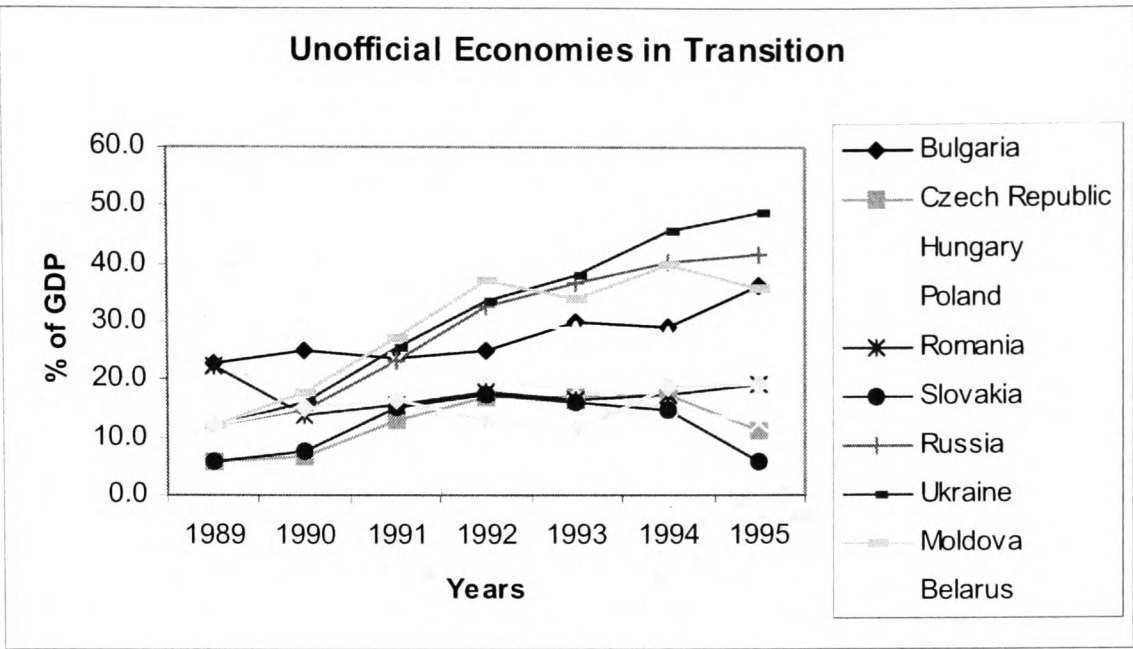


Figure 3.7: Informal sector in selected European transition countries (source: Johnson *et al.*, 1997)



Figure 3.8: Informal sector in selected FSU countries (source: Johnson *et al.*, 1997)

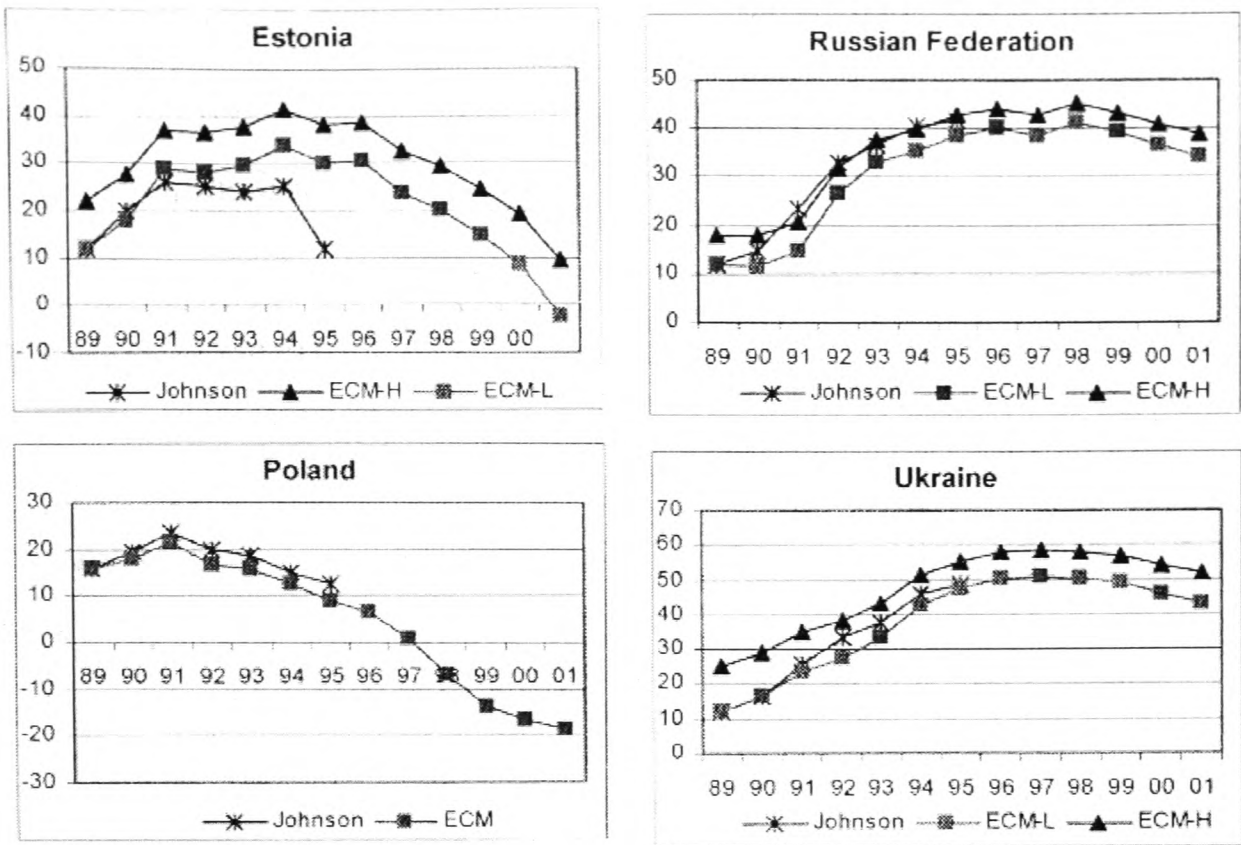


Figure 3.9: Dynamics of the share of informal economy in GDP of selected transitional countries – estimates obtained by various methods (courtesy of Ed Feige, source: Feige and Urban, 2003)

lower¹⁸ (see, for example, Stiglitz, 2002, p.152, who also points to the controversy over the GDP decline in Russia).

Has Russia switched?

Above we have suggested that the lagging reformer scenario is what probably describes the Russian or Ukrainian experience as opposed to the East European one. However, in reality a purely informal equilibrium where shadow firms dominate the post-transition economy is difficult to imagine. Moreover, if we return to Fig.3.1 it is obvious that the divergence between the CIS and CEEC took place before 1998, and after that both groups

¹⁸We do not provide here the results of simulations for the intermediate case with both new and informal sectors present in the steady state, which happens whenever condition (3.27) is met. However, the dynamics observed for such sets of parameters are very similar to the case of the lagging reformer where the informal economy prevails in the end: the economy exhibits a non-recovering drop in formal activity, while the evolution of formal output resembles a L-shaped curve.



Figure 3.10: Simulations: a case of the switching reformer

of countries have probably followed similar paths of development. This suggests that round about that time there was a switch in policies, friendliness of economic environment, and so forth, that put Russia and, possibly other countries in the region lagging in reforms, on a trajectory leading to a "good" equilibrium without the informal sector. This possibility is simulated with the help of our model and results are presented in Fig.3.10. In doing that the initial values of parameters are taken as corresponding to the lagging reformer scenario. Then starting from the 6th iteration we gradually change the values of α , β , and μ until they match up with the advanced scenario values. The change is implemented in the course of 5 periods. We still keep the value of γ , the rate of reduction in the old sector employment, at the level of the lagging reformer, i.e. 0.06. The results illustrate a case of mixed dynamics, what we call a "switching reformer" scenario. It can be seen that the

slow initial development of the new sector goes over to a comparatively buoyant growth at some point after the change in parameters took place. At the same time, the informal sector employment, gaining weight in the economy in first few years of transition, starts to decrease as does the share of the informal output. Formal output also starts growing only after the economy switches to the advanced reformer scenario. In general Fig.3.10 pictures a very slow but successful transition in terms of convergence to a steady state without the informal sector. Whether or not it properly describes or predicts the course of the Russian and Ukrainian transformation - we shall see in the years to come. However, there is some indication that such a switch to the convergence pattern, i.e. from condition (3.26) to condition (3.3.8), could happen in the Russian economy since 1998. Indeed, as we have already mentioned, the divergence between the CIS and CEEC stopped at that time, as can be seen in Fig.3.1. Meanwhile, Fig.3.9 shows that both in Russia and Ukraine the share of the informal sector also started to decline right after 1998 according to two estimates implemented by Feige and Urban (2003). That is, both Fig.3.1 and Fig.3.9 indicate that 1998, the year of the crisis of devaluation of the ruble, is a watershed in the CIS transition. The devaluation brought about better conditions for the development of some export-oriented industries and ended a period of demonetisation of the Russian economy. All this has probably contributed to the switch we have been talking about. Still, the important point made in this study is that before 1998, i.e. at the start of transition, the economies in CEEC and the CIS had probably pursued two different trajectories that are explained by convergence to two different equilibria as suggested by our model. It echoes well the conclusions made by Johnson *et al.* (1997) who argued that in the first half of the 1990s Russia and Ukraine, as opposed to the countries in Eastern Europe, were

likely stuck in a bad equilibrium with poor institutions, a large unofficial economy and an ineffective state.

3.3.10 Further Discussion and Policy Implications

Obviously, there are stylised facts that our model cannot explain. For example, a drop in real wages in the old sector, or a drop in total labour productivity observed both in Russia and CEEC. However, its main purpose is to account for the evolution of the informal sector in transition and its effect on the emerging new sector. The main message that it provides is that the asymmetries in the development of the CIS and CEEC may well be sought in the different degree of presence of the informal economy as well as the policies towards the new emerging sector.

The model does identify the parameters that make the economy converge either to a "good" equilibrium where the new sector is present in the end of transition or to a "bad" equilibrium where the informal sector crowns the system. Interestingly, the speed of old sector closure, γ , in the setting that we have considered (without embedded fiscal externalities) does not affect the eventual outcome of transformation, although it should have a bearing on distribution of gains. In general, from condition (3.27) and the subsequent analysis of the effects of changes in parameters, it follows that the higher the ratio of α to β , the lower mark up c in relation to m (or, more generally, the bigger is the difference between the new and informal sector surpluses, $S_n - S_i$) the more chances the economy has to end up in a "good" equilibrium. This has an unequivocal implication that the governments in transition should create a favourable climate for the new sector to develop, if the transformation is to be successful. This seems to be more important than the choice

of a right speed of transition - the topic in the very centre of the OST debate.

Parameters α , β , c , and m in our model reflect institutional features of the economy and inefficiencies in the new and informal sectors arising *inter alia* from specific investment. They impact on the development of the emerging sectors by affecting job creation through the amount of firms' profits available for investment. So, to ensure that the new sector takes off and thrives to succeed in transition the government must address the institutional ills. Such an implication is very much in line with earlier arguments of Caballero and Hammour (1996). In particular, in respect of the aforementioned hold-up problem they write: "When appropriability problems cannot be overcome at the microeconomic level, they receive a highly inefficient macroeconomic "solution" in the form of depressed investment and thus more difficult employment opportunities for workers. By restraining the bargaining position of workers in their transactions with firms, this endogenous response allows the economic system to guarantee the required return on investment." In our model the reduced bargaining power of workers in the new sector would imply a lower mark-up c , which raises the odds that new firms succeed. Similarly, if c reflects firms' investment that is held up because of regulations (e.g. redundancy pay), the reduction in the burden of such regulations should provide an additional incentive for the new sector firms.

In general, our result that unfavourable institutions stifle successful growth during adjustment periods resembles findings of some previous work. On the one hand, it echoes what Caballero and Hammour (1996, 2000a,b) call "sclerosis" - the inefficient survival of low-productivity jobs. In this essay we show that poor institutions lead to the development of the informal economy at the expense of the formal one, although the productivity of

both are equal. However, it is not difficult to think of a less productive informal sector in our context too, as what matters for the conclusions is not the level of productivity but the size of surpluses in the new and informal sector (see Section 3.3.8). On the other hand, our model suggests that the higher is the surplus in the new sector (i.e. the lower various costs and inefficiencies facing it), the more intense is the "entry" of new sector jobs and the more likely for the transition to be successful. This may be seen as a complement to the result of Berkowitz and Cooper (1997): in their model decreasing the cost discrimination against the entrant makes it more likely for the economy to converge eventually to a high development equilibrium. That is, the equilibrium results where *de novo* firms supply high quality goods as opposed to the low development equilibrium in which start-ups provide lower quality goods and the overall supply of goods is lower. The parallels between these two types of steady state and our "good" and "bad" outcomes are clear. Still, however, despite these similarities the distinct feature of our model is the explicit development of the informal economy.

Finally, once a set of parameters favourable to the development of the new sector is chosen, the government can think of ways to speed up the transition. Above we have shown that the presence of the informal sector in fact prolongs the convergence to a "good" equilibrium. So, the efforts may be taken to ward it off faster by increasing monitoring of firms, μ , or punishment measures for tax evasion (captured by the catch-all variable ξ) at the start of transition when the rise in unemployment makes it more likely for the informal sector to grow. At the same time, taxes should be reduced to attract more entrepreneurs in the new sector. As soon as the economy is over the hump in informal sector development the taxes on formal firms could be raised, provided that they still are

held at such a level that does not switch the economy towards a "bad" equilibrium. In general, what is important for the new sector development is not the burden created by taxes or other factors *per se*, but their effect *in comparison* to the burden on the informal sector. Thus, it is likely that in reality governments have enough flexibility to achieve a desirable outcome of transition.

3.3.11 Departures from Basic Model

The model that we have been considering so far is in fact very basic and lacks a few useful features that can make it more realistic. Here we briefly discuss two such complications: forward-looking firms in the new and informal sectors, and feedback mechanisms introducing externality effects.

Forward-looking firms

From (3.4) it is obvious that workers in our model are forward-looking. This is, however, not true of the firms in the new and informal sectors. Expressions (3.10) and (3.11) indicate that all the current profits per worker, corrected only for reactivity coefficients α and β , are myopically invested by firms. However, Aghion and Blanchard (1994) notice that job creation is likely to be at least in part forward-looking. They emphasise that many new formally established firms, and especially foreign direct investors in transitional countries, will not invest if they expect conditions to worsen and profits to get lower in the future. One should expect that similar reasoning can be applied to the newly established firms in the informal economy too.

To patch our model up to take account of the expectations on firms' side we can modify expressions (3.10) and (3.11) for job creation per worker to read:

$$rJ_n = \alpha(\tau, \zeta)(y - w_n) + \frac{dJ_n}{dt} \quad (3.30)$$

and

$$rJ_i = \beta(\mu, \xi)(y - w_i) + \frac{dJ_i}{dt}. \quad (3.31)$$

Then total job creation in the new sector will be rJ_nN , while in the informal economy it will equal rJ_iI . So, instead of investing the current profits into new job creation, firms invest the present value of their profits. Here we implicitly relax the assumption of capital market constraints.

In the model so modified expressions for wages (3.7) and (3.8) stay the same, while the transition probabilities are now redefined as $p_n = \frac{rJ_nN}{U}$ and $p_i = \frac{rJ_iI}{U}$. Substituting the expressions for wages into the Bellman equations (3.30) and (3.31) one can reduce them to:

$$\frac{dJ_n}{dt} = rJ_n \left(\frac{U + \alpha cN}{U} \right) + rJ_i \frac{\alpha mI}{U} - \alpha S_n, \quad (3.32)$$

$$\frac{dJ_i}{dt} = rJ_i \left(\frac{U + \beta mI}{U} \right) + rJ_n \frac{\beta cN}{U} - \beta S_i. \quad (3.33)$$

Together with four differential equations (3.20) they determine dynamics of our economy. As before, the same three types of steady state of this system are defined by the following conditions: 1) $O^* = 0$, $rJ_n = \lambda$ while $rJ_i < \lambda + \mu$ and $I^* = 0$ (when only the new sector is present); 2) $O^* = 0$, $rJ_i = \lambda + \mu$, while $rJ_n < \lambda$ and $N^* = 0$ (when only the informal sector is present); 3) $O^* = 0$, $rJ_n = \lambda$, $rJ_i = \lambda + \mu$ (when both the new and informal sectors are present). It is easy to show that in steady states:

$$rJ_n = \alpha \left(\frac{US_n - \beta mI(S_i - S_n)}{(U + \alpha cN + \beta mI)} \right) \quad (3.34)$$

and

$$rJ_i = \beta \left(\frac{US_i - \alpha cN(S_n - S_i)}{(U + \alpha cN + \beta mI)} \right), \quad (3.35)$$

which are essentially the same expressions as in (3.14) and (3.15). Thus, the conditions for the steady states are the same as in the case of myopic firms.

The introduction of forward-looking job creation affects only paths leading to an equilibrium point, but not the steady states themselves. Aghion and Blanchard (1994) show that convergence to a steady state will in such a case depend not only on the initial levels of unemployment and employment in sectors, but also on the initial level of investment into new job creation. In such a situation expectations matter a lot and they become self-fulfilling: if job creation is low at the beginning of transition, the steady state may never be reached.

Feedback mechanisms and multiple equilibria

Leaving aside the forward-looking firms, we move on to another important departure from our basic model. This would be an introduction of fiscal or another type of externalities leading to multiple equilibria for each set of parameters. Algebraically, it would imply that functions $\alpha(\tau, \zeta)$ or/and $\beta(\mu, \xi)$ depend on stocks of workers N , U , and I in addition to policy parameters such as τ and μ . This may happen if, for example, the gov-

ernment always runs a balanced budget such that $\tau J_n N = b_u U$. An alternative feedback mechanism can be at work when the effectiveness of monitoring of the informal sector directly depends on the funds available in the state budget, i.e., generally speaking, on the amount of taxes collected: $\mu = \tau J_n N$.

In such cases the above expressions for w_n , w_i , J_n , and J_i become very complicated. The signs of their derivatives with respect to U (or/and N and I) become ambiguous as a result of additional externality effects brought in. Then solutions to conditions $J_n = \lambda$ or/and $J_i = \lambda + \mu$ could well no longer be unique. The number of possible steady states of the dynamic system could well increase. So, the problem of choosing a correct combination of policy parameters becomes more involved. In particular, one should expect that the choice of the speed of old sector closure, γ , is no longer neutral with respect to the eventual outcome: through the effect on the size of unemployment it would have a direct bearing on the convergence properties of the system.

3.4 Concluding Remarks

In this work we provided an extensive review of the literature taking up the issue of finding the Optimal Speed of Transition. In our critical comments we highlight that the obvious weak spot of these studies is their ignorance about informal sector employment widely observed in transitional countries. On the policy front, they excessively focus on the proper engineering of job destruction in the old inefficient sector, and the reasonable choice of social support programmes (unemployment benefits, wage floors, etc.). The job creation part of the transition process and policies towards the emerging new sector have received hardly any attention in the theory. The major failure of this literature, pretending

to provide a theoretical model of transition, is the inability to explain convincingly the noticeable differences in the development of East European countries, on the one hand, and the countries of the Former Soviet Union (apart from the Baltic states), on the other hand.

In our attempt to close the gap we have presented a model of transition in the presence of the informal sector, where qualitatively different outcomes of transformation depend on the choice of various parameters, many of which can be affected by the government. In particular, we have showed that transition can fail if the pressure of various factors on new formal firms outweighs the pressure on the informal sector. In such a case it is the informal firms that dominate the post-transition economy. We argue that the convergence to qualitatively different equilibria helps explain the varying experience of CEE and CIS countries. This point should be seen as complementing rather than substituting the previous attempts to explain the CEE and CIS divide made by Garibaldi and Brixiova (1998) and Boeri and Terrell (2002), who argue that it is the differences in unemployment benefits, minimum wages, and other labour market institutions that have contributed to the divergence in question.

In general, our model highlights three main aspects theoretically vitally important to shaping the outcome of transition.

First, we mostly support the point made earlier by many other researchers that creation of institutions favourable to the development of the new sector should be the paramount objective of transition. While the speed of transition - the darling of many previous OST studies - does make the difference for the distribution of gains, it may not be at all important for the qualitative result, as our model shows. At the same time, the burden

of taxes, labour market regulations, legal infrastructure, and other factors captured by the "reactivity coefficients" measuring friendliness of the economic environment towards particular sectors, defines the proportion of profits directed to investment purposes and the rate of job creation.¹⁹ What is interesting, however, is that the model suggests that it is not, for instance, the high taxes *per se* that may be blamed for snarling up the successful development of the new sector, but rather the burden that they create on new firms in comparison to the burden on their informal counterparts.

Second and getting more specific, we argue that it is the labour market conditions in the formal and informal sectors that make the difference. In particular, we consider the mark-ups that new formal and informal firms pay their workers over the value of being unemployed. As discussed in the text these mark-ups can be given a variety of explanations, but we suggest seeing them as resulting from the appropriability problem raised in the developing economics context by Caballero and Hammour (1996, 2000a,b). Reducing the opportunities for investment hold-ups in the new sector in comparison with the informal sector is one of the policy implications of the model for successful transition.

Finally, the presence of the informal sector undoubtedly prolongs the timing of transition. However, the upsurge in unemployment at the start of transition as well as the initial conditions can lead to an increase in size of the shadow economy regardless of the type of steady state that the economy eventually converges to. Thus, efforts could be made to reduce the extent of informality as quickly as possible provided that in general the economic environment is favourable to the development of the new formal sector. It is suggested that such a reduction can be achieved, for example, through more efficient

¹⁹EBRD (2000) reports that in CEE setting up a business was not unusual even before the transition, whereas in other countries, for example, Ukraine, it was discouraged by a punitive use of taxation and other measures.

monitoring of informal activities that would affect the rate of death of informal firms. This has a direct impact on the labour turnover in the sector and, thus, the development of informality. By increasing the monitoring efficiency, leading to the higher incidence of detection of informal activities and their liquidation, the authorities can hinder the take-off of the informal economy and ensure the success of transformation.

All in all, the main message of our model is that costs and the eventual success of transformation are determined not only by the speed of restructuring of the old sector, but also, importantly, by the policies towards emerging new businesses. If such policies are highly regulatory and restrictive, especially in comparison to the policies towards the notorious informal economy, it is the latter not the formal new firms that eventually prevail in the economy. More generally, the model presented here supports the point recently made by Stiglitz (2002, p.57) that "privatization needs to be part of a more comprehensive program, which entails creating jobs in tandem with the inevitable job destruction that privatization often entails."

3.A New and Informal Sector Mark-ups:

Cost of Separation Interpretation

Consider new sector firms and unemployed workers. The values of working in the new sector, V_n , and of being unemployed, V_u , are given in (3.1) and (3.3). In the main text we do not explicitly consider the values of an open vacancy and of a filled job to the firm. Let us denote these by V_v and V_j , respectively. Then the total surplus of a match is given by $S = V_n - V_u + V_j - V_v$. The generalised Nash bargaining solution implies that the wage is chosen so as to split the surplus as $V_n - V_u = \epsilon S$ and $V_j - V_v = (1 - \epsilon) S$, where ϵ is seen as a parameter reflecting impatience of the two sides, or bargaining power of workers.

Under the assumption of no matching problem for firms creating a new vacancy the surplus of having a job filled rather than vacant is equal to zero (Blanchard, 1998). That is, if a candidate for a job decides not to take it up another one can be hired on the spot. Then whatever is the bargaining power of workers, they are effectively held down to their reservation value while haggling over the wage: $V_n - V_u = 0$. However, suppose that potential separation implies a cost to the firm. This may be interpreted as initial training (lasting one instant) that employer provides the worker with at some cost k , or a state imposed redundancy cost. In both cases it is supposed that worker can renegotiate after being officially hired. Then the cost to the firm of separating a worker is k , which is independent of labour market conditions because of the no-matching-problem assumption (*ibid*). Then Nash bargaining implies:

$$V_j - V_v = (1 - \epsilon) S = k$$

and

$$V_n - V_u = \epsilon S = \frac{\epsilon k}{(1 - \epsilon)}.$$

Thus, as we can see from the two equations above, constant c , defined in the main text as equal to $V_n - V_u$, depends on the value of separation costs in the new sector and the relative bargaining power of workers and firms: $c = \frac{\epsilon k}{(1 - \epsilon)}$.

Similarly, in the informal sector constant m , equal to $V_i - V_u$, depends on the value of separation costs and the relative bargaining power of workers in the informal economy.

3.B Derivation of Basic Results

3.B.1 Wages and Job Creation in the New and Informal Sectors

Wages in the new and informal sectors are given by, respectively,

$$w_n(U, N, I) = \left(y - \frac{US_n - \beta m I(S_i - S_n)}{(U + \alpha c N + \beta m I)} \right)$$

and

$$w_i(U, N, I) = \left(y - \frac{US_i - \alpha c N(S_n - S_i)}{(U + \alpha c N + \beta m I)} \right).$$

Job creation rates per worker employed in the new and informal sectors are, respectively:

$$J_n(U, N, I) = \alpha(\cdot)(y - w_n) = \alpha\left(\frac{US_n - \beta m I(S_i - S_n)}{(U + \alpha c N + \beta m I)}\right)$$

and

$$J_i(U, N, I) = \beta(\cdot)(y - w_i) = \beta\left(\frac{US_i - \alpha c N(S_n - S_i)}{(U + \alpha c N + \beta m I)}\right),$$

where $S_n = (y - b_u - (r + \lambda)c)$, $S_i = (y - b_u - (r + \lambda + \mu)m)$, $\alpha \equiv \alpha(\cdot)$ and $\beta \equiv \beta(\cdot)$.

As Aghion and Blanchard (1994) we assume that N (or I , for that matter) is 0 if profits in the sector are negative. In other words, emerging firms close if they are losing money. Then, from proportionality of $J_n(U, N, I)$ and $J_i(U, N, I)$ to profits it follows that the necessary condition for both sectors to exist is both $J_n(U, N, I)$ and $J_i(U, N, I)$ being positive. Otherwise, sector profits are negative, as by assumption both $\alpha \equiv \alpha(\cdot)$ and $\beta \equiv \beta(\cdot)$ belong to $[0, 1]$. It is easy to show that whenever U , N , and I are positive, the necessary and sufficient conditions for $J_n(U, N, I)$ and $J_i(U, N, I)$ to be positive is that either $\frac{(r + \lambda + \mu)}{(r + \lambda)}m < c < \frac{y - b_u}{(r + \lambda)}$ or $c < \frac{(r + \lambda + \mu)}{(r + \lambda)}m < \frac{y - b_u}{(r + \lambda)}$ hold. Then it is also true that wages never rise above y .

Properties of functions $J_n(U, N, I)$ and $w_n(U, N, I)$

1. $J_n(U, N, I) = \alpha\left(\frac{US_n - \beta m I(S_i - S_n)}{(U + \alpha c N + \beta m I)}\right) \geq 0$ by simple logic: otherwise the sector ceases to exist as the negativity of J_n also implies negative profits.

2. $\frac{\partial J_n(U, N, I)}{\partial U} = \alpha \left(\frac{S_n(\alpha c N + \beta m I) + \beta m I(S_n - S_i)}{(U + \alpha c N + \beta m I)^2} \right) = \alpha \left(\frac{\alpha c N S_n + \beta m I S_i}{(U + \alpha c N + \beta m I)^2} \right) > 0$, whenever $S_n > 0$ and $S_i > 0$ (otherwise one of the sectors simply does not exist, because, again, sector profits become negative).
3. $\frac{\partial J_n(U, N, I)}{\partial N} = -\alpha^2 c \left(\frac{U S_n - \beta m I(S_i - S_n)}{(U + \alpha c N + \beta m I)^2} \right) \leq 0$.
4. $\frac{\partial J_n(U, N, I)}{\partial I} = -\alpha \beta m \left(\frac{U S_n - \alpha c N(S_n - S_i)}{(U + \alpha c N + \beta m I)^2} \right) \leq 0$ (follows from the fact that $J_i(U, N, I) \geq 0$).
5. $w_n(U, N, I) = \left(y - \frac{U S_n - \beta m I(S_i - S_n)}{(U + \alpha c N + \beta m I)} \right) \geq 0$.
6. $\frac{\partial w_n(U, N, I)}{\partial U} = -\frac{\alpha c N S_n + \beta m I S_i}{(U + \alpha c N + \beta m I)^2} \leq 0$.
7. $\frac{\partial w_n(U, N, I)}{\partial N} = \alpha c \frac{U S_n - \beta m I(S_i - S_n)}{(U + \alpha c N + \beta m I)^2} \geq 0$, as $J_n(U, N, I) \geq 0$.
8. $\frac{\partial w_n(U, N, I)}{\partial I} = \beta m \frac{U S_n - \alpha c N(S_n - S_i)}{(U + \alpha c N + \beta m I)^2} \geq 0$, as $J_i(U, N, I) \geq 0$ (see below).

Properties of functions $J_i(U, N, I)$ and $w_i(U, N, I)$

1. $J_i(U, N, I) = \beta \left(\frac{U S_i - \alpha c N(S_n - S_i)}{(U + \alpha c N + \beta m I)} \right) \geq 0$, by analogy to the case of $J_n(U, N, I)$ above.
2. $\frac{\partial J_i(U, N, I)}{\partial U} = \beta \left(\frac{\alpha c N S_n + \beta m I S_i}{(U + \alpha c N + \beta m I)^2} \right) \geq 0$, whenever $S_n > 0$ and $S_i > 0$ (otherwise one of the sectors simply does not exist).
3. $\frac{\partial J_i(U, N, I)}{\partial N} = -\alpha c \beta \left(\frac{U S_n - \beta m I(S_i - S_n)}{(U + \alpha c N + \beta m I)} \right) \leq 0$, as $J_n(U, N, I) \geq 0$.
4. $\frac{\partial J_i(U, N, I)}{\partial I} = -\beta^2 m \left(\frac{U S_i - \alpha c N(S_n - S_i)}{(U + \alpha c N + \beta m I)^2} \right) \leq 0$, as $J_i(U, N, I) \geq 0$.
5. $w_i(U, N, I) = \left(y - \frac{U S_i - \alpha c N(S_n - S_i)}{(U + \alpha c N + \beta m I)} \right) \geq 0$.
6. $\frac{\partial w_i(U, N, I)}{\partial U} = -\frac{\alpha c N S_n + \beta m I S_i}{(U + \alpha c N + \beta m I)^2} \leq 0$.
7. $\frac{\partial w_i(U, N, I)}{\partial N} = \alpha c \frac{U S_n - \beta m I(S_i - S_n)}{(U + \alpha c N + \beta m I)^2} \geq 0$, as $J_n(U, N, I) \geq 0$.
8. $\frac{\partial w_i(U, N, I)}{\partial I} = \beta m \frac{U S_i - \alpha c N(S_n - S_i)}{(U + \alpha c N + \beta m I)^2} \geq 0$, as $J_i(U, N, I) \geq 0$.

3.B.2 Steady States

Only the new sector is present

Conditions for this steady state are $J_n = \lambda$, while $J_i < (\lambda + \mu)$, $I^* = 0$, and $O^* = 0$.

Rewriting the condition $J_n = \lambda$ as $\alpha \left(\frac{U^* S_n + \beta m I^* (S_n - S_i)}{(U^* + \alpha c N^* + \beta m I^*)} \right) = \lambda$, and recalling that $U^* = 1 - N^*$ in steady state, yields:

$$N^* = \frac{(\alpha S_n - \lambda)}{(\alpha S_n - \lambda + \lambda \alpha c)}$$

and

$$U^* = \frac{\lambda \alpha c}{(\alpha S_n + \lambda(\alpha c - 1))} = \frac{\lambda \alpha c}{(\alpha S_n - \lambda + \lambda \alpha c)}.$$

The necessary and sufficient condition for N^* and U^* to be positive in the steady state is $\alpha S_n > \lambda$.

The existence condition Rewriting the condition $J_i < (\lambda + \mu)$ we obtain

$$J_i = \beta \left(\frac{U S_i - \alpha c N (S_n - S_i)}{(U + \alpha c N + \beta m I)} \right) |_{ss} < (\lambda + \mu).$$

This is equivalent to

$$\beta \left(\frac{\lambda \alpha c S_i - \alpha c (\alpha S_n - \lambda) (S_n - S_i)}{\lambda \alpha c + \alpha c (\alpha S_n - \lambda)} \right) < (\lambda + \mu)$$

or

$$\beta < \frac{(\lambda + \mu)}{\left(\frac{\lambda \alpha c S_i - \alpha c (\alpha S_n - \lambda) (S_n - S_i)}{\lambda \alpha c + \alpha c (\alpha S_n - \lambda)} \right)}.$$

Then $\alpha > \frac{\lambda \beta}{((\lambda + \mu) + \beta(S_n - S_i))}$.

So, this steady state exists whenever $\alpha > \frac{\lambda \beta}{((\lambda + \mu) + \beta(S_n - S_i))}$.

The effect of parameter changes on equilibrium value of unemployment

1. $\frac{\partial U^*}{\partial \alpha} = \frac{\partial}{\partial \alpha} \left(\frac{\lambda \alpha c}{(\alpha S_n + \lambda(\alpha c - 1))} \right) = -\frac{\lambda^2 c}{(\alpha S_n + \lambda(\alpha c - 1))^2} < 0$.
2. $\frac{\partial U^*}{\partial \mu} = \frac{\partial}{\partial \mu} \left(\frac{\lambda \alpha c}{(\alpha S_n + \lambda(\alpha c - 1))} \right) = 0$, unless α depends on μ .
3. $\frac{\partial U^*}{\partial c} = \frac{\partial}{\partial c} \left(\frac{\lambda \alpha c}{(\alpha S_n + \lambda(\alpha c - 1))} \right) = \lambda \alpha \frac{\alpha(y - b_u) - \lambda}{(\alpha S_n + \lambda(\alpha c - 1))} > 0$ if the necessary and sufficient condition for U^* to be positive holds.
4. $\frac{\partial U^*}{\partial m} = \frac{\partial}{\partial m} \left(\frac{\lambda \alpha c}{(\alpha S_n + \lambda(\alpha c - 1))} \right) = 0$.

Only the informal sector is present

Conditions for this steady state are $J_i = (\lambda + \mu)$, while $J_n < \lambda$, $N^* = 0$, and $O^* = 0$.

Then by rewriting $J_i = (\lambda + \mu)$ as

$$\beta \left(\frac{U^* S_i - \alpha c N^* (S_n - S_i)}{(U^* + \alpha c N^* + \beta m I^*)} \right) = (\lambda + \mu)$$

with $U^* = 1 - I^*$ we get

$$I^* = \frac{(\beta S_i - (\lambda + \mu))}{(\beta S_i - (\lambda + \mu) + (\lambda + \mu)\beta m)},$$

and

$$U^* = \frac{(\lambda + \mu)\beta m}{(\beta S_i - (\lambda + \mu) + (\lambda + \mu)\beta m)} = \frac{(\lambda + \mu)\beta m}{(\beta S_i + (\lambda + \mu)(\beta m - 1))}.$$

It follows that the necessary and sufficient condition for I^* and U^* to be positive in the steady state is $\beta S_i > (\lambda + \mu)$.

The existence condition The condition $J_n < \lambda$ implies

$$J_n = \alpha \left(\frac{U S_n + \beta m I (S_n - S_i)}{(U + \alpha c N + \beta m I)} \right) |_{ss} < \lambda.$$

From this formula it follows that the steady state exists whenever

$$\alpha < \frac{\lambda}{\left(\frac{(\lambda + \mu)\beta m S_n + \beta m (\beta S_i - (\lambda + \mu))(S_n - S_i)}{(\lambda + \mu)\beta m + \beta m (\beta S_i - (\lambda + \mu))} \right)}$$

or

$$\alpha < \frac{\lambda \beta}{((\lambda + \mu) + \beta(S_n - S_i))}.$$

The effect of parameter changes on equilibrium value of unemployment

1. $\frac{\partial U^*}{\partial \beta} = \frac{\partial}{\partial \beta} \left(\frac{(\lambda + \mu)\beta m}{(\beta S_i - (\lambda + \mu) + (\lambda + \mu)\beta m)} \right) = -\frac{(\lambda + \mu)^2 m}{(\beta S_i - (\lambda + \mu) + (\lambda + \mu)\beta m)} < 0.$
2. $\frac{\partial U^*}{\partial \mu} = \frac{\partial}{\partial \mu} \left(\frac{(\lambda + \mu)\beta m}{(\beta S_i - (\lambda + \mu) + (\lambda + \mu)\beta m)} \right) = \frac{\beta^2 m (y - b_u - \mu m)}{(\beta S_i - (\lambda + \mu) + (\lambda + \mu)\beta m)} > 0.$
3. $\frac{\partial U^*}{\partial c} = \frac{\partial}{\partial c} \left(\frac{(\lambda + \mu)\beta m}{(\beta S_i - (\lambda + \mu) + (\lambda + \mu)\beta m)} \right) = 0.$
4. $\frac{\partial U^*}{\partial m} = \frac{\partial}{\partial m} \left(\frac{(\lambda + \mu)\beta m}{(\beta S_i - (\lambda + \mu) + (\lambda + \mu)\beta m)} \right) = (\lambda + \mu) \beta \frac{\beta(y - b_u) - (\lambda + \mu)}{(\beta S_i - (\lambda + \mu) + (\lambda + \mu)\beta m)} > 0$ if the necessary and sufficient condition for U^* to be positive maintains in this steady state.

Both new and informal sectors are present

The conditions for this steady state to take place are $J_n = \lambda$ and $J_i = (\lambda + \mu)$, while $O^* = 0$. The conditions can be rewritten as:

$$\alpha \left(\frac{US_n + \beta m I(S_n - S_i)}{(U + \alpha c N + \beta m I)} \right) = \lambda$$

and

$$\beta \left(\frac{US_i - \alpha c N(S_n - S_i)}{(U + \alpha c N + \beta m I)} \right) = (\lambda + \mu),$$

where $U^* = 1 - N^* - I^*$.

The existence condition It is easy to see that in fact J_n can be reexpressed through J_i :

$$J_n = \alpha \left(\frac{US_n + \beta m I(S_n - S_i)}{(U + \alpha c N + \beta m I)} \right) = \frac{\alpha}{\beta} \beta \left(\frac{US_i - \alpha c N(S_n - S_i)}{(U + \alpha c N + \beta m I)} \right) + \alpha (S_n - S_i) = \frac{\alpha}{\beta} J_i + \alpha (S_n - S_i).$$

So, if $J_i = (\lambda + \mu)$ in the steady state then $\frac{\alpha}{\beta} (\lambda + \mu) + \alpha (S_n - S_i) = \lambda$. This last equality implies that the steady state exists whenever

$$\alpha = \frac{\beta \lambda}{((\lambda + \mu) + \beta (S_n - S_i))},$$

i.e. it is the knife-edge condition (3.27) in the main text.

Multiplicity of equilibria When the knife-edge condition holds it is possible to show that the equilibria are linked by the relation

$$N^* = \frac{(\beta S_i - (\lambda + \mu))}{(\beta S_i - (\lambda + \mu) + \beta c \lambda)} - I^* \frac{(\beta S_i - (\lambda + \mu) + \beta m (\lambda + \mu))}{(\beta S_i - (\lambda + \mu) + \beta c \lambda)}.$$

The concrete size of the formal and informal sectors in such an equilibrium depends on initial conditions as well as various combinations of policy parameters. Indeed, in equilibrium we have

$$N^* = \exp \left(\int_0^{t^*} (J_n - \lambda) dt \right) N(0)$$

and

$$I^* = \exp \left(\int_0^{t^*} (J_i - \lambda - \mu) dt \right) I(0).$$

where t^* is time necessary to reach the steady state. It is also possible to show that the condition (3.27) implies $(J_n - \lambda) = \frac{\alpha}{\beta} (J_i - (\lambda + \mu))$. Then

$$N^* = \exp \left(\frac{\alpha}{\beta} \int_0^{t^*} (J_i - (\lambda + \mu)) dt \right) N(0),$$

so that

$$\frac{N^*}{I^*} = \exp \left(\left(\frac{\alpha}{\beta} - 1 \right) \int_0^{t^*} (J_i - \lambda - \mu) dt \right) \frac{N(0)}{I(0)}.$$

First, as we can see from the factor before the integral in the expression above, the relative size of N^* and I^* depends on the ratio of $\alpha(\cdot)$ and $\beta(\cdot)$ even if the integral is the same for various sets of policy parameters satisfying the knife-edge condition, and if $N(0) = I(0)$. That is, by changing $r, c, m, \lambda, \mu, \alpha(\cdot)$, and $\beta(\cdot)$ so that the condition (3.27) holds, the value of $\frac{N^*}{I^*}$ depends *ceteris paribus* on the ratio of $\alpha(\cdot)$ and $\beta(\cdot)$. So, for any set of initial conditions the knife-edge equilibrium is not generally unique.

Second, the value of the integral in turn depends on the initial conditions, namely $U(0), N(0)$ and $I(0)$, in addition to the values of $r, c, m, \lambda, \mu, \alpha(\cdot)$, and $\beta(\cdot)$, through job creation function $J_i(U, N, I)$. Thus, for each knife-edge condition the corresponding steady state is not generally unique either, and depends on the initial conditions.

From the multiplicity of equilibria it follows that we clearly have a coordination problem in this case: under the restriction that $\alpha = \frac{\beta\lambda}{((\lambda+\mu)+\beta(S_f-S_i))}$ the government still can affect the relative level of informal employment and achieve various levels of welfare by changing correspondence between parameters, in particular $\alpha(\cdot)$ and $\beta(\cdot)$.

The effect of parameter changes Obviously, whenever the knife-edge condition is violated we immediately end up in one of the two cases considered above - the system starts converging to another steady state equilibrium. This implies that these steady states are very sensitive to parameter changes.

For example, consider what happens if λ rises. The derivative of the right hand side of (3.27) is

$$\frac{\partial}{\partial \lambda} \frac{\beta\lambda}{((\lambda+\mu)+\beta(S_f-S_i))} = \frac{\partial}{\partial \lambda} \frac{\lambda\beta}{((\lambda+\mu)+\beta((r+\lambda+\mu)m-(r+\lambda)c))} = \frac{\beta\mu+\beta^2((r+\mu)m-rc)}{((\lambda+\mu)+\beta((r+\lambda+\mu)m-(r+\lambda)c))^2}.$$

The sign of this expression is ambiguous, unless one makes an additional *ad hoc* assumption regarding the relative value of m and c . However, by assumption $\beta < 1$ so that the second term in the numerator is likely to be very low. Then it is likely that an increase in λ leads to an increase in the right hand side of the knife-edge condition, thus putting the economy in the situation where it converges to the steady state without the new sector.

3.C Behaviour of Job Creation Functions

In this appendix we pay some more attention to the behaviour of job creation rates in the neighbourhood of no-growth points $J_n(U, N, I) = \lambda$ and $J_i(U, N, I) = \lambda + \mu$. This provides some insight into the shape of trajectories as the economy converges to a steady state.

Above we have shown that

$$J_n(U, N, I) = \frac{\alpha}{\beta} J_i(U, N, I) + \alpha(S_n - S_i),$$

where S_n and S_i are the surpluses in the new and informal sectors, respectively.

Also, it is straightforward to derive that

$$\frac{dJ_n(U, N, I)}{dt} = \frac{\alpha}{\beta} \frac{dJ_i(U, N, I)}{dt}.$$

Consider the case when the policy parameters are favourable to the existence of the steady state with the sole presence of the new sector in the end of transition (other cases can be analysed by analogy). That is, suppose that

$$\alpha > \frac{\lambda\beta}{((\lambda+\mu)+\beta(S_n-S_i))}.$$

It can readily be shown that under this condition we have

$$J_n - \lambda > \frac{\alpha}{\beta} (J_i - (\lambda + \mu)).$$

In other words, whenever the informal sector is growing, i.e. $J_i > (\lambda + \mu)$, the new sector is growing too. Likewise, whenever the new sector is declining, $J_n < \lambda$, the informal economy is also in recession. Assume that from the start of transition both sectors are

growing. We know that in the steady state $J_n = \lambda$. This means that rates J_n and J_i must be decreasing. What happens to the size of the two sectors?

Let us denote by t_n and t_i the times necessary to achieve the point where the sector sizes start to decrease. That is, if at time moment t_0 we have $J_n = J_n^* > \lambda$ and $J_i = J_i^* > \lambda + \mu$, and J_n and J_i are decreasing, as $\frac{\partial J_n}{\partial N} < 0$, $\frac{\partial J_n}{\partial I} < 0$, $\frac{\partial J_i}{\partial N} < 0$, and $\frac{\partial J_i}{\partial I} < 0$, it takes t_n and t_i for the new and the informal sectors, respectively, to reach the points where $J_n = \lambda$ and $J_i = \lambda + \mu$. Then t_n and t_i satisfy the equations

$$\int_{t_0}^{t_n} \frac{dJ_n(U, N, I)}{dt} dt = J_n^* - \lambda \quad (3.36)$$

and

$$\int_{t_0}^{t_i} \frac{dJ_i(U, N, I)}{dt} dt = J_i^* - \lambda - \mu, \quad (3.37)$$

respectively.

Applying the above relations between J_n and J_i and their full derivatives to the first integral yields

$$\int_{t_0}^{t_n} \frac{dJ_n(U, N, I)}{dt} dt = \frac{\alpha}{\beta} \int_{t_0}^{t_n} \frac{dJ_i(U, N, I)}{dt} dt = \frac{\alpha}{\beta} J_i^* + \alpha(S_n - S_i) - \lambda. \quad (3.38)$$

From $\frac{\alpha}{\beta} > \frac{\lambda}{(\lambda + \mu) + \beta(S_n - S_i)}$ it follows that $(\lambda + \mu) > \frac{\beta}{\alpha} \lambda - \beta(S_n - S_i)$, i.e.

$$J_i^* + \beta(S_n - S_i) - \frac{\beta}{\alpha} \lambda > J_i^* - \lambda - \mu.$$

Thus, from (3.37) and (3.38) it follows that $t_i < t_n$.

This implies that even when initially both sectors grow, the informal sector is the first to reach the edge when it starts to decrease, while the new sector will still be growing until it gets into the steady state (meanwhile, the informal sector must completely disappear as the steady state is reached). So, the shape of the trajectory followed by the new sector resembles an inverted L-curve, while the path taken by the informal sector mimics an inverted J-curve.

Now let us assume that the initial conditions are such that $J_n^* < \lambda$ and $J_i^* < \lambda + \mu$, i.e. both sectors are decreasing at the start of transition. As they decrease and unemployment increases, $\frac{dJ_n(U, N, I)}{dt}$ and $\frac{dJ_i(U, N, I)}{dt}$ become positive, so that the decline in size of the new and informal sectors gets more moderate. By applying a reasoning similar to the one above, it can be shown that the new sector is the first to achieve the point where $J_n > \lambda$ and it could start growing, while the informal sector will still be shrinking.

Numerical simulations in Section 3.3.9 provide an illustrative example of the dynamics of our economy.

3.D The Effect of the Informal Sector on the Timing of Transition

In the main text we argue that the presence of the informal sector slows down the transition. To see this, let us assume that a set of parameters is chosen so that condition (3.23) favourable to the prevalence of the new sector in the end is satisfied. The fixed set of parameters defines the resulting steady state level of new sector employment, N^* , and unemployment, U^* .

Let us consider two paths followed by new sector employment (see Fig.3.11). The first trajectory is given by solution $N_n(t)$ to system (3.20) with initial conditions $O(0) = O_0$, $N(0) = N_n(0) = N_0$, $I(0) = 0$, $U(0) = U_0$. The second trajectory is given by solution $N_i(t)$ to the dynamic system with initial conditions $O(0) = O_0$, $N(0) = N_i(0) = N_0 - \varepsilon$, $I(0) = \varepsilon$, $U(0) = U_0$, where ε is some infinitesimally small quantity. The question is whether or not the new sector converges to N^* faster in the absence rather than in the presence of the informal sector? We will say that the former case takes place if $N_n(t) > N_i(t)$ for all $t > t_c$, whereas the latter case takes place when $N_n(t) < N_i(t)$ for all $t > t_c$, where t_c is some moment of time. This is appropriate as from (3.20) it follows that the convergence is asymptotic and N^* is reached when $t = \infty$. Suppose that $N_i(t) > N_n(t)$ for any $t > t_c$, i.e. the economy in the presence of the informal sector converges faster to steady state N^* . This implies that, as $N_0 - \varepsilon < N_0$, path $N_i(t)$ must cross path $N_n(t)$ from below at least once. Without the loss of generality suppose it happens at the moment of time t_c and not before. Then $N_n(t_c) = N_i(t_c) =$

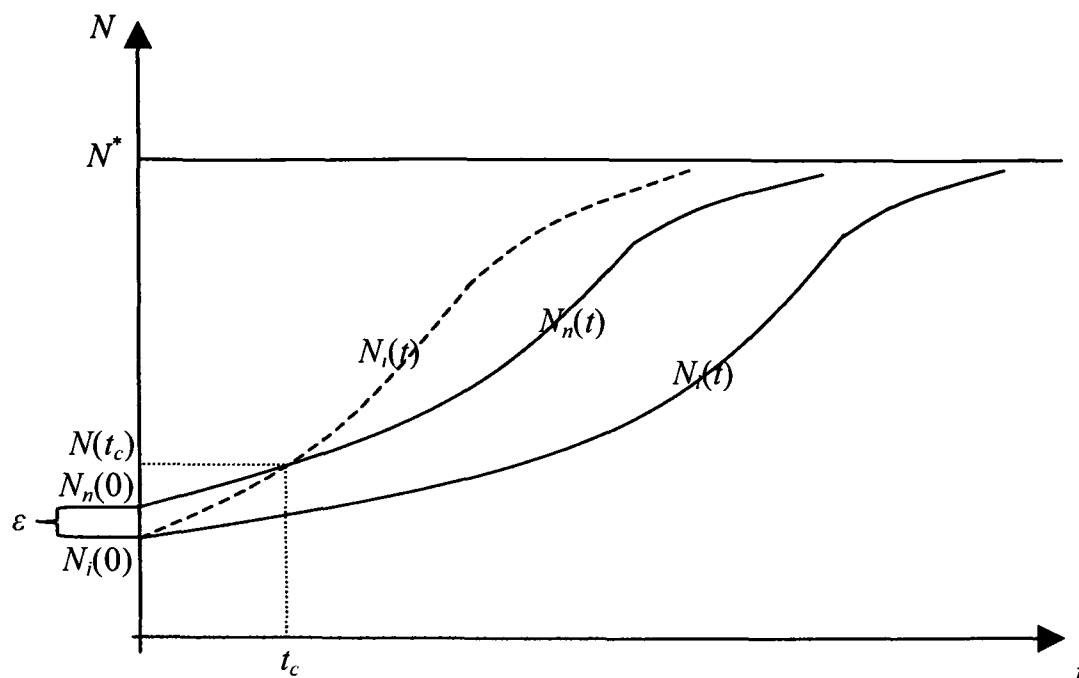


Figure 3.11: Timing of transition

$N(t_c)$, while the level of old sector employment is $O(t_c)$ (it is the same for both sets of initial conditions, as development of the old sector does not depend on N , U or I). Let $U_n(t_c) = U_n$ and $U_i(t_c) = U_i$, be the levels of unemployment, and $I_n(t_c) = 0$ and $I_i(t_c) = I_i$ be the levels of shadow employment for the two sets of initial conditions (note that the informal sector does not take off if $I_n(0) = 0$). From the consistency conditions evaluated at t_c we have that $O(t_c) + N(t_c) + U_n = O(t_c) + N(t_c) + U_i + I_i$, or, equivalently $U_n = U_i + I_i$. The fact that $N_i(t)$ crosses path $N_n(t)$ from below implies that $J_n(U_i, N(t_c), I_i) > J_n(U_n, N(t_c), 0)$. However, from $\frac{\partial J_n(\cdot)}{\partial U} > 0$, $\frac{\partial J_n(\cdot)}{\partial I} < 0$ and $U_n > U_i$, $I_i > 0$ we can conclude that $J_n(U_i, N(t_c), I_i) < J_n(U_n, N(t_c), 0)$. The contradiction proves that path $N_n(t)$ lies above path $N_i(t)$ for any t , and that the new sector converges to the steady state faster in the absence of its informal counterpart.

Chapter 4

State Regulations, Job Search and Wage Bargaining: A Study in the Economics of the Informal Sector

4.1 Introduction

An increase in the size of informal sectors all over the world has recently been the focus of a debate in many studies. The situation in OECD countries since 1960 has been analysed by Schneider (2000, 2001) and Schneider and Enste (2000) who point to the fact that for all countries investigated the informal economy has reached a remarkably large size. Other authors note that in most transitional countries of Eastern Europe (CEE) and the former Soviet Union (FSU) the irregular sectors have been growing over the last 15 years too (see, e.g. Johnson *et al.*, 1997; Lackó, 2000; Feige and Urban, 2003). In such countries as Georgia, Russia, and Ukraine an increase in the share of the informal sector

has been especially notable and its persistent character is clearly observed. As regards CEE and FSU countries, the primary motivation for this essay, it has been argued that the increase may well be a transitional feature en route to the market economy, prompted by an increase in unemployment at the start of economic reforms in the region (see Chapter 3). At the same time, long-run strengthening of informality should not be excluded. In the main the literature on the informal sector is yet to do much work in dotting the i's and crossing the t's as regards preconditions and mechanisms leading to stable coexistence of the formal and informal sectors in the longer term.

Generally, it is held that it is the burden of governmental regulations of various nature that forces firms and entrepreneurs to move underground. The ratio of reported to unreported activities depends largely on costs and benefits of operating in each economy (Kaufmann, 1997), which often are derivatives of governmental actions as can be seen from the discussion in, for example, De Soto (1989) and Loayza (1996). Schneider and Enste (2000) and Boeri and Garibaldi (2001) point, in particular, to the constraints on formal firms in labour markets - the fact that leads to an increase in size of the underground labour force (see Schneider, 2000, 2001).

According to Castells and Portes (1989), considerations of labour costs are among the most important factors forcing entrepreneurs to "go shadow" across the globe. Significant wage differentials between formal and informal sectors are a notable stylised fact (see Mønsted, 2000, and Gindling and Terrell, 2004, for evidence from some developing countries, while Kolev, 1998, and Roshchin and Razumova, 2002, report on the situation in Russia). Part of these differences is explained by minimum wage laws and productivity differentials.¹ However, in many FSU countries minimum wages are not binding, while

¹Productivity differentials are traditionally used in modelling of the formal-informal segmentation -

both formal and informal jobs can often coexist in the same enterprises. and workers can receive part of their salary in black cash - "under the table", so that the "productivity gap" explanation is not applicable. Such facts suggest that the interaction of firms and employees in the labour market is especially worthy of attention in addressing the questions of emergence and development of the irregular sector. Nonetheless, as noted in Kolm and Larsen (2004), the previous theoretical research on informal economies has been mainly conducted within the public finance tradition. In that literature labour markets are competitive, while wages are either assumed fixed or determined by market clearing. In such a framework the burden of regulations cannot cause formal-informal segmentation unless specific assumptions are made about preferences or risk attitudes of workers, heterogeneity of the labour force, built-in technological externalities, etc. Modelling of those aspects has received all the attention of researchers, while the issue of wage formation is effectively left out. As discussed later on in this work, such *ad hoc* assumptions are not always justified by evidence. However, dropping them would imply that economic agents just choose the sector where the effect of regulations is least onerous. Thus, in that literature a non-corner equilibrium with both sectors is effectively presupposed, while the role of labour costs in the formal-informal split is neglected.

Recently it has become popular to invoke the theory of search in the labour market to model the formal-informal duality (Kolm and Larsen, 2001, 2004; Boeri and Garibaldi, 2001; Bouev, 2002; Fugazza and Jacques, 2004). The focus of those studies has been mainly on the effect of various governmental policies on the size of the informal sector and the level of involuntary unemployment. The models suffer from a great number of parameters, and are often built around the same specific assumptions that were made in

for examples see Agénor and Aizenman (1999), Friedman *et al.* (2000), or Boeri and Garibaldi (2001).

earlier studies, which sometimes adds a lot to complication of the work. At the same time, in our opinion, they camouflage a rather simple mechanism that can make for the emergence of the long-term formal-informal split of the labour market, even when workers are assumed to be homogeneous, risk-neutral, and there are no presupposed technological externalities.

In this essay we look into the interaction of firms and workers in the non-Walrasian labour market to see a) if wage bargaining and search can be conducive to the emergence of informal labour markets in the long term, and b) where the government regulations blamed for being a main cause of informality fit in with this framework. In such labour markets productive matches of firms and workers are costly, they take time to accomplish, while wages are determined in bilateral negotiations. Following Loayza (1996) we distinguish two types of government regulations that affect the result of the bargaining and, hence, the equilibria in our model. First are the measures that impact on the costs of functioning in a particular sector. Such policies, as for example, taxes or social security contributions in the formal economy, and penalties for running business underground in the informal sector, determine the size of the surplus generated by a productive match and subject to sharing during wage bargains. In addition, auditing of informal firms by the government generates asymmetries in match duration across sectors, which affects the values of expected or averaged surpluses. Second are the activities of the low tier of the government, such as bureaucracy that, if corrupt, can through red tape, license fees, extortion of bribes, and so forth, erect artificial barriers to entry into the formal sector, and thus raise relative costs of access to legality (see, also, De Soto, 1989; Djankov *et al.*, 2002). In the presence of search frictions these venal practices increase opportunity costs of vacancy posting

for firms looking for workers in the labour market and weaken firms' outside option in wage negotiations. We show that as a consequence, when entry costs differ and/or match lifetimes are not the same in the two sectors, wage differentials can ensue in long-run equilibrium, thus leading to labour market segmentation. Search and rent sharing are very important for this result, because without them the system would produce only corner solutions. However, these features are inherent in labour markets of many countries and have been confirmed for Eastern Europe, in particular (see Smirnova, 2003b, and Roshchin and Markova, 2004, for evidence on time-consuming job search, and Grosfeld and Nivet, 1999, and Shakhnovich and Yudashkina, 2001, on rent sharing). Thus, it can be concluded that wage bargaining in the presence of costs of entry can be one of the main channels through which informality is brought about. Having said that we compare our result with the previous studies of the informal economies, stressing its independence of preferences of workers and other assumptions mentioned above. This work can certainly be extended to incorporate a great deal of those additional features which would not, however, diminish the role performed jointly by government regulations and wage bargaining in the presence of costly search in splitting the labour market.

Having described the workings of and equilibria in our model, we turn to consideration of its implications for policy making and welfare. The aforementioned studies of informality featuring the non-competitive labour market with job search and costly matching often attempt a normative analysis of policies aimed at the reduction in size of the shadow economy (see, e.g. Kolm and Larsen, 2001; Bouev, 2002; Fugazza and Jacques, 2004). However, they do not take into account a number of inefficiencies arising in the labour market that may well affect the conclusions of such exercises as regards welfare improving

measures. We show that labour market externalities arising in such environments should not be expected to be internalised. In general, no equilibrium is efficient in our model. One of the sources of welfare losses in this work is a version of Grout's (1984) hold-up problem whereby workers appropriate part of return on firms' start-up investment. It is shown that while the first-best solutions are not likely, a benevolent government can achieve sub-optimal allocations of resources. However, the upshot of standard policies, such as variation in the tax rate, efficiency of monitoring of the informal firms, and the penalty rate, depends upon the state of the labour market. In particular, the relation between the bargaining power of workers and the elasticity of the matching function prominently figured in the Hosios efficiency condition (Hosios, 1990) affects the ultimate effect on economic welfare. This point has been completely overlooked in the previous research.

The essay is organised as follows. The next section provides a quick overview of the previous theoretical literature on the informal sector, highlighting a few important soft spots, the main of which is the absence of a proper account of the labour market, especially the wage determination process. Then Section 4.3 introduces a two-sector search model à la Pissarides (2000), solves it by deriving steady state equilibria and discusses how state regulations lead to formal-informal wage differentials and, hence, labour market segmentation. Implications for policies and their welfare impact are discussed in Section 4.4. Section 4.5 concludes.

4.2 The Informal Sector:

A Glimpse of the Literature

There exists an extensive literature concentrating on various aspects of informality. For the most recent review the reader is referred to Gërxhani (2004), while the effects of regulations on the emergence and development of the informal sector both from theoretical and empirical perspectives are discussed in *inter alia* Kaufmann and Kaliberda (1996), Loayza (1996), Fortin *et al.* (1997), Johnson *et al.* (1997), Friedman *et al.* (2000), etc. The large body of previous theoretical research, however, has suffered from a few significant deficiencies, in our opinion. First, it does not make clear whether the informal sector can exist in the long run, i.e. whether or not it is just a short-run product of adjustment in the economy, after some sort of a shock has pulled it out of an equilibrium state. Second, in many both static and dynamic models an interior equilibrium with both formal and informal economies is often possible only due to a number of restrictive assumptions about the utility function of workers, penalties for concealing income, etc. Finally, as regards mechanisms whereby governmental regulations affect the segmentation of the economy, the literature has mainly ignored the fact that the decision to "go underground" is essentially a result of both employers and employees interacting in the labour market. We briefly discuss these points below.

4.2.1 Long-run Informality

Empirically the existence of the informal sector provokes no doubt, whereas theoretical substantiation of its existence, especially in the long run, has been not satisfactory. When

modelling informality researchers often restrict their attention to those *ad hoc* combinations of parameters alone that generate interior equilibria with an informal sector in their models (see, e.g. Kolm and Larsen, 2001). Their inattention to corner equilibria (which often are the most probable result), i.e. equilibria with formal or informal sectors alone, is understandable: on the one hand, an equilibrium with the formal sector alone does not allow analysis of informality, and, on the other hand, an equilibrium with the informal sector alone is not conceivable as a realistic long-run outcome.² At the same time, the reasons or possibilities for emergence of interior equilibria (let alone their stability) with both formal and informal sectors are not explained nor explored. However, the existence of the informal sector can be a transitional phenomenon of adjustment in the economy as shown, for example, in Chapter 3 of this thesis. The underground sector can be around for some time, even when an economy converges to a long-run steady state, where informality is not eventually present. Still, it begs the question of whether and when the informal sector can stably coexist with the formal one in the long term. The conditions for such coexistence, if any, are of great interest, in our opinion. A recent strand of endogenous growth literature allowing for the informal sector (e.g. Loayza, 1996; Sarte, 2000) has partially succeeded in showing that long-run mixed equilibria are indeed possible. Nevertheless, it either imposes *ad hoc* restrictions leading to the existence of such outcomes

²In general, it is possible to think of a number of reasons why governments may be interested in increasing the number of official firms. Shleifer and Vishny (1998), for example, suggest that such reasons may emanate from properly organised fiscal systems, politicians' desire to win greater support for elections, or direct financial interests of politicians (shareholding). On the other hand, while the state itself can have stakes in enterprises, the firms can repeatedly interact with public officials in their own turn. Such interaction may result from historical ability of some firms to influence the government so that they enjoy considerable private gains. Other, *de novo* firms can engage in attempts to capture the state, i.e. make private payments to state officials to affect the rules of the game as a strategy to compete with influential incumbents. In other words, powerful firms can collude with state authorities to extract rents through manipulation of state power (Hellman *et al.*, 2000). Thus, all this suggests that in general the conditions leading to the emergence of corner equilibria with the informal sector alone in our model should be considered as implausible.

(Loayza, 1996, models the effective penalty rate for producing informally as an endogenous function of the relative size of the informal sector), or is still lacking in a proper account of the labour market (Sarte, 2000).

4.2.2 Restrictive Assumptions

Dependence of the interior equilibrium on specific assumptions has also characterised other, less recent, branches of the theoretical literature concerned with informality. For instance, in tax evasion studies (for recent reviews see, e.g. Andreoni *et al.*, 1998, or Slemrod and Yitzhaki, 2002), evasion (and hence, the existence of underground activities) arises in a gamble where a risk-averse tax-payer trades off the utility from tax savings and disutility of extra risk taken on of having her income understatement detected by the authorities and penalised. The seminal Allingham and Sandmo (1972) tax evasion model, for example, predicts that in a situation where individuals are risk-neutral only corner solutions are possible - an individual would either do no evasion or remit no tax at all. In the work on unrecorded activity emanating from Allingham and Sandmo (1972) the equilibrium with coexisting recorded and unrecorded activities is possible only under certain assumptions about the utility function, i.e., in particular, risk aversion of an individual.

In a similar vein, static models of labour supply to the formal sector and the underground economy are often based on restrictive assumptions about the utility function. This may include imperfect substitutability of output from the compliant and evading sectors, heterogeneity of workers in evasion costs (Kesselman, 1989) or skill levels (Sandmo, 1981). Interior equilibria in models with home production or moonlighting (see, e.g. Becker, 1965; Gronau, 1977) are also a product of the choice of a specific utility function,

namely preferences over consumption, work in a particular sector and leisure.

A specific choice of other functions, such as, for example, a probability of detection, that can be made an endogenous function of the amount of unrecorded activities (Slemrod and Yitzhaki, 2002), have both characterised interior equilibria in the tax evasion literature mentioned above, and featured in more recent work on underground economies (see, again, Loayza, 1996). The main problem with this approach, however, as well as with the one where an interior equilibrium hinges upon specific non-economic costs of evasion - moral considerations (Kolm and Larsen, 2001) or psychic costs (Fugazza and Jacques, 2004) - is that viability and implications of such analyses depend on the precise way that the concepts are formalised (Andreoni *et al.*, 1998; Slemrod and Yitzhaki, 2002).

All in all, although the contribution of the literature briefly considered here is undisputable, especially because it has provided a useful framework for an investigation into the effect of various governmental policies on the relative size and growth of the informal sector, the assumptions made there do not always stand up to the evidence. For example, as regards different preferences over formal and informal output, it can be noted that some goods are produced in both the formal and informal sectors, or/and individuals may have no clear idea if the supplier is operating in the formal or the irregular sector (see, e.g. Thomas, 1992, Ch.8). Even when it is claimed that in countries of, for example, Western Europe the informal sector is concentrated within particular industries, such as services or construction, so that, on a large scale, different goods are produced in the formal and informal sectors (Kolm and Larsen, 2004), in countries of Eastern Europe, and, particularly Russia, this assumption may not be correct as the practice of informal contracts is often widespread (Ingster, 2003). In relation to the dependence of an interior equilibrium

with formal and underground activities on the existence of moral or social considerations it should be stressed that, although there is little dispute that those factors are important in individual compliance decisions, little is agreed upon on how best to incorporate these effects in a theoretical analysis (Andreoni *et al.*, 1998).

4.2.3 A Need for Labour Markets

Having said that, another substantial weakness in theory of the informal sector is still its lack of proper attention to labour markets. Empirical facts such as a drop in participation rates (for a discussion of the situation in Eastern Europe see Boeri, 2000b), widespread informal (not registered) contracts (e.g. Haltiwanger and Vodopivec, 2002, and Ingster, 2003, mention such practices in Estonia and Russia, respectively), significant formal-informal wage differentials (for Russian experience see Kolev, 1998; Roshchin and Razumova, 2002), beg for more research to be done in the area. Mention by Castells and Portes (1989) of labour costs as one of the key factors causing informality points to a special interest that should be attracted to revealing the role played by wages in propagating the effects of governmental regulations and their impact on formal-informal segmentation. However, as has been noted in the introduction, in the large body of the previous work on informality wages are either treated as exogenous or assumed to be determined through market clearing, i.e. no proper theoretical foundation has as yet been established in regard to that role.

A few recent studies (Boeri and Garibaldi, 2001; Kolm and Larsen, 2001, 2004; Bouev, 2002; Fugazza and Jacques, 2004) have made an attempt to incorporate the theory of search and matching functions into the models with the informal sector. The focus of that

literature is implications for policies aimed at the reduction in informality and involuntary unemployment. At the same time, they provide a hint that wage bargaining in the presence of costly search in the labour market may serve as an important channel through which preconditions for the formal-informal duality emerge.

In the next section we present a model of the informal sector which serves to illustrate three important moments either not clearly stated or absent completely in the previous studies of the informal sector. First, it shows that under a broad set of conditions the long-run equilibrium with both the formal and informal sectors is possible. Second, it highlights the role of the wage determination process in shaping the equilibrium outcome. Finally, all the results are obtained in the absence of many restrictive assumptions characterising much preceding work.

4.3 A Model of Informal Employment

The model developed in this section captures the impact of governmental regulations and labour market institutions, such as wage bargaining, on sectoral reallocation of jobs and workers as well as wage rates in an economy with the formal and informal sectors. It is assumed that the labour market in such an economy is characterised by risk-neutral firms and workers searching for each other to form a match to start production. Search and rent sharing in the process of wage bargains are crucial to the results we obtain. The approach is similar to that used by Acemoglu (2001) who studied reallocation of labour across jobs with different capital costs. In this work we abstract from goods and capital markets (both of which are assumed to clear) in order to highlight the joint effect of state regulations, search frictions, and rent sharing on job composition, rather than on prices

of both capital input and final output.³

4.3.1 The Main Idea

The informal sector is seen as representing productive (not rent-seeking⁴) activities that are not associated with crime or household production.^{5,6} Thus, we take the approach that views informal employment as resulting from efforts of entrepreneurs to trade off costs and benefits of functioning in compliance with formal regulations.

It is assumed that goods produced both in the formal and informal sectors are perfect substitutes, while marginal productivity of formal and informal matches is the same. We do not go along the lines of the prevalent view of the informal economy (see, for example, Agénor and Aizenman, 1999; Boeri and Garibaldi, 2001) that assumes underground jobs to be less productive and, hence, paying lower wages. We shall see that what is important for the conclusions that we draw is not the differentials in productivity but the differences in surpluses that formal and informal matches generate.

Both formal and informal firms have to sink some costs before opening a vacancy, meeting a worker, and starting production. Those can be capital costs, vacancy advertisement costs, or bribes and other extortionary payments that firms have to bear before

³See Kolm and Larsen (2004) for a general equilibrium model with wage bargaining and costly search where prices absorb part of the effect of governmental policies.

⁴Acemoglu (1995) and Acemoglu and Verdier (1998) study the allocation of talent between productive and rent-seeking activities. Vostroknutova (2003) extends their models to include an underground sector.

⁵In the literature on informal activities it is normal to distinguish between household activities, the informal sector, the irregular sector and the criminal sector (see, for example, Thomas, 1992). While the idea behind home production and criminal activities should be obvious, one may become confused over the difference between the informal and irregular sectors. Usually it is the small workshops and self-employment which are regarded as the informal sector. It can also comprise home production that is traded in the market. All these activities are not illegal. The sector that we consider in this model is indeed irregular which comprises production of legal output but involves tax evasion and avoidance of formal regulations. However, we will use both terms "irregular" and "informal" interchangeably. Other synonyms used here are the "shadow" or "underground" economy.

⁶For models of crime see *inter alia* Becker (1968), and Fiorentini and Peltzman, eds. (1995); for household production see Becker (1965), and Gronau (1977).

starting their businesses. We assume that these costs are greater in the formal sector than in the informal one, which can be explained by higher entrance barriers into the formal sector or access costs to legality (Loayza, 1996) associated with bribery, license fees and registration requirements (De Soto, 1989; Djankov *et al.*, 2002). In the appendix we muse on departures from this set-up. Another main conceptual difference between the two sectors relates to the effect of official regulations and costs associated with them. On the one hand, firms producing formally and, hence, abiding by the rules and regulations imposed by the state face additional costs of production such as, for example, taxes, social security contributions, etc. (in what follows we refer to all such costs as "taxes" to keep things simple). On the other hand, functioning informally does not involve those expenses. Jobs can be undeclared in order to avoid costs of functioning openly. Although such concealment of production is possible, it is prosecuted by officials. Thus, each hiding firm faces some positive probability of being caught, fined, and closed as a result of monitoring or audit by the government. This, in turn, implies that informal matches on average last for a shorter time.

Workers in the model can either work formally or informally or be unemployed. We neglect possibilities of moonlighting, so workers can perform only one activity at a time. Aggregate labour supply is inelastic.

Once having met, workers and firms bargain over wages and, as a result, employees can appropriate some rents. Given different entrance and production costs and varying average match duration across sectors, rent sharing leads to equilibrium wage differentials. In turn, different labour costs and different production surpluses in the two sectors provide an opportunity for the formal and informal sectors to coexist in the long run. The equilibrium

allocation of jobs and workers in steady state is eventually determined by zero profit conditions as free entry in each sector is assumed.

4.3.2 Matching Technology

In the absence of on-the-job search it is only the unemployed workers who look for jobs. We assume that search is random or undirected, i.e. workers search for any employment and accept the first job that offers them the prospects at least as good as their currently expected life-time income. In the presence of undirected search both formal and informal vacancies have the same probability of meeting workers. Then it is the total number of vacancies that enters the matching function.

The number of job matches is given by $M(n, v)$, where n is the number of workers seeking jobs (i.e. the number of the unemployed) and v is the number of vacancies created in the economy.

With constant returns to matching the instantaneous probability that a vacant job meets a job-seeker is given by

$$\frac{M(n, v)}{v} = M\left(\frac{n}{v}, 1\right) = q(\theta),$$

where $\theta \equiv \frac{v}{n}$.

The first derivative of the flow rate of matching for a vacancy, $q'(\theta)$, is negative, because the greater is the value of θ the more difficult for firms to fill the job. In the matching literature θ is referred to as market tightness from the firms' standpoint (see, for example, Pissarides, 2000).

Similarly, the flow rate of matching for an unemployed worker is given by

$$\frac{M(n, v)}{n} = M\left(1, \frac{v}{n}\right) = \alpha(\theta) = \theta q(\theta),$$

where $\alpha'(\theta) > 0$.

When $q(\theta) < \infty$ and $\alpha(\theta) < \infty$ then matching is not instantaneous and takes some time. We will also make the additional Inada-type assumptions that $\lim_{\theta \rightarrow \infty} q(\theta) = 0$, $\lim_{\theta \rightarrow 0} q(\theta) = \infty$, $\lim_{\theta \rightarrow \infty} \alpha(\theta) = \infty$, and $\lim_{\theta \rightarrow 0} \alpha(\theta) = 0$.

4.3.3 Formal and Informal Jobs

Jobs are created in either the formal or the informal sector. We do not necessarily define one job as one firm by assuming constant returns in production. Before opening a vacancy a risk-neutral firm has to decide in which sector the potential match will produce and, at this point, will have to bear some costs. These costs are either k_f or k_i , if the firm is to open a vacancy in the formal economy or underground, respectively. These start-up costs are incurred before the firm meets its employees. They can be thought of as capital expenditure, job advertisement costs as well as a registration fee or bribes to be paid, for example, in order to prevent a delay in registration in the formal sector or to guarantee security of the job in the informal sector.⁷ The important assumption that we make is that access to legality is more costly than access to informal production, i.e. $k_f > k_i$. In other words, we postulate that the presence of extortion costs at the moment of entry in the formal sector implies higher instant start-up costs to entrepreneurs (see De Soto, 1989, and Loayza, 1996, for justification of this assumption). The latter are generally thought of as being wealthy enough to meet the start-up costs without resorting to external credit.

⁷Görxhani (2004) points out that, although the ease of entry is used by various researchers as one of the criteria for defining the informal sector, entry costs into informality do exist.

All matches in the economy, either formal or informal, die at rate δ , in which case the job is destroyed while the worker becomes unemployed.⁸

Both formal and informal jobs are equally productive. Wages are paid out of the match product, y . In addition to wages, formal jobs have to pay a lump sum tax, τ , whereas informal jobs enjoy tax evasion.

It is implicitly assumed that in the model there are some taxation authorities, such as the tax police, whose aim is to collect taxes and reveal cases of tax evasion. So, there is an exogenous flow probability, m , that an employer gets caught in engaging in underground business and fined by amount F . When m strikes the informal match is liquidated and the burden of the fine is borne by the employer, not the employee. An alternative to match liquidation may be its continuation or transformation into a formal match. However, if detected parties fear that continuing the match either formally or informally would result in more frequent visits by the tax police, our assumption of match destruction is reasonable^{9,10} (see, e.g. Kolm and Larsen, 2004).

The Bellman equation¹¹ for a formal job is

⁸Alternatively, one can consider a situation when δ (or m - see below) strikes, the match is destroyed but the job is not. That is, the job turns into a vacancy rather than is liquidated. However, such alteration does not change the qualitative results of the analysis.

⁹Here we also exclude the possibility that firms can avoid penalties and liquidation by paying a bribe to the tax inspector. However, in reality the agents directly carrying out monitoring may often side-contract with firms, thus allowing the latter to evade payment of fines (see, e.g., Chander and Wilde, 1992, and Wane, 2000, for the models of collusion between tax inspectors and tax evaders).

¹⁰Safavian *et al.* (2001) note that visits of firms by the tax police are closely linked to corruption - regulatory inspections are positively correlated with the amount of bribes paid. Interestingly, tax authorities can often change a regulation without notifying entrepreneurs, and then pay them a visit to obtain a fine or extorting a bribe for avoidance of restrictions implied by the regulation. Evidence suggests, however, that, e.g. in Russia, firms with higher reservation profits (i.e., the revenues that allow them to function just without making losses) are less likely to be charged excessive bribe payments and, hence, less likely to be checked by monitoring bodies. Above we have assumed that k_f is higher than k_i , which in turn implies higher reservation revenues in the formal sector. Thus, in our model the absence of monitoring and fines in the formal sector can be justified not only by the nature of official functioning: it can also be interpreted in the light of the results obtained by Safavian *et al.*

¹¹Hereafter we consider only steady state values of the Bellman equations since the focus of the chapter is the irregular sector in the long run. Out of steady states each Bellman equation should be augmented to include a first time derivative of an appropriate value function.

$$rJ_f = y - w_f - \tau + \delta(0 - J_f), \quad (4.1)$$

where r is the flow rate of return on having the job filled (the interest or discount rate), J_f is the value of the filled formal job to the employer, and w_f is the formal wage. The equation reads that the return to the firm on a filled job in the formal sector is equal to the difference between worker's productivity and costs, plus a potential change in value in the case of the match break-up. At this stage we shall assume very generally that the productivity of a match, y , is high enough to cover wages and taxes, while more exact restrictions on parameters of the model are given in Section 4.3.6.

By analogy, for an informal job we have

$$rJ_i = y - w_i - mF + (\delta + m)(0 - J_i), \quad (4.2)$$

where J_i is the value of the filled informal job and w_i is the informal wage. The equation implies that the return on a filled job in the informal sector is equal to the difference between the product of the match and the worker's wage, less an expected fine in the case of being caught by tax authorities, plus a change in value due to match cessation.

It is assumed that vacancy maintenance in either the formal or the irregular sector involves no flow costs.¹² Then the Bellman equations for vacancies in the formal and informal sectors are:

$$rV_f = q(\theta)(J_f - V_f), \quad (4.3)$$

¹²It can readily be shown that the presence of maintenance costs does not change the model qualitatively.

$$rV_i = q(\theta)(J_i - V_i), \quad (4.4)$$

where $q(\theta)$ is the flow rate of filling a vacancy as defined above.

4.3.4 Workers

There is a fixed (normalised to 1, for convenience) mass of identical workers in the economy. They are risk-neutral, have the same discount rate, r , as firms, and derive utility solely from the wage. Workers can be either employed in one of the sectors or unemployed.

Formal employment provides workers with wage w_f , so that the value of working formally satisfies the Bellman equation

$$rE_f = w_f + \delta(E_u - E_f).^{13} \quad (4.5)$$

It reads that the return on formal employment is equal to the wage income plus a change in unemployment in the case of the match break-up.

As informal employment brings in wage w_i , by analogy we have

$$rE_i = w_i + (\delta + m)(E_u - E_i). \quad (4.6)$$

That is, the return on informal employment is equal to the wage income plus a potential change into unemployment as a result of either match cessation or job closure due to tax evasion detected by the authorities.

¹³In order to keep things simple we neglect the impact of income taxes on the value of being employed in the formal sector. At the same time there is evidence (e.g., Lemieux *et al.*, 1994) on unimportance of taxes for sector choice.

Finally, the Bellman equation for the unemployed is

$$rE_u = b_u + \alpha(\theta)(\phi(E_f - E_u) + (1 - \phi)(E_i - E_u)), \quad (4.7)$$

where b_u is the unemployment benefit, ϕ is the probability of meeting a formal vacancy, $0 < \phi < 1$, and $\alpha(\theta)$ is the flow rate of finding a job in either sector. The equation reads that the return on being unemployed equals unemployment compensation plus a potential change into employment in one of the sectors.

4.3.5 Wage Determination

Wages in the model are determined through a wage bargaining process with the bargaining power of workers, β , given exogenously and such that $0 < \beta < 1$. Then the Nash (1950) bargaining solution implies:

$$(1 - \beta)(E_f - E_u) = \beta(J_f - V_f), \quad (4.8)$$

$$(1 - \beta)(E_i - E_u) = \beta(J_i - V_i). \quad (4.9)$$

The Nash solution in this case assumes that the threat (reservation) points of employers and employees are represented by the value of an unfilled vacancy in an appropriate sector and the value of unemployment, respectively. This implies that bargaining is actually *ex post*, i.e. it takes place before the consummation of a match, but after a producer has opened a vacancy. Thus, firms are assumed to commit to wages over which the consensus was reached: they cannot change the contract once a worker gets employed.

4.3.6 Steady State Equilibria

It is assumed that entry in our economy is free, so that firms' profits have to equal zero in equilibrium. This implies that it should not be possible for an additional vacancy in both the formal and informal sectors to open and make expected net profits. Hence,

$$V_f = k_f, \quad (4.10)$$

$$V_i = k_i. \quad (4.11)$$

That is, the condition of zero profits implies that start-up costs equal to k_f and k_i in the formal and informal sectors, respectively, must be just recouped in equilibrium.

A *steady state* equilibrium in the model is characterised by the labour market tightness, θ , a proportion of formal vacancies, ϕ , and by value functions J_f , J_i , V_f , V_i , E_f , E_i , and E_u , such that equations (4.1)-(4.11) are all simultaneously satisfied. As we have assumed undirected search, in steady state both formal and informal vacancies meet workers at the same rate and both types of job are accepted if they offer a reward at least as large as a worker's outside option.

In order to see what equilibrium allocations of jobs and workers are possible in our economy we shall proceed through the analysis by reexpressing equations (4.10) and (4.11) as functions of θ and ϕ , and then studying their behaviour in the (θ, ϕ) -plane.

Zero profit conditions

By solving (4.1) and (4.2) for J_f and J_i we arrive at

$$J_f = \frac{y - w_f - \tau}{\pi}, \quad (4.12)$$

$$J_i = \frac{y - w_i - mF}{\rho}, \quad (4.13)$$

where $\pi = r + \delta$ and $\rho = r + \delta + m$ are the effective discount rates in the formal and informal sectors, respectively. The rates account both for the interest rate r (equal to the workers' and firms' rate of time preference under risk neutrality) and "depreciation", δ or $\delta + m$, which differs across the two sectors.

By substituting these solutions together with conditions (4.10) and (4.11) for (4.8) and (4.9), and by combining the results with equations (4.5) and (4.6), simple algebra gives

$$w_f = \beta (S_f + b_u) + (1 - \beta)rE_u, \quad (4.14)$$

$$w_i = \beta (S_i + b_u) + (1 - \beta)rE_u. \quad (4.15)$$

For readability of formulae in (4.14) and (4.15), and in the rest of the paper by $S_f = y - \tau - \pi k_f - b_u$ and $S_i = y - mF - \rho k_i - b_u$ we denote the total flow surpluses of a match net of unemployment benefits in the formal and informal economies, respectively. These equations imply that the worker gets share β of the surplus of a match plus $(1 - \beta)$ times her outside option.

Having obtained the expressions for w_f and w_i , by using equations (4.12) and (4.13) together with (4.3) and (4.4), we can define two functions $\Pi_f(\theta, \phi)$ and $\Pi_i(\theta, \phi)$ that

represent profits made in the formal and informal sectors, respectively:

$$\Pi_f(\theta, \phi) = V_f - k_f = \frac{(1 - \beta) q(\theta)}{(r + q(\theta)) \pi} \left(y - \tau - \pi k_f - \frac{r \pi k_f}{q(\theta)} - \frac{\beta r \pi k_f}{(1 - \beta) q(\theta)} - r E_u \right), \quad (4.16)$$

$$\Pi_i(\theta, \phi) = V_i - k_i = \frac{(1 - \beta) q(\theta)}{(r + q(\theta)) \rho} \left(y - mF - \rho k_i - \frac{r \rho k_i}{q(\theta)} - \frac{\beta r \rho k_i}{(1 - \beta) q(\theta)} - r E_u \right). \quad (4.17)$$

From (4.7) it also follows that $E_u = E_u(\theta, \phi)$, i.e. the value of being unemployed is also a function of the market tightness, θ , and the proportion of formal vacancies, ϕ .

Complicated as they are at first sight, expressions (4.16) and (4.17) above allow, in fact, an easy interpretation. The terms in brackets times either $\frac{(1-\beta)q(\theta)}{(r+q(\theta))\pi}$ or $\frac{q(\theta)}{(r+q(\theta))\rho}$ give the expected rents that formal and informal matches will generate when a firm and a worker meet. These rents are shared between the two parties in a bilateral monopoly bargaining game (see, e.g. Shaked and Sutton, 1984), so that in the end the firm gets share $(1 - \beta)$ of the rent according to its bargaining power. Consider for example formal profit (4.16). After consummation the match generates product y . Out of this product the firm has to: a) pay off taxes, τ ; b) cover start-up costs (taking account of "depreciation"), πk_f ; c) cover opportunity costs of having k_f units of resources invested in creation of this particular vacancy, $\frac{r \pi k_f}{q(\theta)}$ (i.e. the vacancy that on average costs πk_f , will be idle until it meets a worker after an average time of search, $\frac{1}{q(\theta)}$, elapses - all this can be invested elsewhere at rate r); d) pay a premium to a hired worker for saving of opportunity costs that the representative firm enjoys when a job is formed, $\frac{\beta r \pi k_f}{(1-\beta)q(\theta)}$ (for a similar intuition see, e.g. Pissarides, 2000, p.17); and, finally, e) the firm has to compensate the worker for her outside option $r E_u$. The remaining surplus is split between the firm and the worker

according to their bargaining powers given by $(1 - \beta)$ and β , respectively. In particular, in the case of $\theta \rightarrow 0$, i.e. $q(\theta) \rightarrow \infty$ (so that firms have no problem finding a match, which is effectively a Walrasian labour market from the firms' standpoint) and when firms expect to make positive profits the expression above is reduced to $(1 - \beta) \frac{(y - \tau - \pi k_f - r E_u(0, \phi))}{\pi}$. That is, firm's (averaged) profits are given by share $(1 - \beta)$ of the expected surplus, while workers capture share β of the expected surplus in addition to being paid their outside option $r E_u(0, \phi)$. In contrast to the case when θ is positive, when $\theta = 0$ workers are not compensated for saving of opportunity costs as matching is instant for firms. If however, $\theta \rightarrow \infty$, i.e. $q(\theta) \rightarrow 0$ (so that workers find a match instantly, while firms on average wait infinitely long), by using the properties of function $E_u(\theta, \phi)$ that are studied below, it is possible to show that formal profits are reduced to $-k_f$. In this case all the match rents are appropriated by workers in the process of bargaining, while firms gain nothing and should not expect to recover even their start-up costs k_f . The expression for informal profit (4.17) can be analysed by analogy.

From (4.16) and (4.17) zero profit conditions (4.10) and (4.11) can be reexpressed as $\Pi_f(\theta, \phi) = 0$ and $\Pi_i(\theta, \phi) = 0$, or, as in general $\frac{(1-\beta)q(\theta)}{(r+q(\theta))\pi} > 0$ and $\frac{(1-\beta)q(\theta)}{(r+q(\theta))\rho} > 0$,

$$y - \tau - \pi k_f - \frac{r\pi k_f}{q(\theta)} - \frac{\beta r\pi k_f}{(1 - \beta)q(\theta)} - r E_u(\theta, \phi) = 0, \quad (4.18)$$

$$y - mF - \rho k_i - \frac{r\rho k_i}{q(\theta)} - \frac{\beta r\rho k_i}{(1 - \beta)q(\theta)} - r E_u(\theta, \phi) = 0. \quad (4.19)$$

Each of equations (4.18) and (4.19) defines θ as a function of ϕ and parameters of the model k_f , k_i , β , r , δ , b_u , τ , m , and F .

To close the circle we now need to analyse properties of $E_u(\theta, \phi)$.

The value of being unemployed

The value of being unemployed follows from (4.7) and equals

$$E_u(\theta, \phi) = \frac{b_u \pi \rho + \alpha(\theta) \beta (\phi \rho (S_f + b_u) + (1 - \phi) \pi (S_i + b_u))}{r (\alpha(\theta) \beta ((1 - \phi) \pi + \phi \rho) + \pi \rho)}. \quad (4.20)$$

For the function $E_u(\theta, \phi)$ it can easily be verified that it is continuous and bounded by $\frac{b_u}{r}$ from below and by $\frac{1}{r} (\max(S_f, S_i) + b_u)$ from above. Also, it is strictly increasing in θ provided that y is big enough.¹⁴ The intuition behind this result is straightforward: the value of being unemployed is increasing in market tightness because it becomes easier to find a job. In contrast, without additional assumptions about the parameters of the model $E_u(\theta, \phi)$ cannot be shown to be increasing or decreasing in ϕ everywhere. The sign of derivative $\frac{\partial E_u(\theta, \phi)}{\partial \phi}$ hinges upon the relative value of employment in the formal and informal sectors, E_f and E_i , respectively. In particular, whenever E_f is greater than E_i , derivative $\frac{\partial E_u(\theta, \phi)}{\partial \phi}$ is positive, and it is negative otherwise. This result implies that the value of being unemployed rises whenever does the proportion of vacancies posted in the sector where the value of employment is higher.

The relative value of E_f and E_i depends on various combinations of the model's parameters and may also depend on the level of market tightness. A formal analysis in Appendix 4.A shows that the variety of parameter combinations is effectively reduced to, on the one hand, the relation between the values of sector flow surpluses,¹⁵ S_f and S_i , and,

¹⁴In particular, to guarantee $\frac{\partial E_u(\theta, \phi)}{\partial \theta} > 0$ we must reasonably claim that at least $y > \tau + k\pi + b_u$, or, equivalently, $S_f > 0$. That is, the product of a match, y , must be at least as large as to pay all the flow costs of functioning in the formal sector and the wage equal to the reservation value, b_u . This condition also implies that $E_f > E_u$ holds. Analogously, to insure $E_i > E_u$, we must guarantee $y > mF + k\rho + b_u$ or $S_i > 0$. Otherwise, whenever any of these conditions is not met, an appropriate sector simply does not exist.

¹⁵Although for convenience we refer to S_f and S_i as surpluses, in fact they are the sector surpluses *net* of unemployment benefits as has been mentioned when the notation was introduced. It is the values on top of b_u that firms and workers worry about when they compare attractiveness of either sector. This

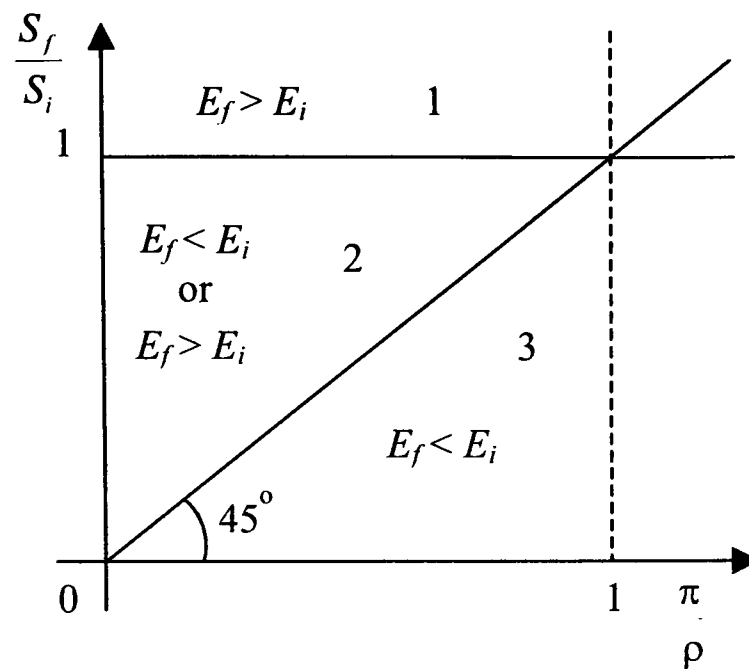


Figure 4.1: Dependence of employment values on model's parameters

on the other hand, the relation between their discounted or expected values, $\frac{S_f}{\pi}$ and $\frac{S_i}{\rho}$. Thus, all possible situations can be graphically represented in the $\left(\frac{S_f}{S_i}, \frac{\pi}{\rho}\right)$ -plane. Fig.4.1 illustrates the cases,¹⁶ while here we provide their intuitive explanation.

Region 1 is restricted by the horizontal line $\frac{S_f}{S_i} = 1$ from below and vertical lines $\frac{\pi}{\rho} = 0$ and $\frac{\pi}{\rho} = 1$ from the left and the right side, respectively. In this region the total surplus of a formal match, S_f , is greater than the total informal surplus, S_i , while the expected value of the former, $\frac{S_f}{\pi}$, is greater than the expected value of the latter, $\frac{S_i}{\rho}$. This is enough to guarantee that $E_f > E_i$. To see why this is so, let us compare two extreme situations.

First, consider a worker having two offers from formal and informal firms and knowing that whatever offer she accepts she will never lose a job thereafter. For such a worker formal employment brings in a higher value than informal employment if and only if $S_f > S_i$ (from the Nash bargaining solution it follows that her wage in either sector is a

is explained by the fact that a match in either sector has to pay workers at least the unemployment compensation so that they do not prefer to stay unemployed.

¹⁶Note that by assumption $\rho > \pi$, so all possible cases are situated to the left of the vertical line $\frac{\pi}{\rho} = 1$ in Fig.4.1 (shaded areas).

constant fraction of the match surplus in the sector). This situation can also be viewed as corresponding to the case when market tightness is infinite, so that offers from both sectors arrive instantly. As a second extreme, let us pick a worker who has two offers from the two sectors, but knows that both matches dissolve soon, and after that she will stay unemployed forever. For such a worker the value of formal employment is greater than the value of informal employment if and only if the *expected* surpluses satisfy $\frac{S_f}{\pi} > \frac{S_i}{\rho}$. This case can also be seen as reflecting a situation when market tightness is equal to nil, so that these two offers are the last chance for the employee. In such a world the discounting is used to take account of different match duration across the two sectors captured by the rates π and ρ . It should be intuitively clear that all other possible cases lie in between these two the "best" and the "worst" scenarios.

Thus, if both $S_f > S_i$ and $\frac{S_f}{\pi} > \frac{S_i}{\rho}$ (as in region 1, Fig.4.1) then for both the "best" and the "worst" scenarios formal employment is better, and hence, unambiguously $E_f > E_i$ and $\frac{\partial E_u(\theta, \phi)}{\partial \phi} > 0$.

By analogy, whenever $S_f < S_i$ and $\frac{S_f}{\pi} < \frac{S_i}{\rho}$ (as in region 3, which is bordered by the 45-degree line, and the lines $\frac{S_f}{S_i} = 0$ and $\frac{\pi}{\rho} = 1$ in Fig.4.1), then $E_f < E_i$ as a worker in both the "best" and the "worst" situations prefers informal employment. Hence, $\frac{\partial E_u(\theta, \phi)}{\partial \phi} < 0$ in this case.

Finally, when $S_f < S_i$ but $\frac{S_f}{\pi} > \frac{S_i}{\rho}$ (region 2, which lies in between regions 1 and 3 in Fig.4.1) the worker has different preferences depending on circumstances. In particular, there must exist a threshold level of the market tightness that separates the effects of the two scenarios on the total value of formal and informal employment. Specifically, it can be shown that as $\theta \rightarrow 0$ (no offers are available, infinite duration of search) $E_f > E_i$

as the effect of the "worst" scenario, $\frac{S_f}{\pi} > \frac{S_i}{\rho}$, dominates. However, as $\theta \rightarrow \infty$ (instant reemployment) $E_f < E_i$ as the effect of the "best" scenario, $S_f < S_i$, dominates the other one. The proposition below extends the result on the sign of derivative $\frac{\partial E_u(\theta, \phi)}{\partial \phi}$ (its proof is relegated to the appendix).

Proposition 1 *There exists some threshold value of the market tightness $\bar{\theta}$, defined by parameters $k_f, k_i, \beta, \tau, \delta, b_u, \tau, m, F$, and parameters of the matching function, such that for parameter values satisfying conditions in region 2, Fig.4.1, and for any $\theta > \bar{\theta}$ derivative $\frac{\partial E_u(\theta, \phi)}{\partial \phi}$ is negative, and for any $\theta < \bar{\theta}$ it is positive.¹⁷*

Having established the properties of function $E_u(\theta, \phi)$ we command all the knowledge necessary to derive and study equilibria in our model.

Finding equilibria

If there exists an equilibrium with both formal and informal jobs then both formal and informal profits must be equal to zero at the equilibrium point. That is, equations (4.18) and (4.19) are simultaneously satisfied. Alternatively, there can exist equilibria with only one type of job. In that case, profits in one of the sectors would be negative and only one of equations (4.18) and (4.19) would hold.

Two loci Two zero profit conditions (4.18) and (4.19) define two loci of formal and informal jobs in the (θ, ϕ) -plane. Both must be evaluated with the expression for $E_u(\theta, \phi)$ (4.20) substituted in.

By using simple algebra and invoking the implicit function theorem it can easily be verified that the locus of formal jobs (4.18) has a slope

¹⁷In fact this value of the market tightness, $\bar{\theta}$, is a bifurcation point that separates two regions with different dynamics of our economy. We reflect on this issue at more length in the appendix.

$$\frac{\partial \theta}{\partial \phi}|_f = \frac{\frac{\partial E_u(\theta, \phi)}{\partial \phi}}{\frac{\partial q(\theta)}{\partial \theta} \frac{\pi k_f}{(1-\beta)q^2(\theta)} - \frac{\partial E_u(\theta, \phi)}{\partial \theta}}. \quad (4.21)$$

Since $\frac{\partial E_u(\theta, \phi)}{\partial \theta}$ is always positive, whereas $\frac{\partial q(\theta)}{\partial \theta} < 0$, it is obvious that the denominator is negative. Then the slope of the locus of formal jobs has a sign opposite to the sign of $\frac{\partial E_u(\theta, \phi)}{\partial \phi}$.

Analogously, the slope of the locus of informal jobs (4.19) in the (θ, ϕ) -plane is

$$\frac{\partial \theta}{\partial \phi}|_i = \frac{\frac{\partial E_u(\theta, \phi)}{\partial \phi}}{\frac{\partial q(\theta)}{\partial \theta} \frac{\rho k_i}{(1-\beta)q^2(\theta)} - \frac{\partial E_u(\theta, \phi)}{\partial \theta}}, \quad (4.22)$$

which, by the same token, has a sign opposite to the sign of $\frac{\partial E_u(\theta, \phi)}{\partial \phi}$.

Remark 1 *Both the locus of formal jobs (4.18) and the locus of informal jobs (4.19) have slopes of the same sign.*

For different combinations of parameters represented by regions 1-3 in Fig.4.1, the two loci will be either positively or negatively sloped. However, the fact that they always have slopes of the same sign implies that only four qualitatively different situations are possible. They are represented in Fig.4.2-4.5 assuming that both loci have negative slopes.

Fig.4.2 and 4.3 show two corner equilibria with formal jobs alone, $(\theta^*, 1)$, and with informal jobs alone, $(\theta^*, 0)$, respectively. In the first case, the locus of formal jobs lies above the locus of informal jobs for all possible values of ϕ . As higher values of the market tightness, θ , imply lower profits for firms in both the formal and informal sectors, the relative position of the two loci indicates that formal firms are more profitable and can cope with more competition from other firms than their informal counterparts can do. In other words, formal firms break even at higher values of the market tightness so they

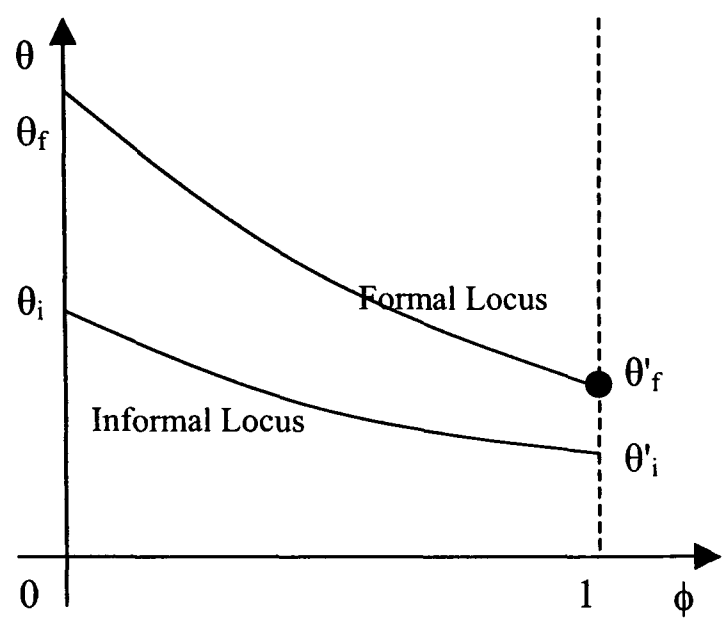


Figure 4.2: No shadow sector equilibrium

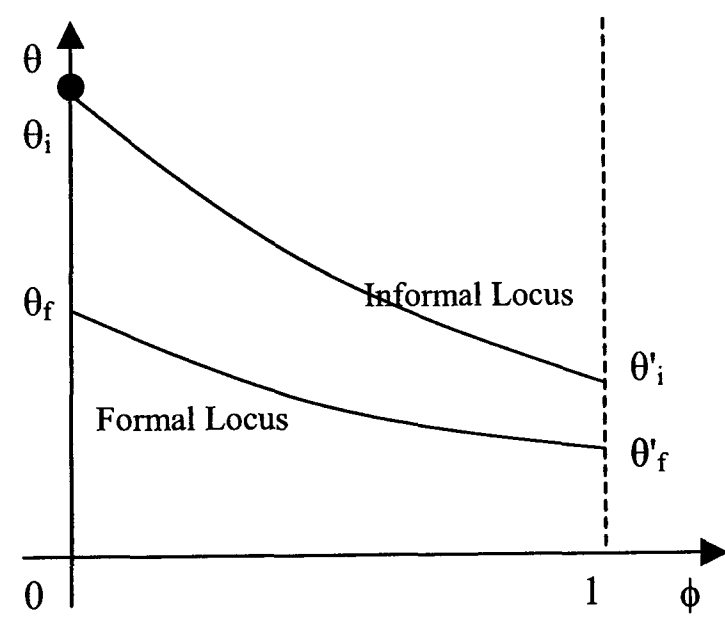


Figure 4.3: Shadow sector equilibrium

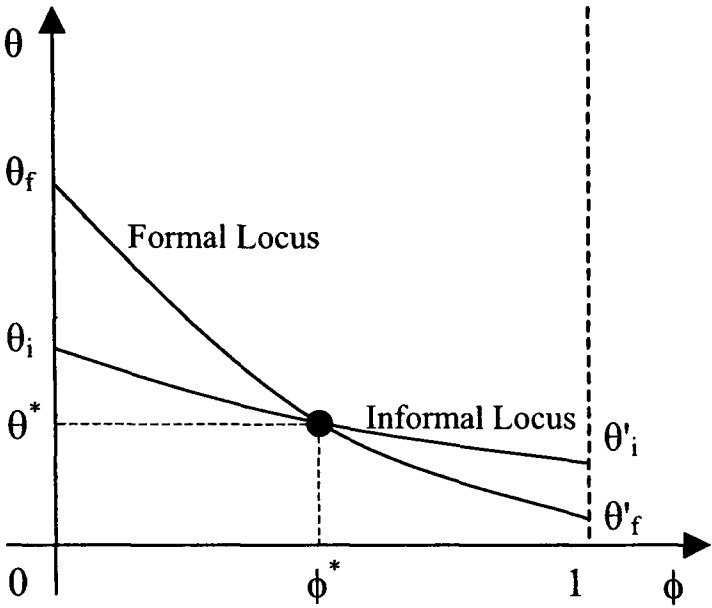


Figure 4.4: Stable interior equilibrium

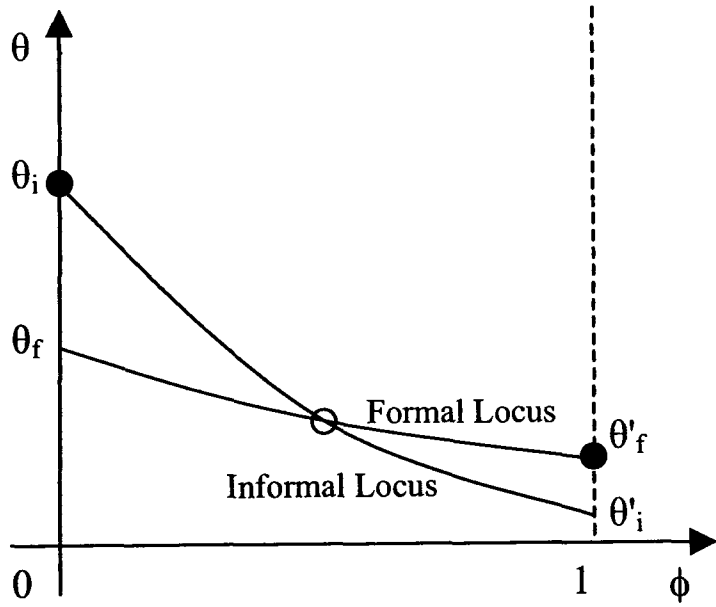


Figure 4.5: Multiple equilibria

can afford higher costs of functioning. Thus, they still can make positive profits while informal firms already make losses. The case with purely informal equilibrium is just a mirror image of the situation just described.

Fig.4.4 and 4.5 present two more interesting cases where mixed or interior equilibria with both formal and informal jobs are possible. In those cases two loci (4.18) and (4.19) intersect at some θ^* and $0 < \phi^* < 1$. Depending on the relative position of the two loci an interior equilibrium can be accompanied by corner equilibria of the types we discussed in the previous paragraph. At the same time, the mixed equilibrium itself can either be stable or unstable, so that in the long run the economy either have chances to end up in a situation with both types of job or is likely to settle in a corner equilibrium with one type of job only. It should be noted that by stability we mean here the ability of the economy to return to the same equilibrium allocation of vacancies if some sort of a shock pulls it out of the equilibrium position. For example, if in the equilibrium shown in Fig.4.4 some random perturbation leads to an increase in the proportion of formal vacancies above ϕ^* , the two zero profit conditions imply that *ceteris paribus* firms opening or holding an informal vacancy are better off than firms opening or holding a formal vacancy. This will attract more firms in the informal sector so that the proportion of formal vacancies will eventually return to equilibrium value ϕ^* . Thus, the interior equilibrium depicted in Fig.4.4 is stable, while the interior equilibrium in Fig.4.5 is unstable. The corner equilibria in Fig.4.5 are stable.

Both Fig.4.4 and 4.5 illustrate the case where there is only one interior equilibrium. However, as both loci have slopes of the same sign, potentially they can intersect in a number of points which would correspond to different interior equilibria. Nevertheless, in the

case of our model it can readily be verified that the interior equilibrium is always unique, whenever it exists. The conditions for existence of equilibria of different types shown in Fig.4.2-4.5 are given in the next subsection. It also provides an intuition underlying our results.

Conditions for existence and stability of equilibria Any equilibrium in our model results from interaction of unemployed workers and firms in the labour market when they bargain over the rents to be generated by the match.

When one of the sides completely dominates the market its preferences unambiguously define which sector of employment exists in equilibrium. For example, when the market tightness is zero a firm has no problem finding a match and enters the sector that provides the highest averaged return on start-up expenditures. Recalling the discussion of profit functions in Section 4.3.6, in such a situation the firm should expect to receive share $(1 - \beta) \frac{S_f}{\pi}$ of rents in the formal sector or share $(1 - \beta) \frac{S_i}{\rho}$ of informal rents. Thus the returns on start-up costs k_f and k_i are given by $(1 - \beta) \frac{S_f}{\pi k_f}$ and $(1 - \beta) \frac{S_i}{\rho k_i}$ in the formal and informal sectors, respectively. If, for instance, $(1 - \beta) \frac{S_f}{\pi k_f} > (1 - \beta) \frac{S_i}{\rho k_i}$, or alternatively, $\frac{S_f}{S_i} > \frac{\pi k_f}{\rho k_i}$, the firm prefers the formal sector.

To the contrary, in a situation when the market tightness is infinite, it is the workers who instantly receive offers from both formal and informal firms. From expressions (4.14) and (4.15) and the properties of function $E_u(\theta, \phi)$ it follows that workers receive wages $w_f = S_f$ and $w_i = S_i$ in the formal and informal sectors, respectively. Thus, while competing for workers firms cannot do any better than offer the whole surplus of the match in any sector. Obviously, in such a situation workers will turn down offers of a lower wage, i.e. if, for example, $S_f > S_i$, workers will never accept offers from the

informal sector.¹⁸

In the process of bargaining the balance of power shifts either to one or the other side depending on the level of market tightness. Thus, it should be intuitively clear that if in the two extreme cases just considered both firms and workers prefer the same sector, then in equilibrium with matching frictions on both sides only jobs in that sector are created. Then the equilibrium market tightness is stabilised at such a level that profits of firms are equal to zero. If, however, given full control of the market, preferences of firms and workers do not coincide, an equilibrium with both types of job can result.

As preferences of both firms and workers over the sector choice when they do not face matching problems are determined by the relative values of sector surpluses S_f and S_i , and respective returns $\frac{S_f}{\pi k_f}$ and $\frac{S_i}{\rho k_i}$, all possible situations can again be graphically represented in the $\left(\frac{S_f}{S_i}, \frac{\pi}{\rho}\right)$ -plane. The possible cases are illustrated in Fig.4.6 that shows four non-overlapping regions 1-4, each of which corresponds to various combinations of parameters and represents a particular equilibrium type. To bear a resemblance to Fig.4.1 we measure the ratio of formal and informal surpluses, $\frac{S_f}{S_i}$, on the vertical axis, and the ratio of the effective discount rates, $\frac{\pi}{\rho}$, on the horizontal axis. The borders of the regions are defined by the two axes, the straight line with slope $\frac{k_f}{k_i}$ and running through the origin,¹⁹ and the straight lines $\frac{S_f}{S_i} = 1$ and $\frac{\pi}{\rho} = 1$.

Region 1 In region 1 in Fig.4.6 $S_f > S_i$ and $\frac{S_f}{\pi k_f} > \frac{S_i}{\rho k_i}$, i.e. both sides of the market prefer formal employment when they have full control over dividing the match surplus.

In such a situation only an equilibrium with formal jobs can satisfy both parties.

¹⁸Also, condition $S_f > S_i$ implies that the value of formal employment is greater than the value of informal employment, $E_f > E_i$, as has been explained in section 4.3.6.

¹⁹It runs above the 45-degree line as $k_f > k_i$ by assumption.

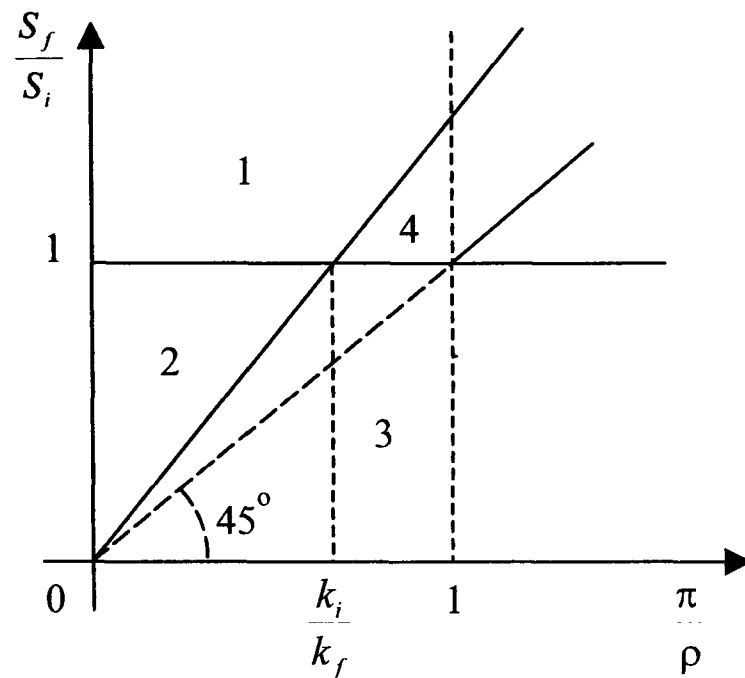


Figure 4.6: Existence conditions for model's equilibria

In this region derivative $\frac{\partial E_u(\theta, \phi)}{\partial \phi}$ is positive, so that expressions (4.21) and (4.22) are negative, and both loci (4.18) and (4.19) are downward sloping in the (θ, ϕ) -plane. The intuition for that is as follows. The value of being non-employed rises as the proportion of formal vacancies, ϕ , increases. As with the rise in ϕ the reservation option of workers, $rE_u(\theta, \phi)$, becomes greater in value, firms have to pay higher wages. This can be seen from the two expressions for wages (4.14) and (4.15). A rise in wages squeezes firms' profit margins so that firms break even at lower levels of market tightness, i.e. the duration of vacancy idleness, $\frac{1}{q(\theta)}$, that can be tolerated by firms before a productive match is formed, is shorter. Thus in the (θ, ϕ) -plane for both loci (4.18) and (4.19) higher values of ϕ must correspond to lower values of θ .

As seen from (4.21) and (4.22) depending on the relation between πk_f and ρk_i the formal job locus can be either flatter or steeper than the informal one. In any case, as $\frac{\partial \Pi_f(\theta, \phi)}{\partial \theta} \big|_{\Pi_f(\theta, \phi)=0} < 0$ and $\frac{\partial \Pi_i(\theta, \phi)}{\partial \theta} \big|_{\Pi_i(\theta, \phi)=0} < 0$ (see Appendix 4.A for proof), the formal

locus must lie above the informal one. The resulting equilibrium without informal jobs is stable, and the market tightness is such that profits in the formal sector are nil, while profits in the informal sector are negative (Fig.4.2).

Region 2 In region 2 in Fig.4.6 restrictions on parameters suggest that the formal surplus is smaller than the informal surplus, $S_f < S_i$, while the ratio of returns on entry costs implies $\frac{S_f}{\pi k_f} > \frac{S_i}{\rho k_i}$. This means that, on the one hand, the formal sector is more appealing to employers when they dominate the market, but, on the other hand, workers would prefer being employed informally if they had no problem landing a job. Thus, there must exist a value of the market tightness, θ^* , and the proportion of formal vacancies, $\phi^* \in (0, 1)$, such that firms are indifferent as to the sector where to place a vacancy. It is easy to verify, that indeed in this case the locus of formal jobs (4.18) and the locus of informal jobs (4.19) have an intersection point for some $0 < \phi^* < 1$, i.e. an interior equilibrium with both types of job exists. From $S_f < S_i$ and $\frac{S_f}{\pi k_f} > \frac{S_i}{\rho k_i}$ it follows that the flow value of formal start-up costs, πk_f , must be smaller than the flow value of informal start-up costs, ρk_i , in this region, i.e. the formal job locus is steeper than its informal counterpart in some neighbourhood of interior equilibrium (θ^*, ϕ^*) . As regards the sign of the slopes of the two loci, by comparing Fig.4.1 and Fig.4.6 it can be seen that in region 2 partial derivative $\frac{\partial E_u(\theta, \phi)}{\partial \phi}$ can be either negative or positive, so from (4.21) and (4.22) the loci can be either positively or negatively sloped. From Proposition 1 we know that the sign of the derivative is negative for any $\theta > \bar{\theta}$, and positive for any $\theta < \bar{\theta}$, where $\bar{\theta}$ is some threshold value of the market tightness. Then for $\theta > \bar{\theta}$ the two loci will both be positively sloped, whereas for $\theta < \bar{\theta}$ they will be downward sloping. The outcome bears on the stability of the interior equilibrium.

Proposition 2 *Let $\bar{\theta}$ be a threshold value of market tightness such that $\frac{\partial E_u(\bar{\theta}, \phi)}{\partial \phi} = 0$. and let (θ^*, ϕ^*) be a point of an interior equilibrium in region 2, Fig.4.6. Then given $k_f > k_i$ and $\rho > \pi$, θ^* is always less than $\bar{\theta}$.*

Proof: see Appendix 4.A.

Proposition 2 implies that if the two loci of formal and informal jobs intersect and an interior equilibrium results we can confine ourselves to the situation with a less tight labour market, i.e. $\theta < \bar{\theta}$. Then $\frac{\partial E_u(\theta, \phi)}{\partial \phi} > 0$, the two loci have negative slopes, and the formal job locus crosses the informal job locus from above. The resulting equilibrium is unique and stable (Fig.4.4).

Region 3 The situation in this region mirrors the one in region 1: both the formal surplus is less than the informal surplus, $S_f < S_i$, and the return on start-up costs in the formal sector is less than the return on entry into the informal sector, $\frac{S_f}{\pi k_f} < \frac{S_i}{\rho k_i}$. Thus, both sides of the labour market favour the informal sector, so that the equilibrium with informal jobs only results. In this region $\frac{\partial E_u(\theta, \phi)}{\partial \phi}$ can potentially be either negative or positive, while πk_f can be either greater or less than ρk_i , so that the relative steepness of the loci and the sign of their slopes are ambiguous in general. However, whatever case comes about the informal locus lies above the formal one, and the outcome is stable to changes in parameters (Fig.4.3). The equilibrium market tightness drives informal profits to zero, while formal profits become negative.

Region 4 Finally, in region 4, $\frac{S_f}{\pi k_f} < \frac{S_i}{\rho k_i}$, while $S_f > S_i$. So, the employers prefer the informal sector when face no problem meeting workers, while the workers are unambiguously after formal jobs when market tightness is infinite. By analogy with the case of

region 2, the locus of formal jobs (4.18) and the locus of informal jobs (4.19) intersect at some $\phi^* \in (0, 1)$, i.e. there exists an interior equilibrium. The restrictions on parameters in region 4 can hold only if $\pi k_f > \rho k_i$, i.e. the formal locus is flatter than the informal locus in the vicinity of (θ^*, ϕ^*) . As $\frac{\partial E_u(\theta, \phi)}{\partial \phi} > 0$, the two loci are negatively sloped, and hence, the formal locus crosses the informal one from below at (θ^*, ϕ^*) . This situation is shown in Fig.4.5 where it can be seen that the interior equilibrium is unique but not stable in this case, while there exist two stable corner equilibria with informal and formal jobs alone.

4.3.7 Discussion

Above we have found general conditions for the existence of equilibria of different types. It has been shown that whenever both sides of the labour market - firms and workers - prefer formal (informal) employment in the situation when they do not face matching problems, the resulting equilibrium will comprise only formal (informal) jobs. However, if given full control of the labour market the demand and supply sides differ in their preferences over the sectors, frictions in matching can ensure that there exists an interior equilibrium with both types of job so that firms become indifferent as to which sector to enter. This last result is especially important for two major reasons.

First, it shows the importance of labour market institutions, such as wage bargaining, for the emergence of equilibria where two sectors coexist in the long run. Wages act as a channel through which asymmetry in governmental regulations in relation to formal and informal firms leads to market duality. Governmental regulations affect the size of match surpluses in the two sectors, as well as the outside options of both firms and workers in

the wage bargaining process.

Second, such an interior equilibrium is obtained in the absence of specific *ad hoc* assumptions about *ex ante* characteristics of workers, their preferences, or various externalities that were simply built in models' technologies in order to provide for formal-informal segmentation of the economy. All externalities present in our model are derived from market interactions. This is in contrast with the previous theoretical literature which overview was given in Section 4.2.

In the rest of this section we examine implications of the above analysis of the model and emphasise the role of various assumptions for the outcomes we obtain. In particular, we expound on the mechanism whereby government regulations, wage bargains and imperfect labour markets make for the emergence of the informal sector in the long run. The section is concluded by touching upon the issue of stability of the long-run equilibrium with formal and informal jobs before summarising the main points.

Important assumptions

An important meaning of the above analysis is that in a standard model of the labour market it indeed reveals the possibility for stable coexistence of the formal and informal sectors in the long term. In the model this result depends almost exclusively on the parameters reflecting the degree of regulations of the economy (i.e. taxes, fines for running business informally, the degree of monitoring), and costs of accessing a particular sector, k_f and k_i . It is likely that many (or even all) of those parameters are effective or potential policy tools in reality. At the same time, the outcome is independent of, for example, preferences of workers over formal and informal goods, heterogeneity of workers, production technology parameters, a form of the monitoring function or the penalty rate. The

result, however, hinges upon four important assumptions.

First, let us take the assumption of wage bargaining. It can be shown that dropping this assumption and, for example, assuming *ceteris paribus* wage posting, leads to corner solutions. Indeed, consider equilibrium wages in our model. Two zero profit conditions (4.18) and (4.19) combined with (4.14) and (4.15), can be solved for equilibrium values of w_f and w_i , which are

$$w_f = \frac{\beta r \pi k_f}{(1 - \beta) q(\theta)} + r E_u(\theta, \phi), \quad (4.23)$$

$$w_i = \frac{\beta r \rho k_i}{(1 - \beta) q(\theta)} + r E_u(\theta, \phi). \quad (4.24)$$

That is, the equilibrium wage differential is

$$w_f - w_i = \frac{\beta}{(1 - \beta)} \frac{r(\pi k_f - \rho k_i)}{q(\theta)}. \quad (4.25)$$

Wage posting can be seen as a situation in which all the bargaining power is vested with firms, or, in other words, $\beta = 0$. From (4.23) and (4.24) above it is clear that putting β equal to 0 eliminates the equilibrium wage differential in (4.25) and, thus, the preconditions for labour market segmentation.

When building our model we kept in mind not advanced but transitional economies, so the question arises of whether the assumption of wage bargaining is reasonable in the context of countries of Eastern Europe. The empirical studies by Grosfeld and Nivet (1999), Luke and Schaffer (1999), Shakhnovich and Yudashkina (2001) provide evidence in full support of the presumption. In the most recent work Basu *et al.* (2004) indicate

that if at the end of the communist period evidence of worker sharing in their enterprise rents and losses was a feature only in some transitional economies, within a year after the start of transition rent sharing became prevalent in all the economies that they study (which are the Czech Republic, Slovakia, Poland and Hungary).

The second important assumption is the presence of search frictions. The modelling of matching between firms and workers hangs upon the form of matching technology. The Inada-type assumptions introduced in Section 4.3.2 are crucial to the existence of an interior equilibrium for parameter values satisfying restrictions in regions 2 and 4, Fig.4.6. Nevertheless, such a specification of the aggregate matching function (which includes a Cobb-Douglas functional form) is generally favoured by the empirical studies (for a review see, e.g. Petrongolo and Pissarides, 2001), while Stevens (2004) provides microeconomic foundations for it.

Third is the irreversibility of firms' decision as to what type of job to open. The irreversibility can first be seen as a consequence of investing into capital of different quality. If formal and informal jobs use different capital then the irreversibility assumption is justified in the absence of a perfect second-hand market for capital goods (as in Acemoglu, 2001). In other words, firms in either sector have to bear some sunk costs before opening a vacancy that can be seen as some sort of start-up irreversible investments in capital. This point is supported by the evidence from the developing world. For example, Loayza (1996) points to the fact that in developing countries with its endemic inefficient capital markets and the lack of proper contracts informal firms face high borrowing rates, are unable to transfer property, and create common stock corporations. In such conditions even if costs of informality grow and incentives to be formal become stronger, the more difficult it is for

informal firms to accumulate the wealth that would enable them to enter the formal sector. However, in the context of our model entry investments are different due to higher entry costs to legality, not the quality of capital. Furthermore, the productivity of a match is the same across sectors. In such conditions the irreversibility of opening an informal vacancy can be justified by the same reason as in the case with capital expenditure: turning formal requires more funds spent on obtaining licenses, registration (we leave aside the question of opening of new vacancies by the firms already existing in the market), etc. At the same time, downgrading from being formal to being informal cannot be done easily due to the fact that, first, some funds will have already been spent on licenses and, thus, effectively sunk, and, second, the merest disappearance of a formal firm from the market not through a bankruptcy procedure is a big deal even in the countries where bankruptcy procedures are not clearly stipulated.

Finally, the fourth important ingredient affecting the results that we have obtained is some degree of heterogeneity which is present in our model. In particular, the formal and informal sectors are different in the access costs. This leads to heterogeneity of firms and, hence, to possibilities for labour market segmentation. However, in contrast to the approach taken in many theoretical studies mentioned in Section 4.2, firms in our model are not *ex ante* heterogeneous. In other words, the heterogeneity is not assumed exogenous but is derived from investment decisions of firms. Moreover, the factor of heterogeneity, i.e. start-up costs, is a product of governmental regulations. It can potentially be affected by structural policies.

Thus, as one can see, all the assumptions that we make are generally empirically justified. So, wages, or, more precisely, the wage determination process in the frictional

labour market can be critical in transmitting the effect of governmental actions on the formal-informal divide. That said, the model can be extended to include preferences of workers, specificity of sector goods (as was done, e.g. in a similar study by Kolm and Larsen, 2004), moral costs (Kolm and Larsen, 2001; Fugazza and Jacques, 2004), and so forth, which will not change the qualitative result but will just distract attention from main mechanisms in action. Formal regulations have a strong direct bearing on the long-term emergence of segmented labour markets when the economy is characterised by search and rent sharing. We take this point further in the next subsection.

Regulations and non-competitive labour markets

From the most general perspective the various government regulations aimed at both the formal and informal sectors determine the size of a surplus that a potential match generates. These rents are to be divided between the two parties forming a match - the firm and the worker. If we assume a perfectly competitive benchmark where matching between firms and workers is instantaneous then firms will hire workers at the same wage equal to their marginal product, and will receive the rest of the surplus. Clearly, there will be no equilibrium wage differences, while firms will prefer the sector with a higher surplus, so that a corner equilibrium will result. As soon as we enter the world of search, however, the fact that matching for both firms and workers takes time implies that there will be some rent sharing in a bilateral monopoly game (as in, e.g. Shaked and Sutton, 1984). In such conditions the worker appropriates part of the match surplus, as the firm now has to compensate the job-seeker for the time it would spend searching for a replacement if the worker quits the negotiations to take her outside option. As our model suggests, depending on the parameters, such as taxes, fines, and so forth, the match surpluses in

the formal and informal sectors are not equal in general, hence nor should potentially be the equilibrium wages. There is the scope for labour market segmentation.

A higher surplus implies a higher wage in equilibrium, as can be seen, for example, from (4.14) and (4.15). At the same time, it implies higher proceeds to firms as well. Thus, the sector which is more attractive for its high surplus, is, at the same time, less attractive for its high labour costs. How does this bear on the resulting equilibrium? The outcome depends on outside options or/and status-quo positions of firms and workers in wage negotiations.²⁰ If firms have the same threat points when bargaining over wages in the formal and informal sectors, the Nash bargaining solution implies that firms will choose a sector with a higher surplus. The same applies to workers. Then preferences of firms and workers over the sector of employment, when they face no matching problem, are driven by the sector surplus size and always coincide, i.e. only corner equilibria will be possible. If, however, either party in the bargaining process has different, i.e. sector specific, outside options then in general it is not necessarily the case that the sector with the highest surplus is preferred by both parties. For the party with sector specific status-quo positions the result depends on the relative value of outside options and the relative value of sector surpluses. In such a situation the preferences over the sector of production given the full bargaining power can potentially be different for firms and workers, depending on the extent of asymmetries. Suppose that the asymmetries are on the side of the party that actually makes the decision as to which sector to enter (in the case of our model it is the firms, whereas workers search randomly without choosing a particular sector in advance).

Then, if in equilibrium these asymmetries in the outside options are levelled off by wage

²⁰The reader can be referred to Sutton (1986) for initial insights into the effect of outside options in non-cooperative bargaining theory.

differentials, an interior equilibrium arises where the entering party is just indifferent as to which sector is better. If, however, the potential wage differentials cannot compensate for the asymmetries in the outside options the entering party will always prefer a particular sector, so that a corner equilibrium results.

In our model the firms do have different outside options depending on the sector when they bargain over wages with workers. In particular, their options are defined by the value of sunk start-up costs, k_f and k_i , and the effective discount rates, π and ρ . The upfront expenditure, k_f and k_i , must be borne before the firm meets a job-seeker, so that when it happens and wage bargaining starts the expected flow values of firm's outside option are $-\pi k_f$ and $-\rho k_i$ in the formal and informal sectors, respectively. Thus, what they care about when deciding as to which sector to enter is not the size of the sector surplus, i.e. not the payoff, but the return on the start-up investments, $\frac{S_f}{\pi k_f}$ and $\frac{S_i}{\rho k_i}$. At the same time, what matters for workers is the size of surpluses, S_f and S_i . In particular, as can be seen from (4.23) and (4.24), wages in the formal and informal sectors depend positively on πk_f and ρk_i , while the difference between the two defines the equilibrium wage differential - see (4.25). In an interior equilibrium as in Fig.4.4 or 4.5 wage differences precisely compensate firms for the differences in returns $\frac{S_f}{\pi k_f}$ and $\frac{S_i}{\rho k_i}$, so firms can make zero profits in both sectors.

Thus, asymmetries in flow values of start-up costs are crucial for an interior equilibrium to ensue. These asymmetries in our model are due to the factors directly affected by the state of the government and its regulations. In particular, a value of k_f higher than a value of k_i is explained by bureaucratic extortion, i.e. by corruption at the lower tier of government officials. A differential between the effective discount rates, π and ρ , is due

to government monitoring activities that crack down on the informal sector. From (4.23) and (4.24) it is easy to see that if, for some reason, $\pi k_f = \rho k_i$, wage differences disappear in long-run equilibrium, so that the relative value of returns $\frac{S_f}{\pi k_f}$ and $\frac{S_i}{\rho k_i}$ is determined by the relative value of surpluses, S_f and S_i . Similarly, if, for instance, $k_f = k_i = 0$ wages in both sectors become equal to workers' outside option rE_u . In all such cases firms' and workers' preferences over the sector of production are the same and only corner equilibria can result.

To sum up, search and rent sharing in the presence of asymmetries in start-up costs, caused by activities of various levels of the government, can bring about a long-term interior equilibrium with formal and informal jobs. From Fig.4.4 and 4.5 we can see that such an equilibrium, when it exists, can be either stable or unstable. The same asymmetries discussed here play a leading role in stability of equilibrium too, which is a focus of the following section.

Notes on stability of an interior equilibrium

The analysis of Fig.4.4 and 4.5 in Section 4.3.6 shows that the stability of an interior equilibrium depends on the sign and the relative value of the slopes of two loci (4.21) and (4.22). Those in turn depend on the sign of $\frac{\partial E_u(\theta, \phi)}{\partial \phi}$ and the relative value of πk_f and ρk_i . Here, we provide insights into the stability issue from the perspective of the effect of the value of being unemployed, $E_u(\theta, \phi)$, and the flow values of start-up costs on wages, and hence, profits in the formal and informal sectors.

As we have seen in (4.23) and (4.24) the equilibrium wages paid by firms in the formal and informal sectors can be represented as a sum of two terms - the basis equivalent to worker's outside option, rE_u , and the "top-up" term paid to a hired worker for saving of

opportunity costs that a representative firm enjoys when a job is formed.²¹

Proposition 1 and Appendix 4.A ascertain that for $k_f > k_i$ and $\pi < \rho$ the value of being unemployed is increasing in the proportion of formal vacancies, ϕ , in regions 2 and 4, Fig.4.6, around the points of interior equilibria. Thus a second term in (4.23) and (4.24) is increasing in ϕ and θ . Either a rise in ϕ or θ (or in both) improves prospects of the unemployed. Such a change affects both the formal and informal wages in the same manner, but the cost of it is always higher to firms in a sector with smaller surplus: the same absolute increase in ϕ or θ eats away a larger percentage share of profits in that sector.

At the same time, the top-up term is independent of ϕ and varies only with the market tightness, θ . Both the absolute effect of a change in θ and the effect of such a change relative to the size of sector surplus are always larger in the sector with a higher flow value of start-up costs, i.e. with a lower outside option to firms once entry costs have been sunk.

As firms enter the economy, market tightness θ rises, so that workers bid wages higher and higher. It can be verified that in such a situation for low values of ϕ profit margins of firms in the sector with smaller surplus are reduced more slowly. This happens because the relative effect on such firms of a wage rise due to changes in the worker's outside option, rE_u , is dominated by the relative top-up effect in wage changes on profits in the sector where outside options of firms are lower.²² Ultimately, as the firms in the latter

²¹The value of the first term in the expressions for equilibrium wages (4.23) and (4.24) depends on the average duration of a vacancy. This represents the worker's compensation for saving firms the opportunity costs. Indeed, from the discussion in Section 4.3.6 it follows that when $\theta \rightarrow 0$ and $q(\theta) \rightarrow \infty$, the equilibrium converges to the Walrasian limit point, and both wages w_f and w_i converge to the value of being unemployed, rE_u . In this case resources invested before entrance never remain idle and workers are not compensated for saving of opportunity costs. If, however, $\theta > 0$ workers have to be compensated as the opportunity costs of idleness are significant. In the other extreme case, i.e. when $\theta \rightarrow \infty$ and $q(\theta) \rightarrow 0$, the opportunity costs of idleness become very high, wages grow large, until workers appropriate all the sector surpluses.

²²Conditions for the existence of an interior equilibrium imply that a sector with a lower surplus and a sector with a lower firms' outside option are necessarily different.

sector break even, the firms in the other sector still make positive profits. even though in absolute value their surplus is lower. Thus, the proportion of vacancies posted by such firms rises.

In contrast, as wages change at high values of ϕ , the relative effect of the worker's outside option on firms in the sector with lower surplus outweighs the relative top-up effect on firms in the sector where their outside options are lower, so that firms in the sector with smaller surplus suffer more as their profits dwindle away faster as θ rises. The proportion of vacancies posted by these firms decreases. The two effects are exactly balanced at the equilibrium level of the proportion of formal vacancies, ϕ^* . Thus, for all $\phi < \phi^*$, the top-up effect dominates the worker's outside option effect, and for all $\phi > \phi^*$, the worker's outside option effect dominates the top-up effect.

The workings of the two effects imply that an interior equilibrium in Fig.4.4 is stable. Indeed, for such an equilibrium the expected return on start-up costs is greater for formal firms, $\frac{S_f}{\pi k_f} > \frac{S_i}{\rho k_i}$, while the size of the match surplus is greater for informal firms, $S_f < S_i$. This implies that the top-up effect is greater in the informal sector, as $\pi k_f < \rho k_i$. Then for any $\phi < \phi^*$, working through wages the top-up effect on informal profit margins dominates the worker's outside option effect on formal profit margins, and thus it is the formal firms that post more and more vacancies in the economy. Thus, ϕ rises. On the other hand, if $\phi > \phi^*$, the worker's outside option effect on formal firms is greater than the top-up effect in the informal sector, so that informal firms post more vacancies than their formal counterparts, i.e. ϕ decreases. Once ϕ reaches the value of ϕ^* firms in both sectors break even at precisely the same level of market tightness, θ^* , which is, thus, the equilibrium value. By similar reasoning the interior equilibrium in Fig.4.5 is unstable.

The main implication of the discussion in this subsection is that asymmetries in start-up costs or/and match duration (which is reflected in different value of effective discount rates π and ρ) along with asymmetries in sector surpluses lead through wage bargaining to either stable or unstable outcomes. Again, search and rent sharing are crucial to the results, while specific combinations of various parameters (k_f , k_i , π , ρ , and other parameters determining the size of surpluses S_f and S_i) define the stable outcome.

Summary

The purpose of this section has been to highlight the role of governmental regulations and labour market institutions in the emergence of a stable long-run equilibrium with both the formal and informal sectors. We have deliberately concentrated on that type of equilibrium because corner equilibria with the formal or informal sectors alone are not particularly interesting from a policy perspective. On the one hand, the absence of informality in the long term, implies that informalisation is just a transitional phenomenon of adjustment in the economy. On the other hand, equilibria with the informal sector alone are likely to be unrealistic because governments may have numerous incentives to avoid complete informalisation of the economy: in addition to a number of insights mentioned in the introductory sections, the obvious intuition is that raising revenues through tax collection (i.e. from the tax-compliant sector) implies lower transaction costs than so doing through monitoring of the underground businesses (see, e.g. Slemrod and Yitzhaki, 2002).

We have shown that regulations and policies towards formal and informal firms imply different rents of functioning across sectors. When labour markets are imperfect and characterised by search and rent sharing, surplus differentials can lead to equilibrium

wage differentials and, hence, labour market segmentation. This result is important, because in general it does not depend on further assumptions about preferences of workers, heterogeneity of firms or employees, a specific form of penalty function, etc. All those factors can be successfully incorporated into the model which, however, will keep its main implications intact.

Search and rent sharing have recently been increasingly used for modelling of the interaction between firms and workers through which informality can emerge. Kolm and Larsen (2001, 2004), Boeri and Garibaldi (2001), Bouev (2002), and Fugazza and Jacques (2004) have all studied various versions of a labour market search model à la Pissarides (2000) with an application to the informal sector. However, the main focus of these studies is policy implications and the effect of the informal sector and governmental regulations on the level of employment and unemployment. The role of search and rent sharing in the emergence of the informal sector has not been accentuated nor has it been properly analysed.

While the existence of frictions in the formal labour market should cause no doubts (see, e.g. Acemoglu, 1996, 1999, for references to important sources of evidence), Fugazza and Jacques (2004) point out that there is some indication that favours application of the same approach to modelling of informal labour markets. In this work, however, we do not separate labour markets from the point of view of the worker, who is assumed to search randomly. We postulate that it is the firms that decide in the end in which sector to function. This could well be a feature of transitional labour markets in Eastern Europe, and especially Russia, where job-seekers could well be happy to obtain any offer of employment in the face of rising transitional unemployment.²³

²³This posit has to be verified empirically. However, some general discussions - see, for instance, Boeri

Wage bargaining is also inherent in many advanced European economies. At the same time there is substantial evidence that it is pertinent to transitional countries as well. It is often emphasised that workers can have enough power to push up wages in response to gains in productivity and, thus, appropriate some of the firm-specific rents, which leads to losses of resources otherwise available for further investment (e.g. Grosfeld and Nivet, 1999; Shakhnovich and Yudashkina, 2001). So, rent sharing and its role in wage differentials across the two sectors should be given more close attention.

We have found that differentials in two factors are especially important for equilibrium wage differences and, hence, the existence and stability of an interior equilibrium.

First is the difference in the effective discount rates in the formal and informal sectors that can be due to the government's audit of informal firms. Monitoring of the underground sector leading to a higher death rate of informal matches has featured in the models by Boeri and Garibaldi (2001), Bouev (2002), and Kolm and Larsen (2001, 2004). Empirical evidence from less developed countries also corroborates the view of higher labour turnover in the underground sector (Kaufmann and Kaliberda, 1996; Hoek, 2002). Our analysis shows that differences between the turnover rates across the formal and informal sectors have a direct bearing on the equilibrium wage differential, provided that firms are subject to some costs of posting a vacancy.

Second is the differentials in costs of vacancy posting themselves. The higher costs of access to the formal sector are widely featured in the descriptive literature on the informal sector (De Soto, 1989, is the most popular reference). Here we incorporate them to show that such costs lead to equilibrium wage differences, and hence, labour market

(2000) or Gërxhani (2004) - indicate that in transitional countries informality is often of subsistence nature, thus suggesting that workers hardly purposely direct their search to the informal sector.

segmentation. Interestingly, it can be shown that the conclusions of the model hold even in the absence of the start-up costs, but in the presence of varying *flow* costs of maintaining a vacancy in either sector. Whenever wages are determined by ex post Nash bargaining what is important for the ensuing wage differentials is that firms have to invest some resources, e.g. spent them on preserving the vacancy, before they meet workers, so that they care about opportunity costs of their investments.

Thus, wage differentials and, hence, differences in labour costs are an important feature of formal-informal segmentation. Castells and Portes (1989) have noted that one of the best-known effects of the informalisation process is to reduce costs of labour substantially. It is widely held that such costs can be derived from minimum wages, high social security contributions, redundancy pay, fringe benefits, constraints on free hiring and dismissal (see *inter alia* Loayza, 1996; Schneider and Enste, 2000; Boeri and Garibaldi, 2001). Our model emphasises another aspect of labour costs that assists the emergence of informality. Namely, in the presence of wage bargaining and rent sharing workers may well have a monopoly power in wage negotiations to extract rewards for saving of opportunity costs that result from firms having invested in vacancy creation before meeting with job-seekers. In fact, this points to the presence in our set-up of a version of the seminal hold-up problem (see Grout, 1984; Williamson, 1985), when workers share in returns on investment borne by the other party, i.e. firms. We shall return to this question in the next section when we address some efficiency issues arising in our model.

4.4 Welfare and Policy Making

Now that we have established the role of wage bargaining in the emergence of formal-informal duality we shall consider implications of our model for policy making and address some efficiency questions that arise in such models of labour markets with search frictions.

4.4.1 Implications for Policies

Equilibria in our model emerge as a result of bargaining between workers and firms over the gains that productive matches generate in the two sectors. The government has a direct impact on the outcome of this process by affecting either the size of the surpluses in the sectors or the outside options of negotiating sides (or both), as can be seen from Section 4.3.7. What does this framework imply for policies that the government could implement to drive the economy to a long-run equilibrium of a particular type?

Consider, for example, a policy-maker who wants to choose a combination of policy instruments to support the formal sector. That is, suppose that the policy-maker, or the government for that matter, wants to avoid in the long run the equilibrium with informal jobs (i.e. the one as in Fig.4.3), and aims to attain, if possible, the equilibrium with formal jobs only (Fig.4.2). Also, for simplicity, suppose that all but one parameters of the model are fixed, that and the policy-maker is effectively to decide only on the level of, for instance, the tax rate, τ . What does our model have to say about the level that the tax rate must be set to guarantee convergence to a purely formal equilibrium in the long term? Or, more generally, what is the effect of a variation in the tax rate on the type of the resulting equilibrium? In order to answer these questions we shall proceed in the following way. Section 4.3.6 above has provided us with the conditions on the parameters

of the model that make for different types of equilibrium. Thus, by reformulating the restrictions on sector surpluses S_f and S_i , and their respective returns, $\frac{S_f}{\pi k_f}$ and $\frac{S_i}{\rho k_i}$, as the restrictions on the level of the tax rate, τ , we can determine what values of the tax rate correspond to different equilibria. Having done that we shall see what implications our exercise can suggest for economic policies, which purpose is assumed to be delivering the economy to the equilibrium without informality. In what follows we shall not, however, consider the implications for the state budget - these are touched on in Section 4.4.2.

Restrictions on the tax rate

By carefully examining the conditions corresponding to regions 1-4, Fig.4.6, it becomes clear that given the relative value of the outside options of firms in wage bargains, i.e. $-\pi k_f$ and $-\rho k_i$, *ceteris paribus* not all four situations are equally possible. In particular, it is easy to see that if, for example, $\pi k_f < \rho k_i$, only equilibria in regions 1, 2 and 3, Fig.4.6, are feasible, whereas equilibria in region 4 are not possible (given $\pi k_f < \rho k_i$ one cannot have $S_f > S_i$ and, at the same time, $\frac{S_f}{\pi k_f} > \frac{S_i}{\rho k_i}$). By contrast, if $\pi k_f > \rho k_i$, only equilibria in regions 1, 3, and 4 in Fig.4.6 can come about.

Suppose that $\pi k_f < \rho k_i$, which means that *ceteris paribus* firms in the formal sector are better off than firms in the informal sector in wage negotiations with workers. From the shorthand notation for sector flow surpluses, S_f and S_i , introduced in Section 4.3.6, and the condition that $S_f > 0$ and $S_i > 0$, it follows that equilibria in region 1, Fig.4.6, result when $\tau < \tau_1^H < \tau_2^H$, where $\tau_1^H = (\rho k_i - \pi k_f) + mF$ and $\tau_2^H = \frac{(y-b_u)}{\rho k_i} (\rho k_i - \pi k_f) + mF \frac{\pi k_f}{\rho k_i}$; super-index "H" is introduced to indicate that we deal with the case when formal firms enjoy a higher reservation option than their informal counterparts. By the same token, equilibria in region 2 come about when $\tau_1^H < \tau < \tau_2^H$. Finally, equilibria in region 3 occur

when $\tau_1^H < \tau_2^H < \tau$.

So, the three types of equilibrium result when the tax rate falls in one of the three non-overlapping regions, which borders are defined by τ_1^H and τ_2^H . As we know, equilibria in region 1 have only formal firms present in the market, equilibria in region 2 are mixed, while equilibria in region 3 are purely informal. Thus, the restrictions on the tax rate indicate that while having other parameters fixed, a tax rate above τ_2^H leads to an equilibrium comprising only informal jobs, a rate below τ_1^H makes for an equilibrium with formal jobs alone, while the values of τ in between τ_1^H and τ_2^H generate interior equilibria with both formal and informal firms present.

Consider now a situation when $\pi k_f > \rho k_i$, i.e. formal firms are now worse off in wage negotiations than the informal ones, all other things being equal. In this case only equilibria in regions 1, 3, and 4 in Fig.4.6 can come about. Again, from the definition of S_f and S_i and the requirement of their positivity it follows that equilibria in region 1 result whenever $\tau < \tau_1^L < \tau_2^L$, where $\tau_1^L = \frac{(y-b_u)}{\rho k_i} (\rho k_i - \pi k_f) + mF \frac{\pi k_f}{\rho k_i}$, $\tau_2^L = mF + (\rho k_i - \pi k_f)$, and super-index "L" indicates the case when the outside option of formal firms is lower than that of informal firms. Similarly, equilibria in region 3 are possible under condition $\tau_1^L < \tau_2^L < \tau$, while equilibria in region 4 happen whenever $\tau_1^L < \tau < \tau_2^L$.

Similarly to the case with $\pi k_f < \rho k_i$, a tax rate above the upper bound τ_2^L implies getting into region 3, i.e. where only long-run equilibria with the informal sector alone are possible, while a rate below the lower bound, τ_1^L , brings the economy into region 1, where the informal sector is eliminated in equilibrium. The values of the tax rate in between τ_1^L and τ_2^L make for equilibria in region 4. In this latter case, since the interior equilibrium is not stable, a likely long-run outcome is either a purely formal or a purely

informal equilibrium: much depends on the initial proportion of formal vacancies in the economy when it starts converging to the long-run destination. A higher initial proportion of formal vacancies increases the chances of getting into an equilibrium with formal jobs alone (Fig.4.5).

Implications

The two cases just considered provide several implications for economic policy. We discuss them in turn below. The first one relates, in particular, to the effect of the level of the tax rate on the probability of ending up in the equilibrium with formal jobs alone. Other implications are of a more general interest and highlight specificity of policy making in the two-sector economy where wages are determined by bargaining.

The first finding that can readily be discovered from analysing the restrictions on the tax rate above indicates that lower tax rates raise the odds of ending up in an equilibrium without the informal sector. This echoes a standard conclusion often made in the literature related to our topic. In particular, it is widely held that higher taxes always increase an incentive for informalisation (see, for instance, Johnson *et al.*, 1997; Friedman *et al.*, 2000). Thus, in general governments that aim to weed out informality in the longer term should avoid excessively high tax rates. As seen from the above, however, the margins that allow one to call a tax rate "excessively high" depend on the relation between outside options of firms bargaining over wages in the formal and informal sectors. Obviously, in general, τ_2^L and τ_2^H are not equal to each other, hence the same tax rate may well be excessively high under one condition (say $\pi k_f > \rho k_i$) but not so under the other ($\pi k_f < \rho k_i$). This is a matter of the second implication.

The cases $\pi k_f > \rho k_i$ and $\pi k_f < \rho k_i$ are mutually exclusive. Still, however, from the

analysis of the border values τ_1^L , τ_2^L , τ_1^H , and τ_2^H one can easily ascertain that they always satisfy the relation $\tau_1^L < \tau_2^L < \tau_1^H < \tau_2^H$. This implies that there can exist some tax rate τ^* such that $\tau_2^L < \tau^* < \tau_1^H$, and that leads the economy to an equilibrium with informal jobs alone when $\pi k_f > \rho k_i$, but to an equilibrium without the underground sector when $\pi k_f < \rho k_i$. Moreover, if the economy starts at some $\tau^* < \tau_1^L$, and the government wants to achieve a long-run equilibrium without informal jobs, but needs for some purpose to increase taxes, it should be more cautious about raising the level of τ^* too much when $\pi k_f > \rho k_i$ holds, rather than when $\pi k_f < \rho k_i$ maintains. In other words, a "safety margin" for the tax increase is broader in the latter situation. Thus, in general governments can afford to levy higher tax rates on formal firms when $\pi k_f < \rho k_i$, and still drive the economy to an equilibrium without informal jobs. To put it differently, when the bargaining position of formal firms vis-à-vis workers in wage negotiations is stronger in the formal sector, governments can appropriate a larger part of the match surplus by levying higher taxes, and still eventually eradicate informality. Thus, in this case other things being equal, firms appropriate more rents while haggling with workers over wages, while the government takes these gains away through taxation.

The third implication is derived from the factors affecting border values τ_1^L , τ_2^L , τ_1^H , and τ_2^H . All these values depend *inter alia* on the size of the expected fine for running business informally, mF , and costs of bribing in the formal sector, k_f . Thus, the tax rate cannot be "excessively high" *per se*, but rather in relation to the potential fine levied on informal firms or/and an additional burden of bureaucratic extortion in the formal sector. This is well in line with some empirical findings indicating that higher taxes do not necessarily correspond to a higher share of the informal sector (see e.g. Friedman *et*

al., 2000). The contribution of other factors to an increase (or a decrease for that matter) in the size of shadow economy can prevail over the effect of taxes. For instance, Friedman *et al.* (2000) point out that entrepreneurs go underground not to avoid official taxes but to reduce the burden of bureaucracy and corruption. In our model the heavier such a burden is, the higher is k_f , and the more likely for πk_f to be greater than ρk_i . Hence, higher taxes in such a situation are perceived by formal firms as more onerous which makes their move underground more probable. As regards the relative effect of taxes in the formal sector and expected fines in the informal economy, Chapter 3 has already illustrated that it is not the higher taxes *per se* that lead the economy to an equilibrium with informal firms, but rather the burden that they create on formal firms in comparison with factors affecting productivity growth in the shadow economy.

Finally, the last but not the least implication of our model is that the government itself can affect the regime or situation in which it occurs. In other words, governments do possess ways and means of setting relative values of πk_f and ρk_i . Indeed, this can be done, for example, through some structural or administrative reform that would reduce the scope for corruption in the economy, and thus decrease the value of k_f , and, hence, increase the outside option of formal firms in wage bargains with workers. However, a more straightforward and, possibly, less costly method is to affect the effectiveness of the informal sector monitoring, m , which can be accomplished, for example, through financing of the tax police or other similar authorities. Recall that $\rho = \pi + m$ as introduced in Section 4.3.6. Thus, by varying the value of m the government in fact affects the size of the outside option of firms bargaining with workers over wages in the informal sector. It is easy to verify that for all m^L such that $m^L < m^*$, the case $\pi k_f > \rho k_i$ will

result, while for all m^H such that $m^* < m^H$, the case $\pi k_f < \rho k_i$ will take place, where m^* is such that $\pi k_f = \rho k_i$, i.e. $m^* = \frac{\pi(k_f - k_i)}{k_i}$. In this simple model, after choosing a level of monitoring the government can then easily determine the border values of τ and choose a level of taxation, and thus foreordain a desired type of equilibrium in the long run. In particular, for any m^L and m^H such that $m^L < m^* < m^H$ restriction $\tau_1^L(m^L) < \tau_2^L(m^L) < \tau_1^H(m^H) < \tau_2^H(m^H)$ will still maintain. These relationships and the previous implications suggest, other things being equal, that in the economies with a high level of bureaucratic extortion, or/and where higher taxes are favoured by the public office, governments may well need to spend more on the monitoring of the informal sector or/and levy higher fines for running informal businesses. This would make convergence to equilibria without the informal sector more likely. At the same time, in the economies with *a priori* highly effective audit of the informal sector (high m) governments are likely to be more flexible in setting tax rates, as a wider "safety margin" would allow them to achieve formal equilibria with greater probability. It should be noted, however, that in a more realistic modelling the efficiency of audit or monitoring, m , is itself likely to be a function of collected taxes because it can be seen as representing a public good provided by the state. Then the determination of the border values of τ will be much more complicated and will depend on the form of function $m(\tau)$. Implications for government behaviour will also be more complex.

Summary

Let us now summarise the above discussion. Our model indicates that, while keeping other parameters fixed, each type of long-run equilibrium is univocally related to a range of tax rates which does not overlap with similar ranges for other types of equilibria under

that fixed set of parameters. In general, higher tax rates raise the probability that an equilibrium with informal jobs alone will occur. However, the precise effect (the type of equilibrium) of choosing a particular tax rate depends on relative values of reservation options of formal and informal firms in their wage bargaining with workers. So, whenever formal firms enjoy a stronger position in wage bargaining with job-seekers than their informal counterparts do, governments can, in principle, afford to set higher tax rates while achieving the same type of equilibrium, other things being equal. As the reservation option of firms in the formal sector depends on the costs of bribing, it implies that in more corrupt economies firms are less determined to cope with high taxes and more likely to go underground. In other words, higher taxes are more likely to lead to long-run equilibria with the informal sector if the level of bribing is high. Meanwhile, the reservation option of firms in the informal sector depends on the probability of being caught by monitoring authorities, such as the tax police. So, if the level of monitoring of the underground sector is high, higher taxes are less likely to lead to long-run equilibria with the informal sector, i.e. in general governments should be less restricted as to the choice of tax rates in such economies. Thus, an important implication is that the government can itself affect the likely effect of changing the tax rate by influencing the outside options of firms - it can either be done by pursuing reforms aimed at a reduction in severity of corruption in the economy or, which may be easier to accomplish, by choosing some level of monitoring of the informal sector. Finally, while we have discussed the influence of the government on reservation options of firms, we have not said anything about its impact on reservation options of workers. Obviously, our arguments as regards the tax rate can easily be appropriately reformulated for any of the parameters captured by shorthand notation

S_f and S_i , including the unemployment benefits, b_u , - the instrument which affects the value of worker's outside option. However, it should be noted, that as workers search randomly in our set-up, i.e. they do not differentiate between vacancies coming from the formal and informal sectors, a change in b_u affects wage negotiations in both sectors in the same manner. So, focusing on the ability to influence firms' reservation values is far more important.

In this section we have assumed that elimination of the informal sector is what governments want to achieve in the long run. However, this is not necessarily the objective that governments *do* or *should* pursue in reality. We leave the question of what governments actually *do* want to achieve beyond the scope of this essay - the reader can be referred to Buchanan (1975) and Shleifer and Vishny (1998) for some discussions of possible motivation that governments could have. In what follows we rather attempt to provide some insights into whether the elimination of the informal sector can be justified on the efficiency grounds. In particular, for convenience we mainly focus on an interior equilibrium with both formal and informal jobs (such as the one that results in region 2, Fig.4.6, and is shown in Fig.4.4) - a most complicated stable steady state that can occur in our model in the long run. This steady state is also an interim case between two degenerate equilibria with formal and informal jobs alone. We shall assess its efficiency properties and look into what benevolent governments *should* do to achieve efficient or sub-efficient outcomes if they start out in such an equilibrium.

4.4.2 Implications for Welfare

The welfare analysis below is structured in the following way. We begin it by investigating the general conditions under which equilibria in our economy are efficient, i.e. represent first-best optima in terms of maximisation of a welfare function reflecting gains of private agents, i.e. firms and workers. This provides us with understanding of the underlying causes of potential inefficiency. Then we look into the possibility of achieving first-best solutions or, at least, making them more probable. We are interested in what the benevolent government²⁴ can do for correcting the causes of inefficiency. At the same time, we learn the implications of those measures for the presence or the size of the informal sector in steady state. If, however, attainment of first-best allocations is not realistic, it makes sense to analyse what the government can do to lead the economy to a sub-optimal outcome as opposed to a first-best one. In particular, it is interesting whether or not a reduction in size of informality can improve on welfare if the economy starts out in the interior equilibrium.

Before embarking on a substantive part of the efficiency analysis, however, we need to introduce some measure of economic welfare.

Steady state surplus

Consider a private surplus, as customary in search-in-the-labour-market models (see, e.g. Hosios, 1990; Acemoglu and Shimer, 1999; Pissarides, 2000; Acemoglu, 2001):

$$\Xi(\theta, \phi) = N(y - \tau - \pi k_f) + I(y - mF - \rho k_i) - \theta U(\phi r k_f + (1 - \phi) r k_i) + U b_u. \quad (4.26)$$

²⁴Throughout this section we assume that the government is benevolent. In other words, its aim is seen as maximisation of social welfare represented by gains of firms and workers, rather than pursuit of other objectives, such as, for example, maximisation of the budget revenue.

This measure is what firms and workers care about before entering the economy. The surplus is equal to flows of net output generated in the formal and informal sectors plus the income of the unemployed. In particular, it consists of the number of workers in formal jobs, $N = (1 - U) \phi \frac{\delta + m}{\delta + \phi m}$, times the formal sector output net of taxes and flow costs of creating a vacancy, $y - \tau - \pi k_f$, plus the number of workers in informal jobs, $I = (1 - U)(1 - \phi) \frac{\delta}{\delta + \phi m}$, times the informal sector output less expected fines and flow vacancy costs, $y - mF - \rho k_i$, minus the flow costs of job creation to formal and informal vacancies (respectively, $\theta U \phi r k_f$ and $\theta U(1 - \phi) r k_i$), plus total flow benefits received by the unemployed, $U b_u$. Notice that the proportion of formal vacancies among all vacancies, ϕ , and the proportion of filled formal jobs among all filled jobs, $\phi \frac{\delta + m}{\delta + \phi m}$, do not coincide due to different job duration rates in different sectors. Also, it can be shown that $U = \frac{\delta(\delta + m)}{(\delta(\delta + m) + \alpha(\theta)(\delta + \phi m))}$, the stock of unemployed workers in steady state, is in fact a function of θ and ϕ . The derivation of all these results is relegated to Appendix 4.A. Here we shall just note that the private surplus is a measure of what the social planner should be concerned about in steady state, disregarding a specific way that wages are determined in equilibrium. Indeed, as Pissarides (2000) notes, the social planner is not interested in wages, since wages determine only the distribution of output. Distributional considerations, however, are excluded from the social welfare function.

Inefficiency of corner equilibria: a standard result

In order to provide the reader with a useful background to our main results for the economy with two sectors let us start the analysis of efficiency with examination of a corner equilibrium. In such a somewhat degenerate case, whether it is an equilibrium with formal jobs only (Fig.4.2) or the one with informal jobs only (Fig.4.3), it is effectively a one sector

environment. So, not surprisingly, the implications for efficiency of corner equilibria in our model are standard to the literature on search in the labour market (Pissarides, 2000), that for the most part concentrates on an economy with one type of employment.

In a search economy with one sector the efficiency of an equilibrium would imply that the relative amount of search, represented by the ratio of the number of vacancies to the number of job-seekers, is optimal. As we focus here on the decision of firms to enter the economy, the efficiency of equilibrium can be said to imply that the economy creates the right number of jobs. The standard result in the search-in-the-labour-market literature is that the socially optimal amount of search or job creation, for that matter, is achieved if the so-called Hosios (1990) efficiency condition is maintained. In particular, it is known that in search equilibrium in a one sector economy there will be too little job creation if the bargaining power of workers, β , is greater than the elasticity of the matching function, $\eta(\theta) = -\frac{\partial q(\theta)}{\partial \theta} \frac{\theta}{q(\theta)}$, and there will be too much job creation if β is less than the elasticity of the matching function, $\eta(\theta)$. In order the allocation of jobs and workers to be efficient in models of search, first, matching must exhibit constant returns, and, second, the elasticity of the matching function with respect to unemployment must be equal to the bargaining power of workers (Hosios, 1990; Pissarides, 2000). To see that this also applies to the case of a corner equilibrium with formal jobs in our model (the case with informal jobs is easy to analyse by analogy), consider a problem of maximisation of the private surplus²⁵ (4.26). For the infinitely lived economy it consists in maximising $\max_{\theta} \int_0^{\infty} e^{-rt} \Xi(\theta, 1) dt$ or

²⁵Following Pissarides (2000) the approach that we take here in the comparison of the social and private outcomes is derivation of the social outcome by ignoring wage bargaining equations (4.8) and (4.9). Having done that we investigate whether there are wage rates determined according to (4.8) and (4.9) that make the social and private conditions for job creation identical.

$$\max_{\theta} \int_0^{\infty} e^{-rt} (N(y - \tau - \pi k_f) - \theta U r k_f + U b_u) dt \quad (4.27)$$

subject to restrictions

$$\begin{aligned} \frac{dN}{dt} &= \alpha(\theta)U - \delta N, \\ \frac{dU}{dt} &= \delta N - \alpha(\theta)U, \end{aligned} \quad (4.28)$$

while $N + U = 1$.

The integrand in (4.27) is the private surplus (4.26) discounted back to the initial time and when $I = 0$ (or $\phi = 1$). Restrictions (4.28) are standard equations describing the evolution of unemployment and formal sector employment: an increase in a number of workers in a particular labour market state is equal to the flow into that state minus the outflow. The solution of the programme is relegated to the appendix, while here we just give the condition that the socially efficient level of the market tightness must satisfy:

$$S_f = r k_f \left(\frac{\pi + \alpha(\theta) - \theta \frac{\partial \alpha(\theta)}{\partial \theta}}{\frac{\partial \alpha(\theta)}{\partial \theta}} \right). \quad (4.29)$$

At the same time, from zero profit condition (4.18) and from (4.20) it follows that the level of the market tightness in the wage bargaining equilibrium satisfies

$$S_f = r k_f \frac{(\pi + \alpha(\theta) \beta)}{(1 - \beta) q(\theta)}. \quad (4.30)$$

Two expressions (4.29) and (4.30) are identical if and only if $\frac{\partial \alpha(\theta)}{\partial \theta} = (1 - \beta) q(\theta)$ or, recalling properties of $\alpha(\theta)$ and $q(\theta)$, $\beta = \frac{q(\theta) - \frac{\partial \alpha(\theta)}{\partial \theta}}{q(\theta)} = -\frac{\partial q(\theta)}{\partial \theta} \frac{\theta}{q(\theta)} = \eta(\theta)$. In other words,

the corner equilibrium in our model is efficient if and only if the Hosios efficiency condition is satisfied.

An intuition underlying this result is as follows. The elasticity of the matching function $\eta(\theta)$ can be seen as a measure of relative effectiveness of the unemployed in making contacts, while $1 - \eta(\theta)$ is a measure of relative effectiveness of firms in making contacts. From Shaked and Sutton (1984) we know that in search economies with wage bargaining the relative bargaining position of firms vis-à-vis workers depends on the time that it takes a firm to switch, if necessary, from the worker it currently bargains with to a substitute workforce. In models with matching functions this time is a function of $q(\theta)$. The lower is $q(\theta)$ the longer is the time that firms have to wait for a worker, and the weaker is the bargaining position of firms in bargaining with workers who they have already met. In other words, in labour markets characterised by search workers have more monopoly power over extracting the match surplus the longer firms have to wait to fill the vacancy, i.e. the lower is $q(\theta)$, or the greater is θ , for that matter. Thus, as $\eta(\theta)$ reflects the relative effectiveness of the unemployed vis-à-vis firms in searching for a match, it can also be seen as a measure of worker's monopoly power, implied by the labour market conditions, over the surplus of the match to be shared in wage bargains. Hence, if, for instance, $\eta(\theta)$ is higher than β in equilibrium then firms in practice receive a greater share of the surplus in bargaining than they should have done if the split were arranged by the social planner. The asymmetry between the monopoly power in wage bargains commensurate with labour market conditions, $\eta(\theta)$, and the going bargaining power of workers, β , causes firms to open more vacancies in the equilibrium than it is socially optimal. That is, firms take advantage of this asymmetry by overinvesting in job creation. This, of course, would imply

that at the margin firms cause more congestion to each other than the congestion that the unemployed cause to other unemployed workers. In contrast, if $\beta > \eta(\theta)$, firms will underinvest in job creation as the going bargaining power of workers, β , will be greater than the one that would be justified by labour market frictions. Only in the case when $\beta = \eta(\theta)$ the equilibrium is socially optimal, so the Hosios efficiency condition balances out the going and true bargaining positions of workers in wage negotiations. From a slightly different point of view it also internalises all congestion externalities that firms and workers create to each other in labour markets characterised by time-consuming search.

The moral of this exposition is that in general even the simplest outcome in our model, i.e. the corner equilibrium, should not be expected to be efficient. Efficiency hinges upon the particular choice of the bargaining power of workers. Although it is a standard result, from a perspective of the discussion of the welfare impact of policies reducing the size of informality it is important that inefficiency is not caused by a particular combination of policy parameters set by the government, it is essentially of labour market origins.

Now let us investigate the interior equilibrium and see what lessons on the efficiency of the steady state outcome we can learn from there.

Inefficiency of the interior equilibrium

Our economy has two sectors, so that if a non-degenerate outcome occurs with both formal and informal jobs present in equilibrium the concept of efficiency becomes a little more advanced. With two sectors the efficiency of equilibrium implies not only that the right number of jobs is created, but also that jobs are allocated optimally across the sectors. However, our main result below indicates that in contrast to the corner equilibrium case, in the interior equilibrium the allocation of jobs is generally not optimal even when matching

exhibits constant returns (as we have assumed) and the Hosios condition holds. Indeed, it turns out that in the interior equilibrium with both formal and informal jobs such as in Fig.4.4 or 4.5, the restriction on the bargaining power is no longer enough to guarantee efficiency. There exist other factors responsible for creating another type of inefficiency, and which bring about a non-optimal allocation of jobs and workers across the two sectors. With this type of allocative inefficiency firms underinvest in job creation in a particular sector. As a consequence, the amount of job creation in the economy as a whole is not optimal either.

Consider a stable long-run outcome in Fig.4.4, and continue along the lines of the analysis of the corner case above. The problem of maximisation of the private surplus is then given by $\max_{\theta, \phi} \int_0^{\infty} e^{-rt} \Xi(\theta, \phi) dt$ or

$$\max_{\theta, \phi} \int_0^{\infty} e^{-rt} (N(y - \tau - \pi k_f) + I(y - mF - \rho k_i) - \theta U(\phi r k_f + (1 - \phi) r k_i) + U b_u) dt \quad (4.31)$$

subject to constraints:

$$\begin{aligned} \frac{dN}{dt} &= \alpha(\theta)\phi U - \delta N, \\ \frac{dU}{dt} &= \delta N + (\delta + m)I - \alpha(\theta)U, \\ \frac{dI}{dt} &= \alpha(\theta)(1 - \phi)U - (\delta + m)I, \end{aligned} \quad (4.32)$$

while $N + I + U = 1$.

The presence of the informal sector in the interior equilibrium implies that the number of restrictions is increased as compared to (4.28) to account for the flows into and out of

the informal sector. Otherwise, the problem is similar to programme (4.27)-(4.28). The solution to problem (??) subject to restrictions (4.32) is a socially optimal outcome. By inspecting the first order conditions of this programme it is straightforward to ascertain that the interior equilibrium does not in general belong to the set of optimal outcomes. The detailed analysis is relegated to Appendix 4.A, while here we just discuss its main findings and implications.

Main findings The first result obtained in the appendix is that in the interior equilibrium the relationship between the bargaining power of workers and the elasticity of the matching function matters not only for the total amount of job creation but also for distribution of jobs across the two sectors. In particular, *ceteris paribus* if the going bargaining power of workers, β , is less than the bargaining power of workers implied by the labour market conditions, $\eta(\theta)$, firms will overinvest in job creation in the informal sector. If, however, $\beta > \eta(\theta)$ firms will tend to underinvest in creation of informal jobs.

Second, misallocation of jobs across the two sectors in our economy occurs not only when the Hosios condition fails. Even if it maintains and $\beta = \eta(\theta)$, firms tend to underinvest in creation of formal jobs due to differences in start-ups costs such as $k_f > k_i$.

Third, the total amount of job creation in the interior equilibrium is always below the private optimum. In other words, it can be shown that for any choice of bargaining power β there is too little job creation in the economy.

Finally, it is possible to see that, first, the Hosios condition is necessary, but not sufficient for the efficiency of the interior equilibrium in terms of maximisation of the private surplus, and, second, that the sufficient condition for the efficiency is that both

the Hosios condition is met and $k_f = k_i$.²⁶

An implication of these results is that in the two sector environment in general two factors affect both the total amount of job creation in the economy and the allocation of jobs across the formal and informal sectors. First is the standard relation between the bargaining power of workers and the elasticity of the matching function. The second factor is the relation between the start-up costs that firms bear before entering the economy. Let us consider them in turn.

The effect of the relation between β and $\eta(\theta)$ on the total amount of job creation is standard and similar to the case of a one sector environment discussed above. More interesting is its role in creating allocative inefficiencies across the sectors. Our results suggest, intuitively, that firms tend to take advantage of their bargaining position vis-à-vis workers in the sector where surplus of a match is greater. Indeed, recall that in the stable interior equilibrium, as in Fig.4.4, $S_i > S_f$. At the same time, if $\beta < \eta(\theta)$ then firms gain additional rents because of the discrepancy between the going bargaining power of workers and the one implied by the labour market conditions. Those rents are larger in the informal sector, so firms create too many informal jobs. By contrast, if $\beta > \eta(\theta)$ firms miss out on potential rents because of the too high a going bargaining power of workers, β . Those foregone rents are, again, greater in the informal sector, thus, there is underinvestment into the informal sector job creation.

However, even if the Hosios condition holds, i.e. $\beta = \eta(\theta)$, the equilibrium in our economy does not represent a socially optimal allocation of jobs because of the difference in values of k_f and k_i . In general, the presence of start-up costs that firms have to bear uni-

²⁶Note, however, the proof of Proposition 1, Appendix 4.A, shows that when $k_f = k_i$ the stability properties of the interior equilibrium may change. In fact, it then coincides with the bifurcation point separating two regions with different phase dynamics of θ and ϕ .

laterally before entering one sector or another in fact implies that workers gain additional advantage in wage negotiations. Indeed, while firms start wage bargains with some costs already borne, workers enter into negotiations without previously incurring any costs. So, while firms, in the first place, are concerned with recouping the costs, workers are just interested in receiving as high a wage as possible. In such a situation workers are able to capture excessive rents in the process of wage negotiations, so that firms may underinvest in creation of jobs in the economy in general, and in one of the sectors in particular. Obviously, it may well lead to inefficiency of the equilibrium. This result is closely linked with the presence in our framework of a version of the so-called hold-up problem, already mentioned above briefly.

The hold-up problem: a source of inefficiency In relation to investments the hold-up problem arises when one party pays the cost while others share in the payoff. The problem has attracted attention in numerous studies, beginning from the earliest discussions (Williamson, 1975), first formalisation (Grout, 1984), and coinage of the term (Williamson, 1985), to more recent investigation into the causes and possible remedies (e.g. MacLeod and Malcomson, 1993; Malcomson, 1997). The search-in-the-labour-market literature has addressed the problem both as regards firms (Acemoglu and Shimer, 1999; Acemoglu, 2001) and workers (Acemoglu, 1996), who unilaterally invest into either physical or human capital, respectively.

Grout (1984) has shown that the hold-up problem leads to inefficiency of investment arising in the absence of binding contracts, when workers can negotiate wages and employment once the firm has committed itself to a specific investment.²⁷ Furthermore, as

²⁷MacLeod and Malcomson (1993, p.813) state that "in contrast to general investments, specific investments are valuable for trade only with the chosen partner, not for trade with third parties".

pointed out by Malcomson (1997), in the presence of labour turnover costs there can be hold-up of general, as well as specific investments. The underlying cause of hold-ups is incompleteness of contracts. If contracts were to be complete all the parties who benefit from an investment could be made to pay their share of the cost. Acemoglu and Shimer (1999) note that even when contracts are incomplete an appropriate arrangement of relationship between agents can prevent the problem. However, in a situation when investment must be sunk before agents meet contracts and related arrangements are impossible because at the time they invest agents do not know who their trade partner is going to be (Acemoglu, 1996).

If investments made by firms are exogenous to the economy considered, the hold-up problem does not necessarily lead to inefficiency under an appropriate choice of bargaining power of workers. Acemoglu and Shimer (1999) has shown that in, for example, a one sector search environment the standard Hosios condition guarantees a socially optimal outcome. In that case workers still share in returns, but their ability to appropriate rents does not have a negative externality effect on the decision of firms to invest. The only externality that must be internalised in such an economy is the one of the number of vacancies on worker's outside option in the bargaining process. In other words, it is the externality of congestion that firms and workers cause to its peers. As we have mentioned in Section 4.4.2, the Hosios condition achieves precisely that - it eliminates the congestion externality. However, once the investments are made endogenous the Hosios condition is no longer sufficient: even if it is possible to make the level of wages to be equal to the social shadow value of labour, it is impossible to guarantee that both the level and the slope of the wage function are equal to appropriate social values (Acemoglu and Shimer,

1999). With endogenous investment there is a negative externality of worker's bargaining power on firms' investment, which is not internalised even when the Hosios condition is met.

The hold-up problem in our model In our model firms sink non-specific investments, which are treated as exogenous parameters. These costs, k_f and k_i , are incurred before meeting potential employees. Given the non-competitive labour market and wage bargaining, the presence of search frictions and, hence, switching costs (i.e. costs of finding a new trade partner) implies that workers exert monopoly power in sharing the surplus of a match. During wage negotiations workers capture part of the return on k_f and k_i as indicated by equilibrium wage equations (4.23) and (4.24). Thus, we clearly deal with the hold-up problem.

As investments k_f and k_i are non-specific and, importantly, exogenous in size, following Acemoglu and Shimer (1999) at first sight it may seem that we should not expect hold-ups to cause a problem for the efficiency of equilibrium. Indeed, in the corner solution case in Section 4.4.2 the presence of start-up investments unilaterally borne by firms does not affect optimality of the outcome. There is no externality of the worker's bargaining power on firms' investment. At the same time, there is a standard congestion externality which is internalised when the Hosios condition is met.

However, here it is not the fact that the value of k_f and k_i is exogenous that is important for the efficiency. It is the endogeneity of firms' decision to enter a particular sector (and, hence, invest there) that matters. In the case of the corner equilibrium firms really do not have any other option but invest only in one sector, while the expected profits in the other are negative given the current state of the labour market. By contrast, in the

interior equilibrium firms find both the formal and informal sectors equally profitable. So, their allocation across the two sectors is endogenously determined and in such a situation the presence of the different start-up costs will cause the efficiency problem. To see that consider expressions (4.23) and (4.24) rewritten as

$$\frac{(w_f - rE_u(\theta, \phi))}{\pi} = \frac{\beta r k_f}{(1 - \beta) q(\theta)} \quad (4.33)$$

and

$$\frac{(w_i - rE_u(\theta, \phi))}{\rho} = \frac{\beta r k_i}{(1 - \beta) q(\theta)}. \quad (4.34)$$

The terms on the left hand side represent average or expected wage mark-ups that workers are paid on top of their outside option. It can be seen that as $k_f > k_i$ the formal wage mark-up is larger than its informal counterpart. Thus, intuitively, in equilibrium *ceteris paribus* firms will underinvest into formal job creation (and, hence, in the economy as a whole) as in the formal sector they on average forego more rents to workers than in the informal sector. Obviously, such a situation implies that the interior equilibrium is not efficient. Indeed, it can be verified that with $k_f > k_i$ in the interior equilibrium the value of being formally employed, E_f , is always higher than the value of being informally employed, E_i , regardless of the value of β . Thus, firms do not take into account that creating a formal rather than informal vacancy has a positive externality effect on workers. The Hosios condition cannot internalise this externality which works not through the market tightness, θ , but through the proportion of formal vacancies, ϕ .²⁸ Thus, with $k_f \neq k_i$ the Hosios condition is necessary but not sufficient for the efficiency of the equilibrium with

²⁸The reader can be referred to Acemoglu (2001) for the study of a similar uninternalised allocative externality in a model of labour market segmentation.

two sectors, i.e. the one in which the number of firms (and, hence, investments) in a particular sector is endogenously determined.

If, however, $k_f = k_i$ in equilibrium the expected wage mark-ups will be equal across the two sectors, the value of formal employment to workers will be equal to the value of being informally employed, $E_f = E_i$, and the creation of a formal rather than an informal vacancy will not have any effect on worker's welfare. In fact workers will be indifferent as to what proportion of formal vacancies ϕ is achieved in equilibrium. Thus, the allocative externality will disappear and the Hosios condition will again be sufficient to guarantee the efficiency of the interior equilibrium in terms of maximisation of private surplus (4.26).

Summary Thus, under the assumptions that we have made equilibria in our economy are not efficient. In the particular case of the interior equilibrium with both formal and informal jobs both the total amount of jobs in the economy and their allocation across sectors are generally not optimal. Two factors are responsible for this result. They are, first, the relation between the bargaining power of workers and the elasticity of the matching function, and, second, the relation between the costs of entry that firms pay in the two sectors before meeting their employees.

From the results outlined above and detailed in the appendix it follows that the necessary and sufficient condition for the interior equilibrium to be efficient in our model is that both the Hosios condition holds and $k_f = k_i$. If this is not the case, the effects on job creation are as follows. If $\beta < \eta(\theta)$ firms tend to overinvest in job creation in the informal sector because of the discrepancy between the going bargaining power of workers and the one resulting from actual labour market conditions. At the same time, as $k_f > k_i$ firms tend to underinvest in formal job creation. Thus, both factors influencing the cross-sector

allocation of jobs imply that the optimal proportion of formal vacancies must be higher. Meanwhile, they imply that the impact on the total job creation in the economy depends on the relative size of the two effects. It turns out that the tendency to overinvest in the informal sector is smaller than the tendency to underinvest in the formal sector, so on the whole there is too little job creation in the economy. By contrast, if $\beta > \eta(\theta)$, firms tend to underinvest in the informal sector. At the same time, they still have a tendency to underinvest in the formal sector as $k_f > k_i$. Thus, while the effect on the total amount of job creation in the economy is clear - again too little jobs are created - the impact on the sectoral allocation of jobs is, however, ambiguous. Depending on the size of the two effects there can be either too many or too few formal vacancies.

These results are important not only from the perspective of a better understanding of functioning of economies with a particular formal-informal labour market segmentation. They also shed light on the issue of equilibrium efficiency in two-sector search economies of a more general nature. The previous theoretical literature has mainly studied only one sector economies in that regard, while welfare implications in two-sector search models have largely not been analysed. The few exceptions are Davidson *et al.* (1987) and Acemoglu (2001). In the former work authors suggest a two-sector model with one search and one competitive sectors. They show that in such a framework firms underinvest in the search sector. In this respect our model goes one step further by considering two search sectors instead of one. In the second study Acemoglu (2001) stresses the role of capital investment hold-ups in creating allocative inefficiencies. However, he does not say anything about the role of the relation between the bargaining power of workers and the elasticity of the matching function in misallocation of jobs across the sectors. In

this regard our model contributes by pointing to the fact that both the Hosios condition and the absence of asymmetries in investment hold-ups across sectors are important for efficiency of the steady state equilibrium with two jobs, both from the point of view of the total amount of job creation and the distribution of jobs across sectors. Interestingly, here we do not talk about hold-ups of capital investments, but rather of the costs associated with the extortionary activities of the government.

The interior equilibrium in our model is never efficient because $k_f > k_i$. Exactly this factor makes for the total amount of job creation in the economy being below optimal. This sends a clear message for economic policy that we shall discuss in the next section.

Achieving optimal and sub-optimal allocations

Having established the conditions that make for first-best outcomes in our economy we can now turn to considering efficiency grounds for the reduction in the size of the informal sector.

The first implication that we can derive from the above is that in the interior equilibrium the spectre of inefficiency is raised as compared to a corner equilibrium. This, as we have seen, is explained by the possibility of an additional allocative inefficiency that can be caused either by asymmetries in opportunities for hold-up of start-up investments across sectors or by the violation of the Hosios efficiency condition, or both. Thus, on these grounds for a benevolent government aiming to attain a first-best allocation of jobs in the economy it can be reasonable to accept such a set of policies that would drive the economy to a corner equilibrium with formal jobs only. In that case, the scope for inefficiency of the resulting steady state outcome is narrowed and depends solely on whether or not the Hosios efficiency condition is violated.

Alternatively, the benevolent government can in the first place directly address a source of inefficiency, rather than be concerned with the type of equilibrium resulting in the long term. Indeed, imagine a situation when the economy starts out in the interior equilibrium, while the attainment of the corner equilibrium with formal jobs alone implies, for example, a non-realistic decrease in the tax rate (recall its effect on the long-run outcome as explained in Section 4.4.1). We have seen that the efficiency properties of equilibrium essentially depend on the state of the labour market in general, but not on whether or not the informal sector is eventually present in the steady state. Thus, the question is whether or not governments can create suitable labour market conditions that would make first-best outcomes more probable.

Fighting corruption While there is no guarantee that the Hosios condition will hold in reality (see, e.g. Stevens, 2004), the obvious area where actions of the governments can make a significant contribution and reduce the scope for inefficiency is elimination or mitigation of hold-up opportunities. As follows from the previous section, the interior equilibrium is never efficient because of the differences in values of start-up costs k_f and k_i that lead to misallocation of jobs and workers. Thus, reducing the difference between k_f and k_i would decrease the scope for allocative failures. In the context of our model, the greater value of formal sector entry costs, k_f , in relation to informal sector costs, k_i , is caused by malversation at the low level of government, i.e. by the so-called administrative corruption. Thus, a pursuit of structural administrative reforms aimed at curbing bribery and low-level corruption should level off hold-up opportunities across sectors and, hence, leave less room for inefficiencies in allocation. Importantly, as follows from Section 4.4.1, a reduction in k_f increases the probability of ending up in an equilibrium without the

Table 4.1: Effect of parameter changes in the stable interior equilibrium

Increase in	Effect on				
	ϕ	θ	N	I	U
m	+	-	+	-	-
F	+	-	+	-	-
τ	-	+	-	+	+
b_u	-	0	-	+	+
k_f	-	+	-	+	+

informal sector too. However, notice that from the current perspective on the issue of why k_f should be reduced, a possibility of elimination of the informal sector in the long run accompanies but not causes an increase in economic welfare. Speaking differently, other policies, apart from those affecting k_f , that may result in a decrease in size or complete eradication of the informal sector, may not necessarily bring about an improvement in welfare. We shall shortly illustrate this argument more convincingly.

It should be noted, however, that in practice fighting bribery is likely to demand a great deal of resources and be time-consuming. Some studies of corruption (e.g. Wei, 1999) suggest that its various forms are highly correlated, so that the presence of corruption at the low level of government may well be an indication of a "grabbing hand" (Shleifer and Vishny, 1998) government as a whole. Politicians themselves may enact extortionary taxation laws, mismanage state budget, and so forth, in order to create abilities to extract rents. De Soto (1989) notes that many permits, licenses and regulations exist probably to give venal officials the power to deny them and to collect bribes in return for providing the permits. Thus, interaction between different levels of the public office is likely to imply that it will balk at the prospect of structural reforms.

Reduction in informality and efficiency Attainment of the corner equilibrium with formal jobs alone or fighting corruption make first-best outcomes more likely in our econ-

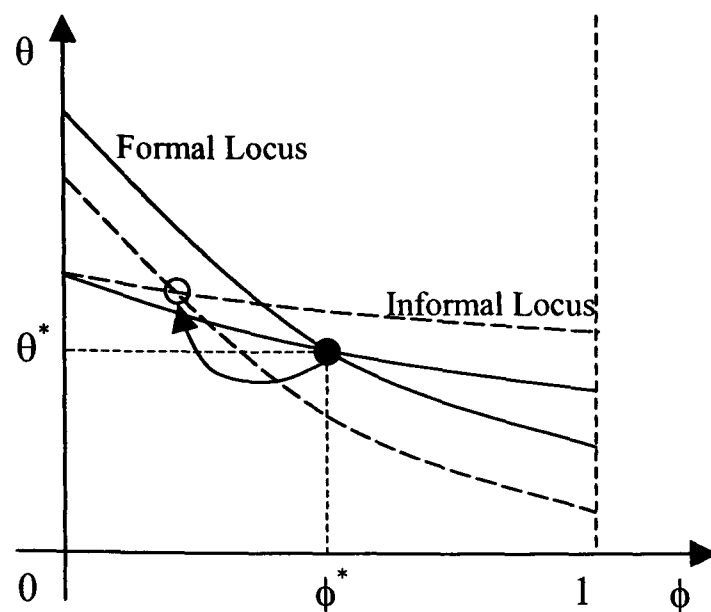


Figure 4.7: The effect of an increase in τ or k_f

omy. However, the complete eradication of the informal sector on the efficiency grounds or the elimination of preconditions for allocative inefficiency may well be unrealistic tasks because of their immensity. Let us consider now a more modest objective of improving on the current allocation of jobs if the economy starts out in the interior equilibrium. The matter of interest is what the benevolent government can do as to achieve at least some sub-optimal level of welfare if first-best solutions are not probable. In other words, should the government try to increase or decrease the size of the informal sector given that initially both types of job exist in steady state?

The policy parameters that affect the size of the informal sector in the interior equilibrium are the tax rate, τ , unemployment benefits, b_u , the fine rate, F , the efficiency of monitoring, m , and the level of entry costs in the formal sector, k_f . Consider, for example, an increase in the tax rate. Intuitively, recalling the analysis in Section 4.4.1, such a change should shift the equilibrium allocation of jobs to the left of the initial equilibrium position, so that the economy gets nearer to the purely informal equilibrium (Fig.4.7).

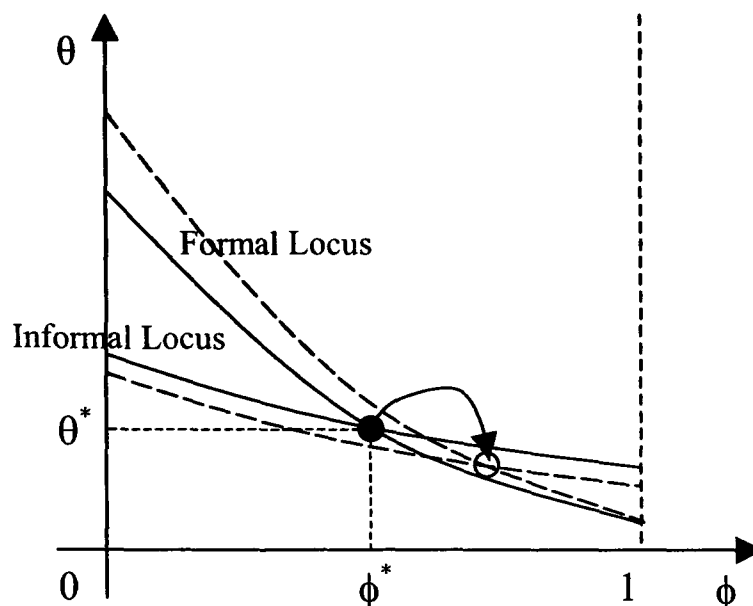


Figure 4.8: The effect of a rise in m or F

If, however, the tax rate is cut down, the economy is to move to the right, so that the equilibrium with formal jobs alone gets closer. Assuming that all other parameters are held fixed, table 4.1, Fig.4.7 and 4.8, and Appendix 4.A illustrate in somewhat greater detail the effects of such changes in taxes; unemployment benefits, the fine rate, monitoring efficiency, and the level of entry costs in the formal sector on the equilibrium level of market tightness, θ , the proportion of formal vacancies, ϕ , the number of the unemployed, U , as well as the formally, N , and informally, I , employed.²⁹

Consider now the question of whether or not a decrease in the share of informal jobs prompted, for example, by a higher F (Fig.4.8), leads to a higher level of welfare. The impact depends on how large is the effect of policies on θ relative to the effect on ϕ . Recall that N , U , and I are functions of θ and ϕ , so substituting for them in (4.26) and using the implicit function theorem gives a relationship between θ and ϕ , drawn as a dashed line

²⁹While the upshot of such policies is generally in line with results obtained in many other studies of formal-informal segmentation, our model, however, indicates that a *reduction* in the level of unemployment benefits would increase the proportion of formal vacancies and decrease unemployment, while keeping the equilibrium level of market tightness constant. See Section 4.A.3 in the appendix for a comparison of this finding to the results obtained in some other models of the informal sector with non-competitive labour markets.

in Fig.4.9, along which the private surplus is constant. An analysis of this relationship at the equilibrium point suggests that shifts of this curve towards northeast bring about a higher surplus. However, this curve can be flatter (the dashed line PS_1 in Fig.4.9), steeper (PS_2) than both the locus of formal jobs (4.18) and the locus of informal jobs (4.19), or it can be flatter than one locus but steeper than the other one. Depending on the case occurring welfare implications of implementing the same policy can be very different. For example, when the private surplus in equilibrium is given by line PS_1 in Fig.4.9, an increase in F would decrease the level of welfare (compare Fig.4.8 and 4.9). By contrast, if the private surplus is represented by line PS_2 , the same policy unambiguously improves on the efficiency of equilibrium. As it is shown in the appendix the slope of the private surplus curve in the (θ, ϕ) -plane depends on the relative values of k_f and k_i , and the value of the bargaining power of workers, β . In particular:

Remark 2 *The higher is k_f as compared to k_i , and/or the lower is β as compared to the elasticity of the matching function, $\eta(\theta)$, the steeper is the PS curve in Fig.4.9. In other words, it is more likely for this curve to cut both loci (4.18) and (4.19) from above in the (θ, ϕ) -plane, so that the policies aimed at the eradication of the informal sector are more likely to improve on the economic welfare measured by the size of the private surplus.*

This remark suggests that when the level of corruption in the economy is high (which reflects in higher k_f), and/or when the bargaining power of firms vis-à-vis workers is strong (lower β), it makes sense to try to reduce the share of the informal sector. Indeed, as follows from the results in Section 4.4.2, both higher k_f and lower β exacerbate underinvestment in creation of formal jobs and overinvestment in creation of informal jobs, respectively. On the one hand, firms avoid the formal sector where they have to give up more rents to

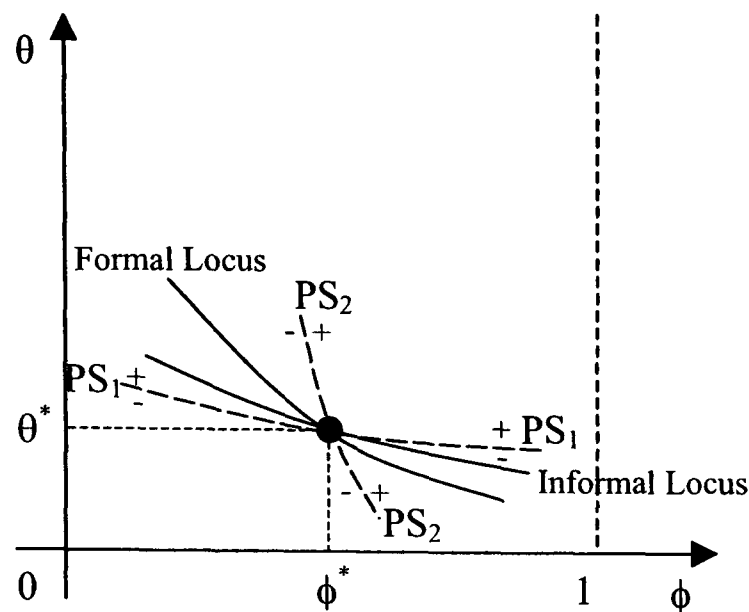


Figure 4.9: Private surplus and interior equilibrium

the workers during wage bargains. On the other hand, they aim to make the most of their stronger bargaining position in the informal sector, other things being equal. Thus, under these conditions the allocative failure to distribute jobs properly towards the formal sector is more acute, so that policies making investments in creation of informal workplaces less attractive would be more likely to increase the level of economic welfare. By contrast, if the difference between k_f and k_i is negligible, whereas β is significantly greater than $\eta(\theta)$ a reduction in size of informality prompted by a change in τ , F , etc., is more likely to be detrimental for economic welfare.

Obviously, a precise set of policies pursued by the government will depend on the fiscal stance and other restrictions.³⁰ By applying the same approach as we did above to the private surplus, it is possible to show that the budget revenue

³⁰So, for example, a nature of the government obviously affects policies it opts for. A "benevolent dictator" (as discussed in Buchanan, 1975) is likely to act differently from a "grabbing hand" government (as in Shleifer and Vishny, 1998). Also, political objectives may and often do differ depending on the part of the Laffer curve that the economy currently sits on (if, for instance, revenue maximisation is assumed). Buchanan and Lee (1982) illustrate this by drawing distinctions between short-run and long-run Laffer curves.

$$R(\theta, \phi) = N\tau + ImF - Ub_u \quad (4.35)$$

can also be represented as a negatively-sloped line in the (θ, ϕ) -plane. Its slope will be the larger in absolute value the larger is the tax rate τ in comparison with the expected value of the fine in the underground sector, mF . Thus,

Remark 3 *The higher is τ as compared to mF the more likely for policies reducing the proportion of informal vacancies to result in the higher budget revenue.*

It should be clear that such policies can lead to an increase in both budget revenue (4.35) and private surplus (4.26), so that the total gain of both the private agents and the government rises. The result will depend on the relative slopes of the budget revenue level curve and the PS curve in the (θ, ϕ) -plane.

Summary

In this section we have generally addressed the issue of what governments should do to improve on the efficiency of equilibrium allocation of jobs and workers across the formal and informal sectors. We have split the analysis in two steps.

Firstly, we have shown that inefficiency of equilibria is inherent in our model. Any steady state is not optimal due to congestion externalities of a standard type. In order to internalise such externalities, the so-called Hosios efficiency condition (Hosios, 1990), equating the bargaining power of workers and the elasticity of the matching function, must hold. However, if the economy ends up in a steady state with both the formal and informal sectors, such a long-run equilibrium suffers from an additional market failure caused by the

so-called hold-up problem (Grout, 1984; Williamson, 1985). The preconditions for hold-ups in our model are exacerbated by the lower tier of the government - corrupt bureaucracy - that extorts rents from firms entering the formal market. In this situation the Hosios condition is no longer enough to guarantee the attainment of first-best optimum, as it can never internalise the allocative inefficiency. In turn, because of that firms underinvest in creation of formal vacancies, while workers gain higher rents in the formal economy during wage bargaining. Importantly, the inefficiency of equilibria that we have considered here is not provoked by a particular size of the informal economy, but rather by particular conditions in the labour market. Thus, in order to make first-best outcomes more probable the government has to address causes of inefficiencies but not to try to scale down the informal sector in the first place. Still, reduction in the size of the latter can come about as a by-product of particular efficiency-improving policies.

As a second exercise we have looked into the ability of the government to attain sub-optimal levels of welfare if the economy starts out in the interior equilibrium. We have especially been interested whether a reduction in informality in such a situation can be efficiency improving. It turned out that welfare effect of such policies as changes in the tax rate, fines for running businesses informally, unemployment benefits, or monitoring efficiency, depends on the same factors which lead to inefficiency of the equilibrium. In particular, the stronger is the position of firms vis-à-vis workers in wage bargains, and/or the better are the opportunities for hold-up in the formal sector, the more acute is underinvestment in formal job creation. So, the more likely for the reduction in the share of informal vacancies to bring about an improvement in economic welfare. Thus, the main conclusion that can be drawn from this discussion is that conditions in the labour

market do affect welfare implications of policies. An important result is that the relation between the bargaining power of workers and the elasticity of the matching function must be taken into account before the decision to reduce the scope of informality. However, the previous literature (Kolm and Larsen, 2001; Bouev, 2002; and Fugazza and Jacques, 2004), that incorporates the search-in-the-labour-market approach to the modelling of the formal-informal segmentation and that assesses welfare implications of various policies, has completely overlooked the considerations outlined here.

4.5 Concluding Remarks

In this work we have broadly pursued two main objectives.

First, it has been shown that labour market institutions, such as wage bargaining may be a very important channel through which state regulations lead to stable emergence of informality in the long run. Labour costs are widely held to be one of the main culprits of causing a firms' drive of moving businesses underground, however, the issue of wage negotiations have not as yet been addressed in this respect.

The model presented in this essay suggests that general conditions for the emergence of long-term mixed equilibria with the formal and informal sectors imply that the two negotiating sides, i.e. firms and workers, have to prefer different sectors if given full power to appropriate the whole surplus of a productive match. Such a difference in preferences may be possible due to differences in both sector surpluses and the outside options of firms and workers. The former are caused by standard government regulations such as taxation, penalties for involvement into irregular activities, etc. The latter are linked with both the impact of the government's auditing activities whereby they destroy informal matches and

cause asymmetries in match lifetimes across the two sectors, and the venal practices of the state bureaucracy extorting bribes and erecting artificial barriers to entry. Costly time-consuming search for trade partners acts as a catalyst of the effects of government audit and bureaucratic extortion, as it raises opportunity costs of idleness of sunk investments and leads to rent sharing.

The role of wage bargaining and rent sharing in the emergence of equilibria with the informal sector has important implications for policy-makers - a second major issue analysed in this essay. We have illustrated many implications by considering such a standard policy tool as the tax rate. It has been found that, in general, higher tax rates raise the probability of an equilibrium with informal jobs alone. However, the precise effect (i.e. the type of equilibrium that the economy eventually attains) of choosing a particular tax rate depends on relative values of reservation options of formal and informal firms in their wage bargaining with workers. More specifically, other things being equal governments can afford setting higher tax rates and still achieve a long-run equilibrium without informality whenever formal firms enjoy a stronger position in wage bargains with job-seekers than their informal counterparts do. This significance of firms' reservation options in bargaining for the effect of policies is an important implication for policy making that has hardly been given any attention in the previous literature on the topic. As we have seen, governments can affect these outside options of firms in both formal and informal parts of the economy, and thus predetermine the effect that a change in the tax rate will have on the long-term outcome. Such an impact on the reservation options can be achieved either through policies aimed at the reduction in the scope for bribing in the formal sector or through policies raising efficiency of monitoring of informal jobs.

Another issue of major importance for consideration of potential policies is the welfare impact of various actions of the government, in particular those aimed at achieving a reduction in size of informality. In analysing this impact we have ascertained that first-best outcomes are generally not guaranteed in our model. It has been shown that inefficiency of equilibria is caused by general labour market conditions and not linked to the presence or a particular size of the informal sector as such in the steady state. The inefficiency is brought about through rent sharing and *inter alia* related to the presence of the so-called hold-up problem in our set-up. This problem arises when workers have enough bargaining power to share in payoffs on investments unilaterally made by firms before these meet with their potential employees. The novelty of this work is that it shows that hold-ups can effectively emerge not only in the presence of investments in capital, but in the presence of expenses that firms have to pay at the stage of entry to suborn venal officials. In our economy hold-ups cause allocative inefficiencies, i.e. misallocation of jobs and workers across the formal and informal sectors when they both exist in equilibrium. Thus, we suggest that fighting corruption is more important than reducing the size of the informal sector in the first place - the reduction, however, can come about as a by-product of policies aimed at administrative reforms and preventing bribing practices. This supplements other theoretical work (see, e.g. Murphy *et al.*, 1993; Shleifer and Vishny, 1998; Sarte, 2000; etc.) that also indicates that reduction in corruption brings more entrepreneurs into the formal sector. Here we have especially highlighted the efficiency-improving aspect of such policies.

It is also interesting that while in general governmental actions scaling down informality can lead the economy to a better, though not a first-best outcome, it is not necessarily

always the case. We have shown that conditions in the labour market must be taken into account when evaluating the effect of such policies. In particular, it has been found that the relation between the bargaining power of workers and the elasticity of the matching function, as well as differences in hold-up opportunities across sectors bear on welfare implications of identical sets of policies. If, for example, the bargaining power of workers is relatively low (less than the elasticity of the matching function), while the level of corruption in the economy is high (which is reflected in the relative values of start-up costs in the formal and informal sectors), then it is more likely for the policies cutting down the size of the informal sector to result in a higher level of economic welfare. This is explained by the fact that the stronger the firms are vis-à-vis workers in wage bargaining, and the more onerous bribing is, the stronger the firms' drive to open informal rather than formal jobs is. Thus, they overinvest in job creation in the informal economy and cause greater misallocation, which should be corrected through a reduction in size of the shadow sector. Interestingly, in many states, successors of the former Soviet Union the bargaining power of workers in wage negotiations with firms is believed to be negligible. Although there is evidence on rent appropriation by workers (Shakhnovich and Yudashkina, 2001), employees often have nearly no vote and are bound to accept the conditions of their employer. In such a situation our model does suggest that policies reducing the share of the shadow economy are likely to somewhat increase the private sector surplus.

To conclude we note that this work has also provided some useful hints on directions of future research. In particular, although not discussed at length in the main text, it has been found that higher unemployment benefits in the model with the formal and informal sectors and wage bargaining may lead to an increase in size of the informal sector - a

surprising result at first sight, which, however, concurs well with some evidence from Eastern Europe. We believe that analysis of the effects of unemployment compensation on economic welfare, government revenue, formal and informal employment within the framework put forward in this essay deserves further attention and should form the basis for investigation in the future.

4.A Proofs, Main Technical Results, and Further Analysis

4.A.1 Some Core Elements of the Model

This section of the appendix presents some core technical results that are referred to in the main text.

Bellman functions for workers

The system of Bellman equations for filled jobs and vacancies (4.1)-(4.3), workers (4.5)-(4.7), wage determination rules (4.8) and (4.9), and zero profit conditions (4.10) and (4.11) imply:

$$\begin{aligned} rE_f &= \frac{(r\beta(\rho S_f + \beta(1-\phi)\alpha(\theta)(S_f - S_i)) + \pi\rho b_u + \beta\alpha(\theta)(\rho\phi S_f + \pi(1-\phi)S_i))}{\pi\rho + \alpha(\theta)\beta((1-\phi)\pi + \phi\rho)}, \\ rE_i &= \frac{(r\beta(\pi S_i - \beta\phi\alpha(\theta)(S_f - S_i)) + \pi\rho b_u + \beta\alpha(\theta)(\rho\phi S_f + \pi(1-\phi)S_i))}{\pi\rho + \alpha(\theta)\beta((1-\phi)\pi + \phi\rho)}, \\ rE_u &= \frac{\pi\rho b_u + \beta\alpha(\theta)(\rho\phi S_f + \pi(1-\phi)S_i)}{\pi\rho + \alpha(\theta)\beta((1-\phi)\pi + \phi\rho)}. \end{aligned}$$

In the analysis in the main text we are interested in the properties of function $E_u(\theta, \phi, \cdot)$. It is easy to verify that $\frac{\partial E_u(\theta, \phi)}{\partial \theta} > 0$, namely:

$$\frac{\partial E_u(\theta, \phi)}{\partial \theta} = \frac{\partial \alpha(\theta)}{\partial \theta} \beta \pi \rho \frac{\phi \rho S_f + (1-\phi)\pi S_i}{r(\pi\rho + \alpha(\theta)\beta((1-\phi)\pi + \phi\rho))^2} > 0,$$

as $S_f > 0$ and $S_i > 0$ by assumption. At the same time, the sign of $\frac{\partial E_u(\theta, \phi)}{\partial \phi}$ depends on the relative value of E_f and E_i . The derivative is equal to

$$\frac{\partial E_u(\theta, \phi)}{\partial \phi} = \frac{\alpha(\theta)\pi\rho(E_f - E_i)}{r(\pi\rho + \alpha(\theta)\beta((1-\phi)\pi + \phi\rho))} = \alpha(\theta)\beta\pi\rho \frac{(\rho S_f - \pi S_i) + \alpha(\theta)\beta(S_f - S_i)}{r(\pi\rho + \alpha(\theta)\beta((1-\phi)\pi + \phi\rho))^2}$$

Consequently,

- a) if $S_f - S_i > 0$ then $\frac{S_f}{\pi} > \frac{S_i}{\rho}$ and $E_f > E_i$, and hence $\frac{\partial E_u(\theta, \phi)}{\partial \phi} > 0$ (region 1, Fig.4.1);
- b) if $\frac{S_f}{\pi} < \frac{S_i}{\rho}$ then $S_f - S_i < 0$ and $E_f < E_i$, and hence $\frac{\partial E_u(\theta, \phi)}{\partial \phi} < 0$ (region 3, Fig.4.1);

c) if $\frac{S_f}{\pi} > \frac{S_i}{\rho}$ and $S_f - S_i < 0$ then $E_f < E_i$ when $\theta \rightarrow \infty$. and $E_f > E_i$ when $\theta \rightarrow 0$. so that $\frac{\partial E_u(\theta, \phi)}{\partial \phi} < 0$ when $\theta \rightarrow \infty$, and $\frac{\partial E_u(\theta, \phi)}{\partial \phi} > 0$ when $\theta \rightarrow 0$ (region 2. Fig.4.1).

Propositions 1 and 2 pay more attention to the case c) above.

Functions $\Pi_f(\theta, \phi)$ and $\Pi_i(\theta, \phi)$

The possibility of an interior equilibrium with both formal and informal jobs depends on characteristics of the two curves in the (θ, ϕ) -plane, namely the loci of formal (4.18) and informal (4.19) jobs. These in turn depend on the properties of two functions (4.16) and (4.17), reproduced here for ease of reference:

$$\Pi_f(\theta, \phi) = \frac{(1 - \beta) q(\theta)}{(r + q(\theta)) \pi} \left(y - \tau - \pi k_f - \frac{r \pi k_f}{(1 - \beta) q(\theta)} - r E_u(\theta, \phi) \right),$$

and

$$\Pi_i(\theta, \phi) = \frac{(1 - \beta) q(\theta)}{(r + q(\theta)) \rho} \left(y - mF - \rho k_i - \frac{r \rho k_i}{(1 - \beta) q(\theta)} - r E_u(\theta, \phi) \right).$$

Function $E_u(\theta, \phi)$ is given in (4.20). Its properties imply that it belongs to the interval between $\frac{b_u}{r}$ and $\frac{1}{r} \min(S_f + b_u, S_i + b_u)$, with $\frac{\partial E_u(\theta, \phi)}{\partial \theta} > 0$ whenever $S_f > 0$ and $S_i > 0$. The behaviour of $\frac{\partial E_u(\theta, \phi)}{\partial \phi}$ has been briefly analysed above.

$\Pi_f(\theta, \phi) = 0$ is equivalent to $y - \tau - \pi k_f - r E_u(\theta, \phi) = \frac{r \pi k_f}{(1 - \beta) q(\theta)}$. Then from the formal sector zero profit condition it follows that:

$$\frac{\partial \Pi_f(\theta, \phi)}{\partial \theta} \Big|_{\Pi_f(\theta, \phi)=0} = r \frac{\frac{\partial q(\theta)}{\partial \theta} \pi k_f - q^2(\theta) (1 - \beta) \frac{\partial E_u(\theta, \phi)}{\partial \theta}}{q(\theta) (r + q(\theta)) \pi}.$$

So, given $\frac{\partial E_u(\theta, \phi)}{\partial \theta} > 0$ and $\frac{\partial q(\theta)}{\partial \theta} < 0$ we unambiguously have $\frac{\partial \Pi_f(\theta, \phi)}{\partial \theta} \Big|_{\Pi_f(\theta, \phi)=0} < 0$.

By analogy, the derivative with respect to ϕ is

$$\frac{\partial \Pi_f(\theta, \phi)}{\partial \phi} \Big|_{\Pi_f(\theta, \phi)=0} = - (1 - \beta) q(\theta) r \frac{\frac{\partial E_u(\theta, \phi)}{\partial \phi}}{(r + q(\theta)) \pi}.$$

The sign of this partial derivative is opposite to the sign of $\frac{\partial E_u(\theta, \phi)}{\partial \phi}$.

In a similar vein, the informal sector zero profit condition is $\Pi_i(\theta, \phi) = 0$, which is equivalent to $y - mF - \rho k_i - r E_u(\theta, \phi) = \frac{r \rho k_i}{(1 - \beta) q(\theta)}$. Then

$$\frac{\partial \Pi_i(\theta, \phi)}{\partial \theta} \Big|_{\Pi_i(\theta, \phi)=0} = r \frac{\frac{\partial q(\theta)}{\partial \theta} \rho k_i - q^2(\theta) (1 - \beta) \frac{\partial E_u(\theta, \phi)}{\partial \theta}}{q(\theta) (r + q(\theta)) \rho}.$$

Again we unambiguously have $\frac{\partial \Pi_i(\theta, \phi)}{\partial \theta} \big|_{I(\theta, \phi)=0} < 0$, while the partial derivative with respect to ϕ has a sign opposite to the sign of $\frac{\partial E_u(\theta, \phi)}{\partial \phi}$:

$$\frac{\partial \Pi_i(\theta, \phi)}{\partial \phi} \big|_{\Pi_i(\theta, \phi)=0} = - (1 - \beta) q(\theta) r \frac{\frac{\partial E_u(\theta, \phi)}{\partial \phi}}{(r + q(\theta)) \rho}.$$

These results are used in Section 4.3.6 of the main text for determining the relative slopes and positions of two loci (4.18) and (4.19) in the (θ, ϕ) -plane (see Fig.4.2-4.5).

Stocks and proportions

In Section 4.4 of the main text in the formula for private surplus (4.26) we use expressions for the stocks of workers in steady state. From Bellman equations (4.5)-(4.7) it follows that stocks at steady state must satisfy:

$$\begin{aligned} \delta N &= \alpha(\theta) \phi U, \\ (\delta + m) I &= \alpha(\theta) (1 - \phi) U, \\ \alpha(\theta) U &= \delta N + (\delta + m) I, \\ 1 &= I + N + U. \end{aligned}$$

The solution to this system is

$$\begin{aligned} U &= \frac{\delta (\delta + m)}{\delta (\delta + m) + \alpha(\theta) (\delta + \phi m)}, \\ I &= \frac{\delta \alpha(\theta) (1 - \phi)}{\delta (\delta + m) + \alpha(\theta) (\delta + \phi m)}, \\ N &= \frac{\alpha(\theta) \phi (\delta + m)}{\delta (\delta + m) + \alpha(\theta) (\delta + \phi m)}. \end{aligned}$$

Thus, it is straightforward to see that the proportion of filled formal jobs in the total number of filled jobs is not equal to ϕ , but given by

$$\frac{N}{N + I} = \frac{\frac{\alpha(\theta) \phi U}{\delta}}{\frac{\alpha(\theta) \phi U}{\delta} + \frac{\alpha(\theta) (1 - \phi) U}{(\delta + m)}} = \phi \frac{\delta + m}{\delta + \phi m}.$$

Analogously, the proportion of informal jobs in the total number of jobs is

$$\frac{I}{N+I} = \frac{\frac{\alpha(\theta)(1-\phi)U}{(\delta+m)}}{\frac{\alpha(\theta)\phi U}{\delta} + \frac{\alpha(\theta)(1-\phi)U}{(\delta+m)}} = (1-\phi) \frac{\delta}{\delta + \phi m}.$$

All these results are useful for the analysis of dynamic maximisation problems (see below).

4.A.2 Proof of Propositions and Analysis of the Bifurcation Point

The variant of the model considered in the main text assumes that $k_f > k_i$. This assumption bears on the conditions under which an interior equilibrium with formal and informal jobs exists and is stable. Here we provide proofs of the two propositions from the main text that help explain the behaviour of $\frac{\partial E_u(\theta, \phi)}{\partial \phi}$ and, hence, two loci (4.18) and (4.19) as can be seen from (4.21) and (4.22). Then we consider the possible abolition of the assumption that $k_f > k_i$, and look at the consequences of such a change.

Proposition 1

Proposition 1. There exists some threshold value of the market tightness $\bar{\theta}$, defined by parameters $k_f, k_i, \beta, r, \delta, b_u, \tau, m, F$, and parameters of the matching function, such that for parameter values satisfying conditions in region 2, Fig.4.1, and for any $\theta > \bar{\theta}$ derivative $\frac{\partial E_u(\theta, \phi)}{\partial \phi}$ is negative, and for any $\theta < \bar{\theta}$ it is positive.

Proof:

By definition $E_u(\theta, \phi) = \frac{\pi \rho b_u + \beta \alpha(\theta) (\rho \phi S_f + \pi (1-\phi) S_i)}{r(\pi \rho + \alpha(\theta) \beta ((1-\phi)\pi + \phi \rho))}$, with $\frac{\partial E_u(\theta, \phi)}{\partial \theta} > 0$ (see above).

The restrictions on parameter values in region 2, Fig.4.1, imply $\frac{S_f}{\pi} > \frac{S_i}{\rho}$ and $S_f - S_i < 0$, so that $\frac{\partial E_u(\theta, \phi)}{\partial \phi} < 0$ when $\theta \rightarrow \infty$, and $\frac{\partial E_u(\theta, \phi)}{\partial \phi} > 0$ when $\theta \rightarrow 0$.

More concretely, as $\frac{\partial E_u(\theta, \phi)}{\partial \phi} = \alpha(\theta) \beta \pi \rho \frac{(\rho S_f - \pi S_i) + \alpha(\theta) \beta (S_f - S_i)}{r(\pi \rho + \alpha(\theta) \beta ((1-\phi)\pi + \phi \rho))^2}$, we have:

- a) $\frac{\partial E_u(\theta, \phi)}{\partial \phi} > 0$ if $\rho S_f - \pi S_i > -\alpha(\theta) \beta (S_f - S_i)$;
- b) $\frac{\partial E_u(\theta, \phi)}{\partial \phi} < 0$ if $\rho S_f - \pi S_i < -\alpha(\theta) \beta (S_f - S_i)$.

Or, alternatively,

- a) $\frac{\partial E_u(\theta, \phi)}{\partial \phi} > 0$ if $\alpha(\theta) < -\frac{\rho S_f - \pi S_i}{\beta (S_f - S_i)}$, and
- b) $\frac{\partial E_u(\theta, \phi)}{\partial \phi} < 0$ if $\alpha(\theta) > -\frac{\rho S_f - \pi S_i}{\beta (S_f - S_i)}$.

Define $\bar{\theta}$ as $\bar{\theta} = \alpha^{-1} \left(-\frac{\rho S_f - \pi S_i}{\beta(S_f - S_i)} \right)$. The argument in the parentheses is positive, hence from properties of function $\alpha(\theta)$ it follows that $\bar{\theta}$ always exists, and $\frac{\partial E_u(\theta, \phi, \cdot)}{\partial \phi} > 0$ if $\theta < \bar{\theta}$, while $\frac{\partial E_u(\theta, \phi, \cdot)}{\partial \phi} < 0$ if $\theta > \bar{\theta}$.

Q.E.D.

The following proposition establishes that under condition $k_f > k_i$, the point of interior equilibrium falls into the region where $\frac{\partial E_u(\theta, \phi, \cdot)}{\partial \phi} > 0$, i.e. the value of being unemployed rises as the proportion of formal vacancies increases.

Proposition 2

Proposition 2. Let $\bar{\theta}$ be a threshold value of market tightness such that $\frac{\partial E_u(\bar{\theta}, \phi)}{\partial \phi} = 0$, and let (θ^*, ϕ^*) be a point of an interior equilibrium in region 2, Fig.4.6. Then given $k_f > k_i$ and $\rho > \pi$, θ^* is always less than $\bar{\theta}$.

Proof:

From Proposition 1 it follows that $\alpha(\bar{\theta}) = -\frac{\rho S_f - \pi S_i}{\beta(S_f - S_i)} > 0$.

At the point of an interior equilibrium both zero profit conditions (4.18) and (4.19) hold. They can be reexpressed as

$$q(\theta^*) = \frac{r\pi k_f}{(1 - \beta)(S_f + b_u - rE_u(\theta^*, \phi^*))},$$

and

$$q(\theta^*) = \frac{r\rho k_i}{(1 - \beta)(S_i + b_u - rE_u(\theta^*, \phi^*))}.$$

Then equating the two ratios on the right hand side, and solving for $E_u(\theta^*, \phi^*)$ we get

$$E_u(\theta^*, \phi^*) = \frac{\pi k_f (y - mF) - \rho k_i (y - \tau)}{r(\pi k_f - \rho k_i)}.$$

Substituting in turn the expression for $E_u(\theta^*, \phi^*)$ from (4.20) and solving for $\alpha(\theta^*)$ yields

$$\alpha(\theta^*) = \frac{\rho k_i S_f - \pi k_f S_i}{\beta(S_i - S_f)(k_f \phi^* + k_i(1 - \phi^*))} > 0.$$

By comparing $\alpha(\theta^*)$ and $\alpha(\bar{\theta})$, and recalling that $k_f > k_i$ and $\rho > \pi$ by assumption, we have $\alpha(\bar{\theta}) > \alpha(\theta^*)$, which implies $\bar{\theta} > \theta^*$ given the properties of $\alpha(\cdot)$.

Q.E.D.

Bifurcation point

The value $\bar{\theta}$ is in fact a bifurcation point, i.e. the value of market tightness that separates two regions with qualitatively different dynamics. In particular, as follows from Proposition 1, for any $\theta > \bar{\theta}$ derivative $\frac{\partial E_u(\theta, \phi)}{\partial \phi}$ is negative. Hence, in region 2, Fig.4.6, loci (4.18) and (4.19) are positively sloped in the (θ, ϕ) -plane. So, if they intersect, the locus of formal jobs, being steeper in the neighbourhood of the interior equilibrium point, must cross the locus of informal jobs from below. Then, in that particular case, the interior equilibrium is not stable. By contrast, if $\theta < \bar{\theta}$ derivative $\frac{\partial E_u(\theta, \phi)}{\partial \phi}$ is positive and the two loci are negatively sloped, the locus of formal jobs crosses the locus of informal jobs from above, so that the resulting interior equilibrium is stable.

Proposition 2 ascertains that given $k_f > k_i$ and $\rho > \pi$ the interior equilibrium in our model is stable when it is unique, and this situation is depicted in Fig.4.4. What if conditions $k_f > k_i$ and $\rho > \pi$ are violated?

If $k_f < k_i$ or $k_f = k_i$ then, given $\rho > \pi$, it is easy to show that the interior equilibrium in region 2, Fig.4.6, satisfies $\theta^* > \bar{\theta}$ or $\theta^* = \bar{\theta}$, respectively. In other words, in the former case it falls in the region making for a non-stable interior equilibrium (where both loci (4.18) and (4.19) are positively sloped - see above). In the latter case, it falls right in the border of regions with different phase dynamics, i.e. it coincides with the bifurcation point. In such a situation it is impossible to make an unambiguous judgement on what the stability properties of the interior equilibrium will be.

Loayza (1996) notes that along with low labour costs informal firms are also notable for facing high costs of capital. Thus, if one interprets start-up expenditure as costs of capital the case for $k_f < k_i$ can in principle be made. For such a modification of our model it would imply that no stable long-run equilibrium with both formal and informal jobs is possible: while in region 2, Fig.4.6, the equilibrium is not stable as explained above, region 4 ceases to exist altogether. Still, however, it is generally held that total costs of access to legality are higher than costs of entry in informality. Thus, we do believe that the assumption of $k_f > k_i$ is very sensible, so that a stable long-run mixed equilibrium is possible.

Finally, assuming $\rho < \pi$ is not realistic. It has been mentioned in the main text that evidence suggests a labour turnover being higher in the informal sector. This, thus, runs counter to the assumption of $\rho < \pi$.

4.A.3 Welfare and Policy Making

In this part of the appendix we derive the expression for the steady state private surplus, solve the dynamic optimisation problems, and show that in general the equilibria in our model are not efficient.

Derivation of steady state surplus

In derivation of the private welfare function we follow Hosios (1990) and Pissarides (2000). In the absence of capital costs the total private surplus in steady state can be given by the flow of aggregate utility equal to the sum of the steady state utilities as follows:

$$\Xi(\theta, \phi) = NrE_f + NrJ_f + IrE_i + IrJ_i + \theta U (\phi rV_f + (1 - \phi)rV_f) + UrE_u,$$

where $N = \phi \frac{\delta+m}{\delta+\phi m} (1 - U)$ and $I = (1 - \phi) \frac{\delta}{\delta+\phi m} (1 - U)$ are the *numbers* of formal and informal filled jobs, respectively, in steady state. θU gives the total number of vacancies in steady state, of which $\phi \theta U$ are formal vacancies and $(1 - \phi) \theta U$ are informal vacancies; $U = \frac{\delta(\delta+m)}{(\delta(\delta+m)+\alpha(\theta)(\delta+\phi m))}$ (see above).

As it is assumed that firms spend some resources, k_f or k_i , before opening a vacancy, the correct expression for the total private surplus in our model must include the upfront expenditures (i.e. must take into account the total size of the market):

$$\begin{aligned} \Xi(\theta, \phi) = & NrE_f + N(rJ_f - rk_f) + IrE_i + I(rJ_i - rk_i) \\ & + \theta U (\phi(rV_f - rk_f) + (1 - \phi)(rV_i - rk_i)) + UrE_u. \end{aligned}$$

One can simplify this expression by substituting for J_f , J_i , V_f , V_i , E_f , E_i , and E_u , from

(4.1)-(4.7), and by evaluating it in the steady state equilibrium where $V_f = k_f$ and $V_i = k_i$, as follows from two zero profit conditions (4.10) and (4.11). It is important, however, that we never make use of wage bargaining equations (4.8) and (4.9), thus, ignoring the specific way that wages share the surplus of a match. In the end simple algebra yields:

$$\Xi(\theta, \phi) = N(y - \tau - \pi k_f) + I(y - mF - \rho k_i) - \theta U(\phi r k_f + (1 - \phi) r k_i) + U b_u.$$

Thus, the steady state welfare is equal to the flow product minus flow costs of functioning, multiplied by the number of firms in each sector, as appropriate, less flow costs of maintaining vacancies in each sector, and plus the flow benefit of being unemployed.

Solutions to dynamic maximisation problems

In order to illustrate some underlying problems with the efficient allocation of resources in wage bargaining equilibrium, consider first the maximisation programme when the economy ends up in the corner equilibrium such as the one in Fig.4.2.

Corner equilibrium case The maximisation programme is

$$\max_{\theta} \int_0^{\infty} e^{-rt} \Xi(\theta, 1) dt = \max_{\theta} \int_0^{\infty} e^{-rt} (N(y - \tau - \pi k_f) - \theta U k_f r + U b_u) dt$$

subject to

$$\begin{aligned} \frac{dN}{dt} &= \alpha(\theta)U - \delta N, \\ \frac{dU}{dt} &= \delta N - \alpha(\theta)U, \end{aligned}$$

and $N + U = 1$.

Let μ be a co-state variable. Then the Hamiltonian is

$$H = e^{-rt} ((1 - U)(y - \tau - \pi k_f) - \theta U k_f r + U b_u) + \mu (\delta(1 - U) - \alpha(\theta)U),$$

and the current value Hamiltonian is

$$\mathcal{H} = ((1 - U)(y - \tau - \pi k_f) - \theta U k_f r + U b_u) + e^{rt} \mu (\delta (1 - U) - \alpha(\theta) U).$$

The optimal path of unemployment and market tightness satisfies the restrictions on derivatives $\frac{dN}{dt}$ and $\frac{dU}{dt}$, and the following Euler conditions: $\frac{\partial \mathcal{H}}{\partial \theta} = 0$, $\frac{\partial \mathcal{H}}{\partial \chi} = \frac{dU}{dt}$, and $\frac{\partial \mathcal{H}}{\partial U} = r\chi - \frac{d\chi}{dt}$, where $\chi = e^{rt} \mu$. Then in steady state

$$\begin{aligned} \frac{\partial \mathcal{H}}{\partial \theta} &= -U k_f r - \frac{\partial \alpha(\theta)}{\partial \theta} \chi U = 0, \\ \frac{\partial \mathcal{H}}{\partial \chi} &= \delta (1 - U) - \alpha(\theta) U = 0, \\ \frac{\partial \mathcal{H}}{\partial U} &= -(y - \tau - \pi k_f) - \theta k_f r + b_u - \chi (\delta + \alpha(\theta)) = r\chi - \frac{d\chi}{dt}. \end{aligned}$$

By solving this we get $\chi = \frac{-k_f r}{\frac{\partial \alpha(\theta)}{\partial \theta}}$, so that $\frac{d\chi}{dt} = 0$ as θ is constant in steady state, and

$$S_f = r k_f \left(\frac{\pi + \alpha(\theta) - \theta \frac{\partial \alpha(\theta)}{\partial \theta}}{\frac{\partial \alpha(\theta)}{\partial \theta}} \right).$$

The last expression is formula (4.29) given in the main text, which the efficient allocation must satisfy. In turn, it is easy to check that the zero profit condition (4.18) implies

$$S_f = r k_f \frac{(\pi + \alpha(\theta) \beta)}{(1 - \beta) q(\theta)}.$$

In other words, the wage bargaining equilibrium satisfying the expression above, also given by (4.30) in the main text, coincides with efficient allocation (4.29) only if $\frac{\partial \alpha(\theta)}{\partial \theta} = (1 - \beta) q(\theta)$, which is the standard result in the matching function literature (see Hosios, 1990). If this efficiency condition does not hold, the wage bargaining equilibrium is not efficient.

Interior equilibrium case In the case of the interior equilibrium there is even more scope for potential inefficiencies. Consider the maximisation programme

$$\max_{\theta, \phi} \int_0^{\infty} e^{-rt} \Xi(\theta, \phi) dt = \max_{\theta, \phi} \int_0^{\infty} \left\{ e^{-rt} (N(y - \tau - \pi k_f) + I(y - mF - \rho k_i)) \right. \\ \left. - e^{-rt} (\theta U(\phi k_f r + (1 - \phi) k_i r) - U b_u) \right\} dt$$

subject to

$$\begin{aligned} \frac{dN}{dt} &= \alpha(\theta) \phi U - \delta N, \\ \frac{dI}{dt} &= \alpha(\theta) (1 - \phi) U - (\delta + m) I, \\ \frac{dU}{dt} &= \delta N + (\delta + m) I - \alpha(\theta) U, \end{aligned}$$

and $N + I + U = 1$.

Let μ and γ be co-state variables, so that the Hamiltonian is

$$\begin{aligned} H &= e^{-rt} (N(y - \tau - \pi k_f) + (1 - N - U)(y - mF - \rho k_i) - \theta U(\phi k_f r + (1 - \phi) k_i r)) \\ &\quad + e^{-rt} U b_u + \mu (\delta N + (\delta + m)(1 - U - N) - \alpha(\theta) U) + \gamma (\alpha(\theta) \phi U - \delta N), \end{aligned}$$

Then the current value Hamiltonian is

$$\begin{aligned} \mathcal{H} &= (N(y - \tau - \pi k_f) + (1 - N - U)(y - mF - \rho k_i) - \theta U(\phi k_f r + (1 - \phi) k_i r) + U b_u) \\ &\quad + e^{rt} \mu (\delta N + (\delta + m)(1 - U - N) - \alpha(\theta) U) + e^{rt} \gamma (\alpha(\theta) \phi U - \delta N). \end{aligned}$$

The optimal path of formal employment, unemployment and market tightness satisfies the restrictions on $\frac{dN}{dt}$, $\frac{dU}{dt}$, and $\frac{dI}{dt}$, and the following Euler conditions (see Kamien and Schwartz, 2001, p.144): $\frac{\partial H}{\partial \theta} = 0$, $\frac{\partial H}{\partial \phi} = 0$, $\frac{\partial H}{\partial \mu} = \frac{dU}{dt}$, $\frac{\partial H}{\partial \gamma} = \frac{dN}{dt}$, $\frac{\partial H}{\partial U} = -\frac{d\mu}{dt}$, $\frac{\partial H}{\partial N} = -\frac{d\gamma}{dt}$. Or, alternatively, for the current value Hamiltonian we must have: $\frac{\partial \mathcal{H}}{\partial \theta} = 0$, $\frac{\partial \mathcal{H}}{\partial \phi} = 0$, $\frac{\partial \mathcal{H}}{\partial \chi} = \frac{dU}{dt}$, $\frac{\partial \mathcal{H}}{\partial \psi} = \frac{dN}{dt}$, $\frac{\partial \mathcal{H}}{\partial U} = r\chi - \frac{d\chi}{dt}$, and $\frac{\partial \mathcal{H}}{\partial N} = r\psi - \frac{d\psi}{dt}$, where $\chi = e^{rt}\mu$ and $\psi = e^{rt}\gamma$.

Thus

$$\begin{aligned}
\frac{\partial \mathcal{H}}{\partial \theta} &= -U(\phi k_f r + (1 - \phi) k_i r) - \frac{\partial \alpha(\theta)}{\partial \theta} \chi U + \frac{\partial \alpha(\theta)}{\partial \theta} \psi \phi U = 0, \\
\frac{\partial \mathcal{H}}{\partial \phi} &= -\theta U(k_f r - k_i r) + \psi \alpha(\theta) U = 0, \\
\frac{\partial \mathcal{H}}{\partial \chi} &= \frac{dU}{dt} = \delta N + (\delta + m)(1 - U - N) - \alpha(\theta) U = 0, \\
\frac{\partial \mathcal{H}}{\partial \psi} &= \frac{dN}{dt} = \alpha(\theta) \phi U - \delta N = 0, \\
\frac{\partial \mathcal{H}}{\partial U} &= -(y - mF - \rho k_i - b_u) - \theta(\phi k_f r + (1 - \phi) k_i r) - (\delta + m + \alpha(\theta)) \chi + \psi \alpha(\theta) \phi \\
&= r\chi - \frac{d\chi}{dt}, \\
\frac{\partial \mathcal{H}}{\partial N} &= (y - \tau - \pi k_f) - (y - mF - \rho k_i) - m\chi - \delta\psi = r\psi - \frac{d\psi}{dt}.
\end{aligned}$$

By solving the first two equations we get

$$\begin{aligned}
\psi &= r \frac{\theta(k_f - k_i)}{\alpha(\theta)} = r \frac{(k_f - k_i)}{q(\theta)}, \\
\chi &= \left(\phi r \frac{\theta(k_f - k_i)}{\alpha(\theta)} - r \frac{(\phi k_f + (1 - \phi) k_i)}{\frac{\partial \alpha(\theta)}{\partial \theta}} \right),
\end{aligned} \tag{4.36}$$

so that $\frac{d\psi}{dt} = 0$ and $\frac{d\chi}{dt} = 0$, as θ and ϕ are constant in steady state.

Thus, at the point of optimum we must have

$$\begin{aligned}
-S_i - \theta k_i r - \left(\phi r \frac{\theta(k_f - k_i)}{\alpha(\theta)} - r \frac{(\phi k_f + (1 - \phi) k_i)}{\frac{\partial \alpha(\theta)}{\partial \theta}} \right) (\rho + \alpha(\theta)) &= 0, \\
S_f - S_i - m \left(\phi r \frac{\theta(k_f - k_i)}{\alpha(\theta)} - r \frac{(\phi k_f + (1 - \phi) k_i)}{\frac{\partial \alpha(\theta)}{\partial \theta}} \right) - r \frac{\theta(k_f - k_i)}{\alpha(\theta)} \pi &= 0,
\end{aligned}$$

which are obtained from $\frac{\partial \mathcal{H}}{\partial U} = r\chi - \frac{d\chi}{dt}$ and $\frac{\partial \mathcal{H}}{\partial N} = r\psi - \frac{d\psi}{dt}$, respectively. The two equations above can be rearranged to be

$$\begin{aligned}
\left(\phi r \frac{\theta(k_f - k_i)}{\alpha(\theta)} - r \frac{(\phi k_f + (1 - \phi) k_i)}{\frac{\partial \alpha(\theta)}{\partial \theta}} \right) &= -\frac{(S_i + \theta k_i r)}{(\rho + \alpha(\theta))}, \\
(S_f - S_i) + m \frac{(S_i + \theta k_i r)}{(\rho + \alpha(\theta))} - r \frac{\theta(k_f - k_i) \pi}{\alpha(\theta)} &= 0.
\end{aligned} \tag{4.37}$$

Remark on the existence of solutions In general the existence of (interior) efficient solutions depends on further assumptions about the matching function, in particular the second derivatives of $\alpha(\theta)$ and $q(\theta)$. If interior solutions do not exist then only one of the two equations above can be satisfied at any moment in time. So that either the total number of jobs or the allocation across the sectors is not efficient. However, assuming that solutions to the maximisation problem do exist, we can explore if the equilibrium allocation under wage bargaining belongs to the set of such solutions. From Kamien and Schwartz (2001) we know that the Pontryagin maximum principle provides the necessary conditions for optimality. They become sufficient if, for example, the *maximised* Hamiltonian satisfies the condition of Arrow's generalised version of Mangasarian's theorem (*ibid*, pp.221-222). However, even if the first-order Euler conditions are not sufficient, they are still useful in recognising inefficient allocations (as in, e.g. Acemoglu and Shimer, 1999).

Implications for the wage bargaining equilibrium In order to evaluate efficiency properties of the wage bargaining equilibrium in region 2, Fig.4.6, one needs to work out the signs of two partial derivatives $\frac{\partial \mathcal{H}}{\partial \theta}$ and $\frac{\partial \mathcal{H}}{\partial \phi}$ in the interior equilibrium. In the case of optimal allocation the two derivatives must be equal to zero.

(a) Consider first derivative $\frac{\partial \mathcal{H}}{\partial \theta}$.

From (4.36) we know that $\frac{d\psi}{dt} = \frac{d\chi}{dt} = 0$. Then, $\frac{\partial \mathcal{H}}{\partial U} = r\chi$ and $\frac{\partial \mathcal{H}}{\partial N} = r\psi$. These two equations can be solved for χ and ψ , so that the efficient allocation must satisfy

$$\chi = -\frac{(S_i + \theta k_i r)}{(\rho + \alpha(\theta))},$$

$$\psi = \frac{(S_f - S_i)}{\pi} + m \frac{(S_i + \theta k_i r)}{\pi(\rho + \alpha(\theta))}.$$

Substituting the expressions for χ and ψ above into the expression for $\frac{\partial \mathcal{H}}{\partial \theta}$ yields

$$\frac{\partial \mathcal{H}}{\partial \theta} = U \left(-(\phi k_f r + (1 - \phi) k_i r) + \frac{\partial \alpha(\theta)}{\partial \theta} \frac{(S_i + \theta k_i r)}{(\rho + \alpha(\theta))} \frac{(\pi + \phi m)}{\pi} + \frac{\frac{\partial \alpha(\theta)}{\partial \theta} \phi (S_f - S_i)}{\pi} \right).$$

In the interior equilibrium (by recalling the definition of S_f and S_i , and substituting for them from (4.18)-(4.19) and using (4.20)) this expression becomes

$$\frac{\partial \mathcal{H}}{\partial \theta} |_{\text{equil.}} = U \frac{r \left(\left(\frac{\partial \alpha(\theta)}{\partial \theta} - (1 - \beta) q(\theta) \right) (\phi k_f + (1 - \phi) k_i) \pi (\rho + \alpha(\theta)) + \frac{\partial \alpha(\theta)}{\partial \theta} \beta \alpha(\theta) \phi (\pi + \phi m) (k_f - k_i) \right)}{\pi (\rho + \alpha(\theta)) (1 - \beta) q(\theta)}.$$

The sign of this derivative depends on the two relations between $\frac{\partial \alpha(\theta)}{\partial \theta}$ and $(1 - \beta) q(\theta)$, and k_f and k_i . As $k_f > k_i$, it is easy to see that $\frac{\partial \mathcal{H}}{\partial \theta}|_{equil.}$ is unambiguously positive if $\frac{\partial \alpha(\theta)}{\partial \theta} > (1 - \beta) q(\theta)$. In that situation the total amount of jobs created in the economy is inefficiently low in the interior wage bargaining equilibrium. If, however, $\frac{\partial \alpha(\theta)}{\partial \theta} < (1 - \beta) q(\theta)$ then the sign of derivative $\frac{\partial \mathcal{H}}{\partial \theta}|_{equil.}$ may be ambiguous.

Note, condition $\frac{\partial \alpha(\theta)}{\partial \theta} < (1 - \beta) q(\theta)$ is equivalent to $\beta < \frac{\alpha(\theta) - \frac{\partial \alpha(\theta)}{\partial \theta} \theta}{\alpha(\theta)} = -\frac{\partial q(\theta)}{\partial \theta} \frac{\theta}{q(\theta)} = \eta(\theta)$, where $\eta(\theta)$ is the elasticity of the matching function in the notation introduced in, e.g. Pissarides (2000).

(b) Consider now derivative $\frac{\partial \mathcal{H}}{\partial \phi}$.

Applying the same approach as to the analysis of $\frac{\partial \mathcal{H}}{\partial \theta}$ we obtain the following expression for derivative $\frac{\partial \mathcal{H}}{\partial \phi}$:

$$\frac{\partial \mathcal{H}}{\partial \phi} = U \left(-\theta (k_f r - k_i r) + \left(\frac{(S_f - S_i)}{(\pi + \phi m)} + \frac{m(\phi k_f r + (1 - \phi) k_i r)}{\frac{\partial \alpha(\theta)}{\partial \theta} (\pi + \phi m)} \right) \alpha(\theta) \right).$$

At the point of the interior equilibrium the derivative becomes

$$\frac{\partial \mathcal{H}}{\partial \phi}|_{equil.} = U \frac{\alpha(\theta) r \left(((1 - \beta) q(\theta) - \frac{\partial \alpha(\theta)}{\partial \theta}) m(\phi k_f + (1 - \phi) k_i) + \frac{\partial \alpha(\theta)}{\partial \theta} \beta (\pi + \phi m) (k_f - k_i) \right)}{\frac{\partial \alpha(\theta)}{\partial \theta} (\pi + \phi m) (1 - \beta) q(\theta)}.$$

As $k_f > k_i$ we have that derivative $\frac{\partial \mathcal{H}}{\partial \phi}|_{equil.}$ is always positive if $\frac{\partial \alpha(\theta)}{\partial \theta} < (1 - \beta) q(\theta)$ or $\beta < \eta(\theta)$, for that matter. In that case an increase in the proportion of formal vacancies among all vacancies will increase the value of the Hamiltonian and, hence, welfare. If, however, $\beta < \eta(\theta)$ the sign of the derivative is ambiguous.

(c) Finally, it can also be shown that regardless of the value of β the total amount of jobs created in the interior wage bargaining equilibrium is always inefficiently low.

Indeed, from the second equation in (4.37) it follows that

$$\frac{r}{q(\theta)} (k_f - k_i) \pi \rho = \alpha(\theta) (S_f - S_i) + (\rho S_f - \pi S_i) - \frac{r}{q(\theta)} \alpha(\theta) (k_f \pi - k_i \rho).$$

Let us denote by $\theta^{eff.}$ the efficient value of market tightness that satisfies this equation. Consider now two zero profit conditions (4.18) and (4.19). By using (4.20) and some simple algebra we arrive at

$$\frac{r}{q(\theta)} (k_f - k_i) \pi \rho = \frac{(1 - \beta) \pi \rho}{(\alpha(\theta) \beta ((1 - \phi) \pi + \phi \rho) + \pi \rho)} (\alpha(\theta) \beta (S_f - S_i) + (\rho S_f - \pi S_i)).$$

Let us denote by $\theta^{equil.}$ the level of market tightness that satisfies the formula above in the equilibrium with wage bargaining.

In region 2, Fig.4.6, i.e. where an interior equilibrium is possible, we have $\pi k_f < \rho k_i$. So, in that region for any θ and ϕ it holds that

$$\alpha(\theta)(S_f - S_i) + (\rho S_f - \pi S_i) - \frac{r}{q(\theta)}\alpha(\theta)(k_f\pi - k_i\rho) > \frac{(1-\beta)\pi\rho}{(\alpha(\theta)\beta((1-\phi)\pi + \phi\rho) + \pi\rho)}(\alpha(\theta)\beta(S_f - S_i) + (\rho S_f - \pi S_i)).$$

By making use of this result, as well as of the fact that

$$\frac{\partial}{\partial\theta}\left(\frac{r}{q(\theta)}(k_f - k_i)\pi\rho\right) = \left(-\frac{\partial q(\theta)}{\partial\theta}\frac{r}{q^2(\theta)}(k_f - k_i)\pi\rho\right) > 0,$$

it is straightforward to see that $\theta^{equil.} < \theta^{eff.}$. Thus, firms always underinvest in job creation in the interior wage bargaining equilibrium. In other words, such a conclusion implies that given $k_f > k_i$ derivative $\frac{\partial\mathcal{H}}{\partial\theta}|_{equil.}$ above must be positive regardless of the value of β . The implication for the sign of $\frac{\partial\mathcal{H}}{\partial\phi}|_{equil.}$, however, remains unclear.

Summary Four remarks below summarise the main implications that can be drawn from the analysis of the two derivatives, $\frac{\partial\mathcal{H}}{\partial\theta}|_{equil.}$ and $\frac{\partial\mathcal{H}}{\partial\phi}|_{equil.}$, of the Hamiltonian above.

First, two factors affect the efficiency of equilibrium in our model. On the one hand, it is the relation between the bargaining power of workers, β , and the elasticity of the matching function, $\eta(\theta)$. On the other hand, it is the relation between the values of upfront costs, k_f and k_i . Both relations affect both the total amount of job creation in the economy and the allocation of jobs between the two sectors.

Second, both $\frac{\partial\mathcal{H}}{\partial\theta}|_{equil.}$ and $\frac{\partial\mathcal{H}}{\partial\phi}|_{equil.}$ are equal to zero if and only if $\beta = \eta(\theta)$ and $k_f = k_i$. Hence, this is a necessary and sufficient condition for the efficiency of equilibrium.

Third, it is easy to see that condition $\beta < \eta(\theta)$ implies that firms tend to overinvest in creation of informal jobs and the economy as a whole. Indeed, assuming, for example, that $k_f = k_i$, inequality $\beta < \eta(\theta)$ implies that $\frac{\partial\mathcal{H}}{\partial\theta}|_{equil.} > 0$ and $\frac{\partial\mathcal{H}}{\partial\phi}|_{equil.} < 0$. By contrast, under condition $\beta > \eta(\theta)$ firms *ceteris paribus* tend to underinvest in creation of informal jobs and in the economy as a whole.

Finally, condition $k_f > k_i$ implies that firms tend to underinvest in creation of jobs both in the formal sector and the economy as a whole (it is easy to see by putting $\beta = \eta(\theta)$).

Moreover, whatever the relation between β and $\eta(\theta)$ firms always underinvest in job creation in our economy, as proved in subsection (c) above.

Changes in parameters: the impact on job allocation

This section of the appendix provides an insight into how changes in potential policy instruments (such as tax rate τ , fine rate F , efficiency of monitoring m , and unemployment compensation b_u) affect the allocation of jobs, the level of employment and unemployment in the interior equilibrium (Fig.4.4). We consider effects of such changes for each policy parameter in turn, while holding other parameters of the model constant.

Monitoring and fines for engagement in informal activity Monitoring of firms and fines for engagement in the informal business are the costs of operating in the informal sector. Suppose, first, the government decides to subsidise monitoring authorities (e.g. the tax police) more generously, which results in an increase in m - the variable describing the probability that an employer gets caught working in underground business. Such a rise in m leads, firstly, to a decrease in the informal surplus relative to the formal surplus, and, secondly, to an increase in the rate of death of informal matches. The former effect has a negative impact on the profitability of underground jobs, whereas the latter has a positive externality effect: since underground firms die faster, it becomes easier to fill in the vacancies both for remaining formal and informal businesses.

It is easy to verify that an increase in m turns the locus of formal jobs (4.18) clockwise around the point $(\theta'_f, 1)$, while the locus of informal jobs (4.19) moves downwards. This brings about an increase in the share of formal vacancies, ϕ , as well as reduces the equilibrium value of the market tightness, θ . Apart from that, it can also be shown that not only the proportion of formal vacancies increases, but the total number of formal jobs also rises, while the total number of informal jobs decreases. Moreover, the latter effect is outweighed by the former so that unemployment decreases.

Instead of investing more in monitoring authorities in an attempt to shackle the shadow sector the government may simply increase punishment for involvement in underground business. In our model this would imply a rise in F . The effect of higher F is somewhat similar to that of an increase in m , with the exception that the former does not affect

the effective discount rate in the informal sector, ρ , and, thus, does not create a positive externality on both formal and informal firms competing for workers. Geometrically, an increase in F has absolutely the same effect on both loci (4.18) and (4.19) as a rise in m , so that the upshot of this policy is identical to that in the previous paragraph: θ unambiguously decreases, while ϕ rises (Fig.4.8). The number of formal jobs increases, while the number of informal jobs drops. Unemployment is also reduced.

Taxes Empirical literature has suggested that taxes levied in the formal sector are one of the main possible reasons that drives potential employers underground as they represent one of the costs of legal production (see, e.g. Johnson *et al.*, 2000a). In our model an increase in the value of lump sum taxes on formal employers, τ , reduces the formal sector surplus and results in a downward shift of the locus of formal jobs (4.18) in the (θ, ϕ) -plane, while the locus of informal jobs (4.19) turns anti-clockwise around the point $(\theta_i, 0)$. This implies that the equilibrium value of θ rises, whereas the equilibrium value of ϕ decreases (Fig.4.7). The number of formal jobs decreases, and the number of informal jobs rises together with the number of unemployed. Thus, the effect of a rise in taxes is just opposite to the effect of an increase in punishment rate F or monitoring intensity m .

Unemployment benefits Contrary to taxes, fines and monitoring, the effect of unemployment benefits on both formal surplus S_f and informal surplus S_i is symmetric: both are reduced to the same degree if unemployment compensation is made more generous. An increase in b_u makes both loci (4.18) and (4.19) shift down. In the case of an interior equilibrium it is straightforward to verify that $\frac{d\phi}{db_u} < 0$, while $\frac{d\theta}{db_u} = 0$ by totally differentiating (4.18) and (4.19) with respect to θ , ϕ and b_u , and then solving the system of resulting equations in a stable equilibrium for $\frac{d\phi}{db_u}$ and $\frac{d\theta}{db_u}$. This result implies that in effect for a given θ both loci shift by the same distance, i.e. the equilibrium proportion of formal jobs, ϕ , decreases, whereas the equilibrium value of market tightness, θ , does not change. Thus, in the mixed case the introduction of more generous unemployment benefits leads to crowding out of formal vacancies by informal ones. It will also decrease the number of formal jobs, whereas both the number of informal jobs and unemployment will rise.

Other institutional changes Apart from parameters representing primary policy instruments the allocation of jobs across the sectors in the model depends on the value of entry costs, k_f and k_i . These parameters reflect an institutional climate or environment in the economy: spreading corruption and bribery, for instance, increases k_f relative to k_i .

Consider, for example, a decrease in k_f , which corresponds to less bribing and extortion in the formal sector. This can possibly be achieved by restructuring bureaucracy and by better enforcement of laws cracking down on corruption. With such a decrease in k_f the locus of formal jobs moves up, while the locus of informal jobs turns clockwise around the point $(\theta_i, 0)$. The outcome is a drop in the equilibrium value of θ , and an increase in the equilibrium value of ϕ . The number of formal jobs also rises, whereas the number of informal jobs decreases along with the level of unemployment. By contrast, an upsurge in corruption and rent-seeking in the economy would result in an increase in k_f , scare firms away from the formal sector, and increase unemployment. This would also support informal production and shift the allocation of vacancies and jobs towards informality (Fig.4.7).

Table 4.1 in the main text summarises the effects of all changes considered here.

Some further notes on policies In order not to move the focus of discussion away from the welfare effect of various policies, summarised in table 4.1 in the main text, we use this section of the appendix to look more closely at two particular implications that could be drawn from the analysis of the table. The first is concerned with the effect of various policies on the size of the informal sector. The second is related to the impact on unemployment. Let us consider them in turn.

The table suggests that a decrease in m , F , and/or a increase in τ , k_f , and b_u raise the share of the informal sector in the economy measured both by the proportion of informal vacancies in the total number of vacancies, $(1 - \phi)$, and by the ratio of filled informal jobs to filled formal jobs, i.e. I/N . The effects of a change in the intensity of monitoring, the severity of punishment for concealing business underground, as well as the effect of higher taxes and the impact of corruption in general concur well with results obtained in many other empirical and theoretical studies of informal economies which, however, often lacked a proper account of the labour market (for just a few examples the reader can be referred

to the studies by Murphy *et al.*, 1993; Shleifer and Vishny, 1998; Friedman *et al.*, 2000; Johnson *et al.*, 2000a; Sarte, 2000). It is interesting, however, that our model suggests that an *increase* in the size of the informal sector may be a result of policies aimed at making unemployment compensation more generous. Such an outcome can be contrasted to two recent theoretical results by Boeri (1999, 2000b) and Fugazza and Jacques (2004) as regards the effect of unemployment benefits on the size of the informal economy.

Boeri (1999, 2000b) has observed that overly generous non-employment benefits at the outset of transition from planned to market economy in some countries of Eastern Europe favoured a labour drive to non-employment, which in his model effectively implies employment in the informal sector. He explained that by the effect that open-ended unemployment benefits have on the decision of the non-employed to actively search for jobs in the formal part of the economy. Higher benefits imply that more non-employed workers would prefer to receive the compensation and on top of that enjoy readily available work in the subsistence sector. In other words, higher benefits weaken worker incentives to search for jobs in the formal sector.

A qualitatively different result has been obtained by Fugazza and Jacques (2004) in the dual labour market model where workers have to direct their search towards formal or informal vacancies. In their work higher unemployment benefits raise the worker's value of being attached to the regular sector, so that more individuals are likely to search for a job in the formal economy. The contrast with the Boeri's conclusion arises because of the assumption that unemployed workers searching for jobs in the irregular sector do not receive unemployment compensation.

The crucial difference between our work and the two studies just mentioned is that there a change in the value of benefits does not affect the decision of employers to take off in the informal economy. In our model an increase in that value affects the relative size of surpluses in each sector and strengthens the bargaining position of job-seekers vis-à-vis employers. In the stable interior equilibrium the surplus of a match is smaller in the formal sector, so the relative losses that firms have to bear during wage bargains in the wake of the increase are higher in the formal part of the economy. Hence, this urges more firms into the shadow sector.

The second point that can be seen from table 4.1 is that policies leading to an increase

in informal employment also cause an increase in unemployment.

The existence of links between unemployment and underground employment is well-known. In many developing and transitional countries informal activities have been argued to be of a subsistence nature (Gërxhani, 2004). They often involve the individuals who are affected most severely by changes in the economy and who cannot find a decent option in the formal sector. Not surprisingly, the unemployed, as a most vulnerable cohort in the labour market, are, thus, the primary candidates for supplying labour in the informal sector. Theoretically it has already been suggested (Boeri and Garibaldi, 2001) that policies aimed at reducing unemployment pay off by scaling down the shadow sector. At the same time, a conclusion has been made (*ibid*) that attempts to reduce, in the first place, shadow employment will result in higher open unemployment. Our model highlights that it is not always necessarily true.

Indeed, if one takes into account that labour turnover is higher in the shadow sector (as in this essay is reflected in a shorter duration of an informal match) then crowding out of formal vacancies by informal ones inevitably leads to higher steady state unemployment. If, however, the share of formal jobs in the economy increases, their on average longer duration makes for a decrease in open unemployment. This is a general intuition for the results presented in table 4.1. By contrast, the argument of Boeri and Garibaldi (2001) is related to a specific view of the informal sector. In particular, in their economy all jobs are necessarily formal when they are created. Informal jobs are seen as a means not to sack a worker when some shock hits a formal production unit: under certain conditions firms may find it profitable to convert formal, highly productive jobs into informal, low-productive ones, and maintain them thereafter instead of closing down their business completely. Meanwhile, workers employed in informal jobs have to look for better paid formal jobs and compete in their search with the unemployed. In that case cracking down on the shadow sector leads to an increase in open unemployment. At the same time, policies reducing unemployment imply that it becomes easier for informally employed workers to find a formal job, as competition from the unemployed, also looking for formal jobs, subsides. As a result, the duration of shadow jobs drops and the size of the shadow sector is reduced.

In the absence of on-the-job search in our framework we draw attention specifically

to the moment of entry into the economy, and suggest that it is important to create conditions attracting more firms into the formal sector, which will pay off by decreasing steady state unemployment.

Level curves for surpluses and state revenue

In the main text we consider the level curves of private surplus (4.26) and government revenue (4.35) in the (θ, ϕ) -plane. Here we highlight the factors that their slopes depend upon.

Private surplus The private surplus is given by (4.26). Its partial derivative with respect to θ is given by

$$\frac{\partial \Xi(\theta, \phi)}{\partial \theta} = \frac{\frac{\partial \alpha(\theta)}{\partial \theta} \delta(\delta+m)}{(\delta(\delta+m)+\alpha(\theta)(\delta+\phi m))^2} (\phi(\delta+m)S_f + (1-\phi)\delta S_i) - \frac{\delta(\delta+m)(\phi k_f r + (1-\phi)k_i r)}{\delta(\delta+m)+\alpha(\theta)(\delta+\phi m)} \left(\frac{\delta(\delta+m)+(\delta+\phi m)(\alpha(\theta)-\theta \frac{\partial \alpha(\theta)}{\partial \theta})}{(\delta(\delta+m)+\alpha(\theta)(\delta+\phi m))} \right).$$

In the interior equilibrium it is equal to

$$\begin{aligned} \frac{\partial \Xi(\theta, \phi)}{\partial \theta} \Big|_{\text{equil.}} = & \frac{r\delta(\delta+m)(\phi k_f + (1-\phi)k_i) \left(\frac{\partial \alpha(\theta)}{\partial \theta} (\delta(\delta+m)+(\delta+\phi m)\alpha(\theta)\beta) - (1-\beta)q(\theta)(\delta(\delta+m)+(\delta+\phi m)(\alpha(\theta)-\theta \frac{\partial \alpha(\theta)}{\partial \theta})) \right)}{(\delta(\delta+m)+\alpha(\theta)(\delta+\phi m))^2 (1-\beta)q(\theta)} \\ & + \frac{r\delta(\delta+m) \frac{\partial \alpha(\theta)}{\partial \theta} r((\delta+m)\phi k_f + \delta(1-\phi)k_i)}{(\delta(\delta+m)+\alpha(\theta)(\delta+\phi m))^2 (1-\beta)q(\theta)}. \end{aligned}$$

This derivative is unambiguously positive if $\frac{\partial \alpha(\theta)}{\partial \theta} \geq (1-\beta)q(\theta)$ or, for that matter, $\beta \geq \eta(\theta)$.

The derivative of the private surplus with respect to ϕ is given by

$$\frac{\partial \Xi(\theta, \phi)}{\partial \phi} = \frac{\alpha(\theta)(\delta+m)\delta((\delta+m+\alpha(\theta))S_f - (\delta+\alpha(\theta))S_i)}{(\delta(\delta+m)+\alpha(\theta)(\delta+\phi m))^2} + \theta \frac{\alpha(\theta)m\delta(\delta+m)(\phi k_f r + (1-\phi)k_i r)}{(\delta(\delta+m)+\alpha(\theta)(\delta+\phi m))^2} - \theta \frac{\delta(\delta+m)r(k_f - k_i)}{\delta(\delta+m)+\alpha(\theta)(\delta+\phi m)}.$$

In the interior equilibrium it is always positive and equal to

$$\frac{\partial \Xi(\theta, \phi)}{\partial \phi} \Big|_{\text{equil.}} = \theta \frac{r(\delta+m)\delta(mrk_i + (k_f - k_i)(\alpha(\theta)\beta\phi m + (\delta+m+\alpha(\theta))(r+\beta\delta)))}{(\delta(\delta+m)+\alpha(\theta)(\delta+\phi m))^2 (1-\beta)} > 0.$$

Thus, by using the implicit function theorem, the slope of the level curve for the private surplus in the (θ, ϕ) -plane is

$$\frac{\partial \theta}{\partial \phi} \Big|_{PS; \text{equil.}} = - \frac{\alpha(\theta)(mrk_i + (k_f - k_i)(\alpha(\theta)\beta\phi m + (\delta+m+\alpha(\theta))(r+\beta\delta)))}{A + B + C}, \quad (4.38)$$

where

$$\begin{aligned}
A &= \frac{\partial \alpha(\theta)}{\partial \theta} r((m + \delta) \phi k_f + \delta(1 - \phi) k_i), \\
B &= (\phi k_f + (1 - \phi) k_i) \frac{\partial \alpha(\theta)}{\partial \theta} (\delta(\delta + m) + (\delta + \phi m) \alpha(\theta) \beta), \\
C &= -(\phi k_f + (1 - \phi) k_i) (1 - \beta) q(\theta) \left(\delta(\delta + m) + (\delta + \phi m) \left(\alpha(\theta) - \theta \frac{\partial \alpha(\theta)}{\partial \theta} \right) \right).
\end{aligned}$$

From (4.38) it can be seen that the higher is k_f as compared to k_i , and the lower is β as compared to the elasticity of the matching function, $\eta(\theta)$, the steeper is the level curve in the (θ, ϕ) -plane at the point of the interior equilibrium (see Fig.4.9).

State revenue The government revenue is given in (4.35). Its partial derivative with respect to θ is always positive and equal to

$$\frac{\partial R(\theta, \phi)}{\partial \theta} = \frac{\frac{\partial \alpha(\theta)}{\partial \theta} \delta(\delta + m)}{(\delta(\delta + m) + \alpha(\theta)(\delta + \phi m))^2} (\phi(\delta + m)\tau + (1 - \phi)\delta mF + (\delta + \phi m)b_u) > 0.$$

The partial derivative with respect to ϕ is

$$\frac{\partial R(\theta, \phi)}{\partial \phi} = \frac{\alpha(\theta)(\delta + m)\delta}{(\delta(\delta + m) + \alpha(\theta)(\delta + \phi m))^2} ((\delta + m + \alpha(\theta))\tau - (\delta + \alpha(\theta))mF + mb_u) > 0.$$

It is unambiguously positive in the interior equilibrium as $\tau > mF$. Then the slope of the government revenue level curve in the (θ, ϕ) -plane is

$$\left. \frac{\partial \theta}{\partial \phi} \right|_{R, \text{ equil.}} = -\frac{\alpha(\theta)((\delta + m + \alpha(\theta))\tau - (\delta + \alpha(\theta))mF + mb_u)}{\frac{\partial \alpha(\theta)}{\partial \theta} (\phi(\delta + m)\tau + (1 - \phi)\delta mF + (\delta + \phi m)b_u)} < 0.$$

Thus, the bigger is τ as compared to mF the steeper is the level curve.

Chapter 5

Conclusions

This thesis was concerned with problems of labour reallocation common in many countries of economic transition in Eastern Europe and the former Soviet Union, particularly in Russia. The substantive contents of the work were arranged in three essays. Each piece of writing tackled a separate topic, focusing on a specific group of questions raised in the introduction. However, the essays built on each other and drew conclusions that help create a solid perspective on transitional labour reallocation. Here we summarise and extend those remarks.

In the first essay we employed econometric analysis in order to look into the flows of workers across employment, unemployment and inactivity in Russia, assess their relative importance and understand their determinants. The essay suggested that the Russian transition, similar in main characteristics to the East European experience, had still been notable for several peculiar features mainly related to wage adjustment effects on labour mobility, worker shifts between unemployment and the out-of-the-labour-force state, and the role of irregular markets. In general, the picture of transitional labour movement

painted in the study is different from the predictions and implications of the stylised theory of the reallocation by Aghion and Blanchard (1994) as outlined in the introduction to the thesis. Findings of the essay pointed out the directions in which the benchmark theory should be improved. The second and the third essays drew on these suggestions and put forward models of labour reallocation in transition that include some elements overlooked in the previous theoretical research.

In the second essay we developed an extended version of the Aghion and Blanchard (1994) seminal framework by adding to it the informal sector. While analysing the dynamics predicted by the model we argued that differences in East European and Russian transitional experience could be explained by the different roles that the informal economy had played in labour reallocation in the two regions. In particular, the more noticeable presence of the informal sector in Russia has probably slowed down the transformation process therein by creating additional difficulties for the new private sector to develop successfully. At the same time it is not the presence of the informal economy *per se* that is important, but rather is a precondition for its development, generated by the policies that governments adopt towards new private businesses. We argued that this side to transition and its effect on job creation should have received more close attention in the previous theoretical work on labour redistribution.

Finally, the third essay looked particularly into shifts of workers and entrepreneurs between formal and informal economies. While in the second essay the transition could be portrayed as successful only if in the end new formal firms crowded out the informal sector, in the new model of labour market search and wage bargaining we showed how government regulations can make for long-term equilibria where the informal sector still

exists along with the formal one. As rent sharing had widely become a common feature of interaction of firms and workers in transitional labour markets, we highlighted its importance in the informalisation process. Special attention was given to implications of the model for policy making. In particular, it was shown that inefficiencies in the labour market can make policies aimed at achieving a reduction in size of the shadow sector have a negative effect on economic welfare.

In our view, this research has generally made a contribution in the following areas.

Firstly, it has identified elements that play a substantial role in the course of transitional labour reallocation. It has also pointed to the gaps in the benchmark theory where those elements should be built in. In particular, in the Russian example the first essay emphasises both non-uniformity of the impact of the transition process, especially across different cohorts of individuals in the labour force, and the likely role of labour supply in the irregular sector. We are not the first to draw attention to heterogeneity of the labour force and labour supply, however the previous observations were based mainly on evidence from Eastern Europe, but not Russia (e.g. Boeri, 2000b). While the possibility of employment in the irregular sector was the central topic of the two theoretical essays, we did not attempt to incorporate a heterogeneous population. On the one hand, it was done in order not to overcomplicate the stylised models, where we believed the main point could be made in a simpler set-up (for example, Chapter 3 explains a drop in the participation rate widely observed in CEE and Russia as having been an inevitable feature of transition associated with a growth of transitional unemployment and the informal sector - Boeri, 2000b, draws somewhat similar conclusions in a rather cumbersome model). On the other hand, in the most recent research Commander *et al.* (2004) make an effort to

build an OST-type model with low- and high-skilled individuals to explain persistence of unemployment across transitional countries.

Other important elements that we have highlighted in this thesis include wage-setting mechanisms (in particular, we noted significance of wage arrears in Russia and their link with irregular labour supply and worker movement out of the labour force) and formation of reservation wages of the non-employed. Although we do not incorporate the arrears in either of the theoretical essays (for a somewhat related attempt to model a situation with non-payment of wages in the old sector the reader can be referred to Bouev, 2001), it was shown that wage-setting methods and conditions play a key role in the overall success of transitional redistribution of resources (as captured by wage mark-ups in Chapter 3). They also act as an important channel through which selection into the formal and the informal sectors is made (cf wage bargaining in Chapter 4). As regards the reservation wage of the jobless, both the second and the third essays clearly show that it does depend on the wage in the informal economy in addition to the level of unemployment benefits. As the size of the informal economy changes in the course of transition - so does the reservation wage of the non-employed. This issue has hardly ever been studied either from theoretical or empirical perspectives. An interested reader can find some initial insights in Matveenko *et al.* (1998) who in the model of the Russian labour market show that the wage in one sector adjusted for the time actually worked by employees is equal to the wage in another sector. In other words, workers in the first sector could perceive the wage in the second sector as a fair or reservation wage they deserve to receive and adjust their supply of working hours in the first sector accordingly. Obviously, the dynamics of the reservation wage should have important implications for development of the new

private sector in transition, and, hopefully, more attention will be given to it in future research.

A second contribution of this work is that in the theoretical set-up we have provided a convincing insight into the divergence of Russian and East European economies in the first several years of transition. As it is stressed in the second essay in this volume, creation of favourable conditions to the development of the new formal sector as opposed to the informal sector should be a primary objective of economic reforms. It is widely held, however, that the Russian business environment was or even still is less friendly, generally speaking, than its Western counterpart to emergence of new (formal) enterprises. We have shown that this has probably led to a considerable shift of the Russian private initiative into the underground sector. On the one hand, such a shift is undoubtedly related to the level of corruption and other barriers to successful development. On the other hand, other obstacles may have played a role, for example, labour market rigidities. Widely available empirical evidence suggests that after 1998 the transition trajectories of the two regions have probably pursued two similar rather than different paths. We cannot provide a non-speculative rationale for such a turnaround, apart from suggesting, intuitively, that the recent seeming convergence (or rather the absence of divergence) may well be attributed to changes in the Russian economy provoked by the ruble devaluation crisis in August, 1998, and rising oil prices. That facilitated the development of some export-oriented industries and put the end to the demonetisation of the economy. Also, that change is very likely to be linked with different policy making conducted by the new Russian government since Boris Yeltsin's resignation on the last day of 1999. However, that switch in the evolution of the Russian transition clearly calls for more research to

gain a deeper understanding of the issue.

Related to the questions raised in the previous paragraph are the policy implications derived from the two theoretical chapters of this thesis - a third important point to discuss. The general recommendation, as suggested above, is the creation of favourable conditions to new formal firms. But what does it mean more specifically? Both theoretical essays primarily view that first, and traditionally, as the choice of a proper tax regime, and, second, as the elimination of imperfections in the labour market.

In regard to the taxes, it is found that the higher they are the more likely, other things being equal, for firms to move into the shadow sector - a hardly surprising conclusion. At the same time, it has been emphasised in the text that the effect of variation in the tax rate should be assessed in comparison with the effects of other factors on profitability of the informal sector. Such factors include fines for concealment of activities underground and costs borne by informal entrepreneurs due to monitoring of tax evasion by the authorities. From the third essay it also follows that sunk start-up costs associated with initial investments in capital and so forth define the range of tax values that implies eventual convergence of the economy to equilibria with or without the shadow sector. From this perspective it is not surprising that some studies have found that higher taxes are not necessarily associated with a higher share of the informal sector (Friedman *et al.*, 2000). While the pure effect of taxes is that of raising incentives of moving in shadow, the total effect of increasing the burden of taxation will depend on the burden of other factors that must be taken into account. This sends an important message to researchers making empirical attempts at defining relationship between taxes and the size of the informal sector.

As far as labour market imperfections are concerned, the thesis clearly draws attention to the topic that has hardly been touched upon in the theoretical literature on transition. What is underlined here is that when markets are segmented their relative development hinges on the relative significance of those imperfections. In the second essay it is seen through introducing wage mark-ups in the new and the informal sectors. In the third essay it is conveyed through hold-up opportunities that bear upon investments and, hence, job creation in the two parts of the economy. If the government can reduce such wage mark-ups in the formal as opposed to the informal sector then it will increase chances that the economy converges to equilibria without informality. The third essay shows that in order to do that governments can and need to affect the reservation values of firms in their wage negotiations with workers. On the one hand, these values are most likely drawn down in the formal sector if bureaucracy wrings bribes from entrepreneurs. Thus, it suggests another argument in favour of fighting corruption. On the other hand, as rent sharing and informal contracting become increasingly widespread in transition economies, a necessity for a legal reform of labour market relations is obvious, at least in Russia. It is likely that many of the imperfections and preconditions for hold-ups in the formal sector can be eliminated through a proper contracting (Malcomson, 1997, can be referred to for a discussion of relevant issues), however, this area of research is also underdeveloped as far as transition economies are concerned.

The fourth noticeable finding of this work is that the presence of the informal economy affects the timing of transformation. In this respect, in order to speed up the transition a quick reduction in the size of the shadow sector is crucial, as the second essay suggests. At the same time, the third essay floats an idea that scaling down the informality is not

necessarily welfare improving. Does it mean that there is a clash of arguments here? The answer is no. A quick reduction in the size of the underground sector is advisable if all other economic conditions make for convergence to the equilibrium without informality, i.e. the one where the new and effective formal sector crowns the transformation process. Meanwhile, conclusions as regards the impact on economic welfare were drawn for the economy which long-term destination is a steady state with both the formal and the informal sectors. Thus, it is clearly two mutually exclusive situations that are considered here. The general suggestion is still creation of conditions favourable to the development of the new formal sector, curbing corruption, elimination or mitigation of labour market inefficiencies. That should both lead to an improvement in efficiency properties of the long-run equilibrium and speed up the convergence process.

A final point that we would like to raise is concerned with transitional unemployment. The empirical essay has found that even in the highly dynamic Russian labour market unemployment has become a more stagnant state. Is it an inevitable feature of transition? In a recent study Commander *et al.* (2004) suggest that persistence of transitional unemployment is likely to be linked with differences in the probability of finding a job across various groups of workers. As transition progresses unemployment becomes more concentrated among those whose probabilities are low, for example among low-skilled workers. This is well in line with our empirical findings from the Russian labour market. However, another simpler explanation is possible too. The numerical simulations of the stylised model put forward in the second essay suggest that transitional unemployment will not dwindle away to nothing, as it does in the benchmark theory of labour reallocation by Aghion and Blanchard (1994), so long as the new sector (and the informal one for that

matter) lays off workers. Convergence to a steady state implies that unemployment is bound to stabilise at a non-zero level - a natural rate. This rate will be higher, the heavier, among other things, is the burden of various regulations on new firms. Thus, in the absence of various shocks that could derail the steady convergence to a post-transitional equilibrium, the farther is the economy from the advent of reforms the more persistent unemployment should be and the longer should be the average duration of unemployment. It would be extremely interesting to see more results on this topic, though.

All in all, there is an obvious need for more research to be done on labour reallocation. Despite being a popular topic, it still contains many puzzles. Among other things this thesis has tried to shed some light on possible avenues that investigators should go and explore in order to gain a deeper understanding of the subject. We have outlined suggestions for further work as the time and limits of the DPhil are just not enough to pursue them within this project. It is hoped that those suggestions will not take forever to be implemented.

Bibliography

- [1] Acemoglu D. (1995) Reward Structures and the Allocation of Talent. *European Economic Review*. Vol. 39, pp. 17-33.
- [2] Acemoglu D. (1996) A Microfoundation for Social Increasing Returns in Human Capital Accumulation. *The Quarterly Journal of Economics*. Vol. 111, No. 3, August, pp. 779-804.
- [3] Acemoglu D. (1999) Changes in Unemployment and Wage Inequality: An Alternative Theory and Some Evidence. *The American Economic Review*. Vol. 89, No. 5, December, pp. 1259-1278.
- [4] Acemoglu D. (2001) Good Jobs versus Bad Jobs. *Journal of Labor Economics*. Vol. 19, No. 1, pp. 1-21.
- [5] Acemoglu D., Shimer R. (1999) Holdups and Efficiency with Search Frictions. *International Economic Review*, Vol. 40, No. 4, November.
- [6] Acemoglu D., Verdier T. (1998) Property Rights, Corruption and the Allocation of Talent: A General Equilibrium Approach, *Economic Journal*, Vol. 108(450), pp. 1381-1403.
- [7] Addison J.T., Portugal P. (2003) Unemployment Duration. Competing and Defective Risks. *The Journal of Human Resources*, Vol. 38, No. 1, pp. 156-191.
- [8] Aghion P., Blanchard O. (1994) On the Speed of Transition in Central Europe. In: S.Fischer and J.Rotemberg, eds. *NBER Macroeconomics Annual*, Vol. 9, pp. 283-320. Cambridge, MA: MIT Press.
- [9] Agénor P.-R., Aizenman J. (1999) Macroeconomic Adjustment with Segmented Labor Markets. *Journal of Development Economics*. Vol. 58, pp. 277-296.
- [10] Allingham M.G., Sandmo A (1972) Income Tax Evasion: A Theoretical Analysis. *Journal of Public Economics*, Vol. 1, No. 3-4, pp. 323-338.
- [11] Andreoni J., Erard B., Feinstein J. (1998) Tax Compliance. *Journal of Economic Literature*, Vol. 36, No. 2, June, pp. 818-860.
- [12] Andrienko Y., Guriev S. (2003) Determinants of Interregional Mobility in Russia: Evidence from Panel Data. William Davidson Working Paper No. 551, February. William Davidson Institute, University of Michigan Business School, Ann-Arbor, USA.

- [13] Atkeson A., Kehoe P.J.(1996) Social Insurance and Transition. *International Economic Review*, Vol. 37, No. 2, pp. 377-401.
- [14] Basu S., Estrin S., Svejnar J. (2000) Employment and Wages in Enterprises Under Communism and In Transition: Evidence from Central Europe and Russia. William Davidson Working Paper No. 440, June. William Davidson Institute, University of Michigan Business School, Ann-Arbor, USA.
- [15] Basu S., Estrin S., Svejnar J. (2004) Wage Determination Under Communism and In Transition: Evidence from Central Europe. William Davidson Working Paper No. 717, August. William Davidson Institute, University of Michigan Business School, Ann-Arbor, USA.
- [16] Becker G. (1965) A Theory of the Allocation of Time. *Economic Journal*, Vol. 75(299), September, pp. 493-517.
- [17] Becker G. (1968) Crime and Punishment: An Economic Approach. *Journal of Political Economy*, Vol. 76, No. 2, March - April, pp. 169-217.
- [18] Berkowitz D., Cooper D.J. (1997) Start-Ups and Transition. William Davidson Working Paper No. 84, September. William Davidson Institute, University of Michigan Business School, Ann-Arbor, USA.
- [19] Berkowitz D., DeJong D.N. (1997) Observations on the Speed of Transition in Russia: Prices and Entry. In: J.M.Nelson, Ch.Tilley and L.Walker, eds. *Transforming Post-Communist Political Economies*. Washington D.C.: National Academy Press, National Research Council, US.
- [20] Blanchard O. (1997) The Economics of Post-Communist Transition. Clarendon Press. Oxford. 149p.
- [21] Blanchard O. (1998) Flows, Bargaining and Unemployment. February. MIT Lectures Econ2410d. Mimeo. Available at http://econ-www.mit.edu/faculty/index.htm?prof_id=blanchar&type=books
- [22] Blanchard O., Diamond P. (1992) The Flow Approach to Labor Markets. *American Economic Review*, Vol. 82, No. 2, Papers and Proceedings, May, pp. 354-359.
- [23] Blanchard O., Diamond P. (1994) Ranking, Unemployment Duration, and Wages. *The Review of Economic Studies*, Vol. 61, No. 3, July, pp. 417-434.
- [24] Boeri T. (1998) Labor Market Flows in the Midst of Structural Change. In: S.Commander, Ed. *Enterprise Restructuring and Unemployment in Models of Transition*. EDI Development Studies, Washington DC, the World Bank.
- [25] Boeri T. (1999) Transition with Labour Supply. William Davidson Working Paper No. 274, December. William Davidson Institute, University of Michigan Business School, Ann-Arbor, USA.
- [26] Boeri T. (2000a) Optimal Speed of Transition 10 Years After. CEPR Discussion Paper No. 2384, February. Centre for Economic Policy Research, London, UK.

- [27] Boeri T. (2000b) Structural Change, Welfare Systems, and Labour Reallocation. Lessons from the Transition of Formerly Planned Economies. Oxford: Oxford University Press. 240p.
- [28] Boeri T., Bruno R. (1997) A Short Note on the Characteristics of Labour Turnover in Central and Eastern Europe. In: *Labour Market Dynamics in the Russian Federation*. OECD Proceedings, the Centre for Co-operation with the Economies in Transition, OECD, Paris.
- [29] Boeri T., Burda M., Kollo J. (1998) Mediating the Transition: Labour Markets in Central and Eastern Europe. In: L.Ambrus-Lacatos, M.Schaffer, eds. *Forum Report of the Economic Policy Initiative*. No. 4. CEPR, London.
- [30] Boeri T., Garibaldi P. (2001) Shadow Activity and Unemployment in a Depressed Labor Market. IGIER Working Paper No. 77, 13 June. IGIER, Bocconi University, Milano, Italy.
- [31] Boeri T., Terrell K. (2002) Institutional Determinants of Labor Reallocation in Transition. *Journal of Economic Perspectives*, Vol. 16, No. 1, February, pp. 51-76.
- [32] Bouev M. (2001) Labour Supply, Informal Economy and Russian Transition. William Davidson Working Paper No. 408, October. William Davidson Institute, University of Michigan Business School, Ann-Arbor, USA.
- [33] Bouev M. (2002) Official Regulations and the Shadow Economy: A Labour Market Approach. William Davidson Working Paper No. 524, December. William Davidson Institute, University of Michigan Business School, Ann-Arbor, USA.
- [34] Bouev M. (2004) Diverging Paths: Transition in the Presence of the Informal Sector. William Davidson Working Paper No. 689, May. William Davidson Institute, University of Michigan Business School, Ann-Arbor, USA.
- [35] Bouev M. (2005) State Regulations, Job Search and Wage Bargaining: A Study in the Economics of the Informal Sector. William Davidson Working Paper No. 726, April. William Davidson Institute, University of Michigan Business School, Ann-Arbor, USA.
- [36] Brainerd E. (1998) Winners and Losers in Russia's Economic Transition. *American Economic Review*, Vol. 88, No. 5, December, pp. 1094-1116.
- [37] Bratkowski A., Grosfeld I., Rostowski J. (2000) Investment and Finance in *de novo* Private Firms: Empirical Results from the Czech Republic, Hungary and Poland. *Economics of Transition*, Vol. 8, No. 1, pp. 101-116.
- [38] Brixiova Z. (1997) On the Speed of Transition in Central and Eastern Europe: Does On-the-Job Search Matter? IMF Working Paper No. WP/97/102, August.
- [39] Brown D., Earle J. (2002) Job Reallocation and Productivity Growth Under Alternative Economic Systems and Policies: Evidence from the Soviet Transition. William Davidson Working Paper No. 514, November. William Davidson Institute, University of Michigan Business School, Ann-Arbor, USA.

- [40] Brown D., Earle J. (2003) The Reallocation of Workers and Jobs in Russian Industry. *Economics of Transition*, Vol. 11, No. 2, pp. 221-252.
- [41] Broyden C.G. (1969) A New Double-Rank Minimization Algorithm. *Notices of the American Mathematical Society*, Vol. 16, pp. 670-676.
- [42] Broyden C.G. (1970) The Convergence of a Class of Double-Rank Minimization Algorithms: The New Algorithm. *Journal of the Institute of Mathematics and its Applications*, Vol. 6, pp. 222-231.
- [43] Buchanan J. (1975) A Contractarian Paradigm for Applying Economic Theory. *The American Economic Review*, Vol. 65, No. 2, Papers and Proceedings, May, pp. 225-230.
- [44] Buchanan J., Lee D. (1982) Politics, Time, and the Laffer Curve. *The Journal of Political Economy*, Vol. 90, No. 4, August, pp. 816-819.
- [45] Burda M. (1993) Labour Markets in Eastern Europe. *Economic Policy*, No. 16, April, pp. 101-137.
- [46] Caballero R.J., Hammour M.L. (1996) On the Ills of Adjustment. *Journal of Development Economics*, Vol. 51, pp. 161-192.
- [47] Caballero R.J., Hammour M.L. (2000a) Institutions, Restructuring, and Macroeconomic Performance. NBER Working Paper No. 7720, May. National Bureau for Economic Research, Cambridge, MA, USA.
- [48] Caballero R.J., Hammour M.L. (2000b) Creative Destruction and Development: Institutions, Crises, and Restructuring. NBER Working Paper No. 7849, August. National Bureau for Economic Research, Cambridge, MA, USA.
- [49] Castanheira M. (2003) Public Finance and the Optimal Speed of Transition. *Economics of Transition*, Vol. 11, No. 3, pp. 435-462.
- [50] Castanheira M., Roland G. (2000) The Optimal Speed of Transition: A General Equilibrium Analysis. *International Economic Review*, Vol. 41, No. 1, February, pp. 219-239.
- [51] Castells M., Portes A. (1989) World Underneath: The Origins, Dynamics, and Effects of the Informal Economy. In: A. Portes, M. Castells, L.A. Benton, eds. *The Informal Economy. Studies in Advanced and Less Developed Countries*. The Johns Hopkins University Press, Baltimore and London.
- [52] Chadha B., Coricelli F. (1994) Fiscal Constraints and the Speed of Transition. CEPR Discussion Paper No. 993. Centre for Economic Policy Research, London, UK.
- [53] Chander P., Wilde L. (1992) Corruption in tax administration. *Journal of Public Economics*, Vol. 49, No. 3, December, pp. 333-349.
- [54] Chetvernina T., Lomonosova S. (2001) Sotsialnaya zaschischennost' nayomnykh rabotnikov v novom chastnom sektore: mify i realnost' (Social Security of Employees in the New Private Sector: Myth and Reality). *Voprosy Ekonomiki*, No. 9. In Russian.

- [55] Clarke S. (1999) *The Formation of a Labour Market in Russia*. Edward Elgar, Cheltenham, UK - Northampton, USA. 316p.
- [56] Clarke S. (2000a) Do Russian Households Have Survival Strategies? A Paper Prepared for ESRC Central and Eastern Europe Seminar Series on Household Survival Strategies, 26 February, Warwick University, UK. Available at <http://www.warwick.ac.uk/fac/soc/complabstuds/russia/workinprog.html>
- [57] Clarke S. (2000b) How do Russian Households Make Employment Decisions? A Paper Prepared for ESRC Central and Eastern Europe Seminar Series on Household Survival Strategies, 26 February, Warwick University, UK. Available at <http://www.warwick.ac.uk/fac/soc/complabstuds/russia/workinprog.html>
- [58] Clarke S., Varshavskaya L., Alasheev S., Karelina M. (2000) The Myth of the Urban Peasant. *Work, Employment and Society*, Vol. 14, No. 3, pp. 481-499.
- [59] Cockx B.L.W. (1997) Analysis of Transition Data by Minimum-Chi Square Method: An Application to Welfare Spells in Belgium. *Review of Economics and Statistics*, Vol. 79, No. 3, pp. 392-405.
- [60] Commander S., Kollo J., Tolstopiatenko A. (2004) Explaining Unemployment and Its Persistence in the Transition Countries: Some Conjectures. A Paper Presented at the Economics of Transition Seminar, University of Oxford, 19 October.
- [61] Commander S., Lee U., Tolstopiatenko A. (1996) Social Benefits and the Russian Industrial Firm. In: S.Commander, Q.Fan, and M.Schaffer, eds. *Enterprise Restructuring and Economic Policy in Russia*. Washington D.C.: The World Bank/EDI.
- [62] Commander S., Tolstopiatenko A. (1996) Restructuring and Taxation in Transition Economies. The World Bank Policy Research Working Paper No. 1625, July.
- [63] Commander S., Tolstopiatenko A. (1997) A Model of the Informal Economy in Transition Economies. William Davidson Working Paper No. 122, November. William Davidson Institute, University of Michigan Business School, Ann-Arbor, USA.
- [64] Commander S., Yemtsov R. (1995) Russian Unemployment: Its Magnitude, Characteristics, and Regional Dimensions, The World Bank Policy Research Working Paper No. 1426, February.
- [65] Cox D.R. (1972) Regression Models and Life-Tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 34, No. 2, pp. 187-220.
- [66] Davidson C., Martin L., Matusz S. (1987) Search, Unemployment, and the Production of Jobs. *The Economic Journal*, Vol. 97, No. 388, December, pp. 857-876.
- [67] Davis S.J., Haltiwanger J. (1999) Gross Job Flows. In: O.Ashenfelter, D.Card, eds. *Handbook of Labor Economics*. North-Holland, Amsterdam, pp. 2711-2805.
- [68] De Soto H. (1989) *The Other Path: the Invisible Revolution in the Third World*. Harper and Row.

- [69] Devine T.J., Kiefer N.M. (1991) *Empirical Labor Economics. The Search Approach*. Oxford University Press. Oxford. 343p.
- [70] Djankov S., La Porta R., Lopez-de-Silanes F., Shleifer A. (2002) The Regulation of Entry. *Quarterly Journal of Economics*, Vol. 117, No. 1, February, pp. 1-37.
- [71] Dolton P., Von der Klaauw W. (1995) Leaving Teaching in the UK: A Duration Analysis. *The Economic Journal*, Vol. 105, No. 429, March, pp. 431-444.
- [72] Doornik J.A. (2001) *Ox 3.0: An Object-Oriented Matrix Programming Language*. Fourth edition. London: Timberlake Consultants Press.
- [73] Earle J., Sabirianova K. (2002) How Late to Pay? Understanding Wage Arrears in Russia. Upjohn Institute Staff Working Paper No. 02-077, March. W.E. Upjohn Institute for Employment Research, Kalamazoo, MI, USA.
- [74] Earle J., Spicer A., Sabirianova Peter K. (2004) Community Norms and Organizational Practices: The Legitimization of Wage Arrears in Russia, 1992-1999. William Davidson Working Paper No. 641, January. William Davidson Institute, University of Michigan Business School, Ann-Arbor, USA.
- [75] EBRD (1999) Transition Report. European Bank for Reconstruction and Development. London.
- [76] EBRD (2000) Transition Report. European Bank for Reconstruction and Development. London.
- [77] EBRD (2003) Transition Report. European Bank for Reconstruction and Development. London.
- [78] EBRD (2004) Transition Report. European Bank for Reconstruction and Development. London.
- [79] Feige E.L., Urban I. (2003) Estimating the Size and Growth of Unrecorded Economic Activity in Transition Countries: A Re-evaluation of Electric Consumption Method Estimates and Their Implications. William Davidson Working Paper No. 636, December. William Davidson Institute, University of Michigan Business School, Ann-Arbor, USA.
- [80] Fiorentini G., Peltzman S. eds. (1995) *The Economics of Organised Crime*. Cambridge University Press. 301p.
- [81] Fletcher R. (1970) A New Approach to Variable Metric Algorithms. *The Computer Journal*, Vol. 13, No. 3, August, pp. 317-322.
- [82] Flinn C., Heckman J. (1982) Models for the Analysis of Labor Force Dynamics. NBER Working Paper No. 857, February. National Bureau for Economic Research, Cambridge, MA, USA.
- [83] Flinn C., Heckman J. (1983) Are Unemployment and Out of the Labor Force Behaviorally Distinct Labor Force States? *Journal of Labor Economics*, Vol. 1, No. 1, January, pp. 28-42.

- [84] Foley M. (1997a) Determinants of Unemployment Duration in Russia. Economic Growth Center, Yale University, Center Discussion Paper No. 779.
- [85] Foley M. (1997b) Labor Market Dynamics in Russia. Economic Growth Center, Yale University, Center Discussion Paper No. 780.
- [86] Foley M. (1997c) Multiple Job Holding in Russia During Economic Transition. Economic Growth Center, Yale University, Center Discussion Paper No. 781.
- [87] Fortin B., Marceau N., Savard L. (1997) Taxation, Wage Controls and the Informal Sector. *Journal of Public Economics*, Vol. 66, No. 2, November, pp. 293-312.
- [88] Friebel G., Guriev S. (2000) Why Russian Workers Do Not Move: Attachment of Workers through In-Kind Payments. CEPR Discussion Paper No. 3268. Centre for Economic Policy Research, London, UK.
- [89] Friedman E., Johnson S., Kaufmann D., Zoido-Lobaton P. (2000) Dodging the Grabbing Hand: the Determinants of Unofficial Activity in 69 Countries. *Journal of Public Economics*, Vol. 76, No. 3, June, pp. 459-493.
- [90] Fugazza M., Jacques J.-F. (2004) Labor Market Institutions, Taxation and the Underground Economy. *Journal of Public Economics*, Vol. 88, No. 1-2, January, pp. 395-418.
- [91] Garibaldi P., Brixiova Z. (1998) Labor Market Institutions and Unemployment Dynamics in Transition Economies. IMF Staff Papers, Vol. 45, No. 2, June.
- [92] Gavin M. (1996) Unemployment and the Economics of Gradualist Policy Reform. *Journal of Policy Reform*, Vol. 1, No. 3, pp. 239-258.
- [93] Gërxxhani K. (2004) The Informal Sector in Developed and Less Developed Countries: A Literature Survey. *Public Choice*, Leiden, Vol. 120, No. 3-4, pp. 267-330.
- [94] Gimpelson V. Lippoldt D. (1999) Private Sector Employment in Russia. Scale, Composition and Performance. *Economics of Transition*, Vol. 7, No. 2, pp. 505-533.
- [95] Gindling T.H., Terrell K. (2004) Legal Minimum Wages and the Wages of Formal and Informal Sector Workers in Costa Rica. IZA Discussion Paper No. 1018. IZA, Bonn, Germany.
- [96] Goldfarb D. (1970) A Family of Variable Metric Methods Derived by Variational Means. *Mathematics of Computation*, Vol. 24, No. 109, January, pp. 23-26.
- [97] Greene W.H. (2003) Econometric Analysis. Fifth (international) edition. Prentice Hall, Pearson Education International. 1026p.
- [98] Gregory P., Collier I. (1988) Unemployment in the Soviet Union: Evidence from the Soviet Interview Project. *American Economic Review*, Vol. 78, No. 4, September, pp. 613-632.

- [99] Grogan L. (2003) Worker Flows in the Russian Economic Transition: Longitudinal Evidence from Four Cities. *Economic Development and Cultural Change*, Vol. 53, No. 2, January, pp. 399-425.
- [100] Grogan L., Van den Berg G.J. (1999) The Duration of Unemployment in Russia. Research Memorandum 1999-I5, February. Serie Research Memoranda, Faculteit der Economische Wetenschappen en Econometrie, Vrije Universiteit, Amsterdam, Netherlands.
- [101] Gronau R. (1977) Leisure, Home Production, and Work - the Theory of the Allocation of Time Revisited. *Journal of Political Economy*, Vol. 85, No. 6, December, pp. 1099-1124.
- [102] Grosfeld I., Nivet J.-F. (1999) Insider Power and Wage Setting in Transition: Evidence from a Panel of Large Polish Firms, 1988-1994. *European Economic Review*, Vol. 43, No. 3-4, pp. 1137-1147.
- [103] Grout P. (1984) Investment and Wages in the Absence of Binding Contracts: A Nash Bargaining Approach. *Econometrica*, Vol. 52, No. 2, March, pp. 449-460.
- [104] Haltiwanger J., Lehmann H., Terrell K. (2003) Symposium on Job Creation and Job Destruction in Transition Countries. *Economics of Transition*, Vol. 11, No. 2, pp. 205-219.
- [105] Haltiwanger J., Vodopivec M. (2002) Gross Worker and Job Flows in a Transition Economy: an Analysis of Estonia. *Labour Economics*, Vol. 9, pp. 601-630.
- [106] Harris J., Todaro M. (1970) Migration, Unemployment and Development: A Two-Sector Analysis. *American Economic Review*, Vol. 60, March, 126-142.
- [107] Hausman J., McFadden D. (1984) Specification Tests for the Multiple Logit Model. *Econometrica*, Vol. 52, No. 5, September, pp. 1219-1240.
- [108] Heckman J., Singer B. (1984) A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data. *Econometrica*, Vol. 52, No. 2, March, pp. 271-320.
- [109] Hellman J., Jones G., Kaufmann D. (2000) "Seize the State, Seize the Day". State Capture, Corruption, and Influence in Transition. World Bank Policy Research Working Paper No. 2444, September.
- [110] Hoek J. (2002) Labor Market Institutions and Restructuring: Evidence from Regulated and Unregulated Labor Markets in Brazil. William Davidson Working Paper No. 484, July. William Davidson Institute, University of Michigan Business School, Ann-Arbor, USA.
- [111] Hosios A. (1990) On the Efficiency of Matching and Related Models of Search and Unemployment. *Review of Economic Studies*, Vol. 57, No. 2, April, pp. 279-298.
- [112] Ingster A.Y. (2003) Informal Contracts in Transition Economies. A Dynamic Approach. EERC-Russia Working Paper No. 03/05, Economics Education and Research Consortium, Moscow, Russia.

- [113] Jenkins S.P. (1998) Discrete Time Proportional Hazards Regression (or Estimation of Discrete Time (Grouped Duration Data) Proportional Hazards Models: pgmhaz.) Stata Technical Bulletin STB-39 (insert 'sbe17'). Available in: STB Reprints Volume 7, Stata Press, StataCorp, 302p.
- [114] Jenkins S.P. (2004) Survival Analysis. Unpublished manuscript, Institute for Social and Economic Research, University of Essex, Colchester, UK.
- [115] Johnson S., Kaufmann D., Shleifer A. (1997) The Unofficial Economy in Transition. *Brookings Papers on Economic Activity*, No. 2, pp. 159-239.
- [116] Johnson S., Kaufmann D., McMillan J., Woodruff C. (2000a) Why do Firms Hide? Bribes and Unofficial Activity after Communism. *Journal of Public Economics*, Vol. 76, No. 3, June, pp. 495-520.
- [117] Johnson S., McMillan J., Woodruff C. (2000b) Entrepreneurs and the Ordering of Institutional Reform. Poland, Slovakia, Romania, Russia and Ukraine Compared. *Economics of Transition*, Vol. 8, No. 1, pp. 1-36.
- [118] Jurajda S., Terrell K. (2000) Optimal Speed of Transition: Micro Evidence from the Czech Republic. William Davidson Working Paper No. 355, December. William Davidson Institute, University of Michigan Business School, Ann-Arbor, USA.
- [119] Jurajda S., Terrell K. (2002) What Drives the Speed of Job Reallocation During Episodes of Massive Adjustment. William Davidson Working Paper No. 432, January. William Davidson Institute, University of Michigan Business School, Ann-Arbor, USA.
- [120] Jurajda S., Terrell K. (2003) Job Growth in Early Transition: Comparing Two Paths. *Economics of Transition*, Vol. 11, No. 2, pp. 291-320.
- [121] Kamien M., Schwartz N. (2001) Dynamic Optimization. The Calculus of Variations and Optimal Control in Economics and Management. Second Edition. North-Holland. 380p.
- [122] Kartseva M. (2002) The Duration of Unemployment in Russia: Does Higher Education Decrease Unemployment Spells? NES Working Paper No. BSP/2002/058 E. New Economic School, Moscow, Russia.
- [123] Kaufmann D. (1997) Why is Ukraine's Economy - and Russia's - not Growing? *Transition*. The World Bank. April.
- [124] Kaufmann D., Kaliberda A. (1996) Integrating the Unofficial Economy into the Dynamics of Post-Socialist Economies: A Framework of Analysis and Evidence. The World Bank Policy Research Working Paper No. 1691, August.
- [125] Kehoe P. (1994) A Comment on *On the Speed of Transition in Central Europe* by P.Aghion and O.Blanchard. In: S.Fischer and J.Rotemberg, eds., *NBER Macroeconomics Annual*, Vol. 9, pp. 321-328. Cambridge, MA: MIT Press.

- [126] Kesselman J.R. (1989) Income Tax Evasion: An Intersectoral Analysis. *Journal of Public Economics*, Vol. 38, No. 2, pp. 137-148.
- [127] Keune M. (1997) Youth Unemployment in Hungary and Poland. Action Programme on Youth Unemployment. Employment Strategy Group, Employment and Training Papers No. 20, International Labour Office, Geneva. Available at <http://www.ilo.org/public/english/employment/strat/publ/etp20.htm>
- [128] Kiefer N.M., Neumann G.R. (1979) Estimation of Wage Offer Distributions and Reservation Wages. In: S.A.Lippman and J.J.McCall, eds. *Studies in the Economics of Search*. North-Holland, Amsterdam.
- [129] Kolev A. (1998) Labour Supply in the Informal Economy in Russia during Transition, CEPR Discussion Paper No. 2024, Centre for Economic Policy Research, London, UK.
- [130] Kolm A-S., Larsen B. (2001) Moral Costs, the Informal Sector and Unemployment. Working Paper No. 1-2001, Institut for Nationaløkonomi, Handelshøjskolen i København, Department of Economics, Copenhagen Business School.
- [131] Kolm A-S., Larsen B. (2004) Wages, Unemployment, and the Underground Economy. In: J.Agell and P.B.Sørensen, eds. *Tax Policy and Labour Market Performance*. MIT Press.
- [132] Lackó M. (2000) Hidden Economy - an Unknown Quantity? Comparative Analysis of Hidden Economies in Transition Countries, 1989-1995. *Economics of Transition*, Vol. 8, No. 1, pp. 117-149.
- [133] Lancaster T. (1990) The Econometric Analysis of Transition Data. Cambridge University Press, Cambridge. 352p.
- [134] Layard R., Richter A. (1995) How Much Unemployment is Needed for Restructuring: the Russian Experience. *Economics of Transition*, Vol. 3, No. 1, pp. 39-58.
- [135] Lehmann H., Wadsworth J.(2000) Tenures that Shook the World: Worker Turnover in Russia, Poland and Britain. *Journal of Comparative Economics*, Vol. 28, No. 4, pp. 639-664.
- [136] Lemieux T., Fortin B., Frechette P. (1994) The Effects of Taxes on Labour Supply in the Underground Economy. *American Economic Review*, Vol. 84, No. 1, March, pp. 231-254.
- [137] Lizal L., Svejnar J. (2000) Financial Conditions and Investment during the Transition: Evidence from Czech Firms. CERGE-EI Working Paper No. 153. CERGE-EI, Prague, the Czech Republic.
- [138] Ljungqvist L., Sargent T.J. (1998) The European Unemployment Dilemma. *The Journal of Political Economy*, Vol. 106, No. 3, June, pp. 514-550.
- [139] Loayza N. (1996) The Economics of the Informal Sector: A Simple Model and Some Empirical Evidence from Latin America. *Carnegie-Rochester Conference Series on Public Policy*, Vol. 45, pp. 129-162.

- [140] Luke P.J., Schaffer M.E. (1999) Wage Determination in Russia: An Econometric Investigation. CERT Discussion Paper No. 99/08. Centre for Economic Reform and Transformation, Heriot-Watt University, Edinburgh, UK.
- [141] MacLeod W.B., Malcomson J.M. (1993) Investments, Holdup, and the Form of Market Contracts. *The American Economic Review*, Vol. 83, No. 4, September, pp. 811-837.
- [142] Malcomson J. M. (1997) Contracts, Hold-up, and Labor Markets. *Journal of Economic Literature*, Vol. 35, No. 4, December, pp. 1916-1957.
- [143] Mascarenhas W.F. (2004) The BFGS Method with Exact Line Searches Fails for Non-Convex Objective Functions. *Mathematical Programming, Series A*, Vol. 99, pp. 49-61.
- [144] Matveenkov V., Vostroknoutova E., Bouev M. (1998) Transformational Decline and Preconditions of Growth in Russia. EERC-Russia Working Paper No. 98/03. Economics Education and Research Consortium, Moscow, Russia.
- [145] Meyer B.D. (1990) Unemployment Insurance and Unemployment Spells. *Econometrica*, Vol. 58, No. 4, pp. 757-782.
- [146] Mønsted T. (2000) Wage Differentials Between the Formal and the Informal Sector in Urban Bolivia. Instituto de Investigaciones Socio-Económicas, Universidad Católica Boliviana, La Paz, Bolivia. February.
- [147] Murphy K.M., Shleifer A., Vishny R.W. (1993) Why is Rent-Seeking So Costly to Growth? *American Economic Review Papers and Proceedings*, Vol. 83, No. 2, May, pp. 409-414.
- [148] Namazie C.Z. (2002) Who Bore the Burden of Wage Arrears in the Kyrgyz Republic? Discussion Paper No. DARP 64, November. Distributional Analysis Research Programme (DARP), The Toyota Centre, Suntory and Toyota International Centres for Economics and Related Disciplines, London School of Economics and Political Science, London, UK.
- [149] Narendranathan W., Stewart M.B. (1993) Modelling the Probability of Leaving Unemployment: Competing Risks Models with Flexible Base-line Hazards. *Applied Statistics - Journal of the Royal Statistical Society, Series C*, Vol. 42, No. 1, pp. 63-83.
- [150] Nash J. (1950) The Bargaining Problem. *Econometrica*, Vol. 18, No. 2, April, pp. 155-162.
- [151] Paxson C.H., Sicherman N. (1996) The Dynamics of Dual Job Holding and Job Mobility. *Journal of Labor Economics*, Vol. 14, No. 3, July, pp. 357-393.
- [152] Petrongolo B., Pissarides C.A. (2001) Looking into the Black Box: A Survey of the Matching Function. *Journal of Economic Literature*, Vol. 39, No. 2, June, pp. 390-431.

- [153] Pissarides F., Singer M., Svejnar J. (2003) Objectives and Constraints of Entrepreneurs: Evidence from Small and Medium Size Enterprises in Russia and Bulgaria. *Journal of Comparative Economics*, Vol. 31, pp. 503-531.
- [154] Pissarides C.A. (2000) *Equilibrium Unemployment Theory*. Second edition. Cambridge, MA, London, UK, MIT Press. 252p.
- [155] Popov V. (1999) Belka v kolese (A Squirrel in a Wheel). *Expert-Weekly*, No. 15, April. In Russian.
- [156] Pudney S., Thomas J. (1995) Specification Tests for the Competing Risks Duration Model: An Application to Unemployment Duration and Sectoral Movement. *Oxford Bulletin of Economics and Statistics*, Vol. 57, No. 3, pp. 323-347.
- [157] Puhani P. (1999) Unemployment Benefit Entitlement and Training Effects in Poland during Transition. William Davidson Working Paper No. 226, March. William Davidson Institute, University of Michigan Business School, Ann-Arbor, USA.
- [158] Raiser M. (1995) Governing the Transition to a Market Economy. *Economics of Transition*, Vol. 3, No. 2, June, pp. 215-246.
- [159] Richter A., Schaffer M. (1996) The Performance of De Novo Private Firms in Russian Manufacturing. In: S.Commander, Q. Fan, and M.Schaffer, eds. *Enterprise Restructuring and Economic Policy in Russia*. Washington D.C.: The World Bank/EDI.
- [160] RLMS (2004) A detailed project description is available online at <http://www.cpc.unc.edu/projects/rlms/project.html>
- [161] Rodrik D. (1995) The Dynamics of Political Support for Reforms in Economies in Transition. *Journal of Japanese and International Economics*, Vol. 9, pp. 403-425.
- [162] Roland G. (1996) Political Constraint and the Transition Experience. In: S. Zecchini, ed. *Lessons from the Economic Transition. Central and Eastern Europe in the 90s*. OECD, Kluwer Academic Publishers, pp. 169-187.
- [163] Roshchin S., Razumova T. (2002) Secondary Employment in Russia: Labor Supply Modeling. EERC-Russia Working Paper No. 02/07E, November. Economics Education and Research Consortium, Moscow, Russia.
- [164] Roshchin S., Markova K. (2004) Choice among Different Job Search Channels: The Evidence from Russian Labor Market. EERC-Russia Working Paper No. 04/05. Economics Education and Research Consortium, Moscow, Russia.
- [165] Ruggerone L. (1996) Unemployment and Inflationary Finance Dynamics at the Early Stages of Transition. *Economic Journal*, Vol. 106, pp. 483-494.
- [166] Rutkowski M. (1999) Russia's Social Protection Malaise: Key Reform Priorities as a Response to the Present Crisis. The World Bank Social Protection Discussion Paper No. 9909, April.

- [167] Sabirianova K. (2002) The Great Human Capital Reallocation. A Study of Occupational Mobility in Transitional Russia. *Journal of Comparative Economics*, Vol. 30, No. 1, March, pp. 191-217.
- [168] Sabirianova Peter K. (2003) Skill-Biased Transition: The Role of Markets, Institutions, and Technological Change. William Davidson Working Paper No. 616, October. William Davidson Institute, University of Michigan Business School, Ann Arbor, USA.
- [169] Safavian M, Graham D., Gonzalez-Vega C. (2001) Corruption and Microenterprises in Russia. *World Development* Vol. 29, No. 7, pp. 1215-1224.
- [170] Saint-Paul G. (1996) Dual Labor Markets: A Macroeconomic Perspective. MIT Press, Cambridge, MA, - London, UK. 218p.
- [171] Sandmo A. (1981) Income Tax Evasion, Labour Supply, and the Equity-Efficiency Tradeoff. *Journal of Public Economics*, Vol. 16, No. 3, December, pp. 265-288.
- [172] Sarte P-D.G. (2000) Informality and Rent-seeking Bureaucracies in a Model of Long-run Growth. *Journal of Monetary Economics*, Vol. 46, pp. 173-197.
- [173] Schmidt P., Witte A.D. (1989) Predicting Criminal Recidivism Using 'Split Population' Survival Time Models. *Journal of Econometrics*, Vol. 40, No. 1, January, pp. 141-159.
- [174] Schneider F. (2000) The Value Added of Underground Activities: Size and Measurement of the Shadow Economies and Shadow Economy Labor Force All Over the World. Johannes Kepler University of Linz, Austria. July.
- [175] Schneider F. (2001) The Size and Development of the Shadow Economies and Shadow Economy Labor Force of 18 Asian and 21 OECD Countries: First Results for the 90s. Invited paper prepared for the Asia Pacific Finance Association (APFA) Conference 2001 "Rebuilding Our Financial Architecture" held in Bangkok, Thailand, July 22-25, 2001.
- [176] Schneider F., Enste D.H. (2000) Shadow Economies: Size, Causes, and Consequences. *Journal of Economic Literature*, Vol. 38, No. 1, pp. 77-114
- [177] Shaked A., Sutton J. (1984) Involuntary Unemployment as a Perfect Equilibrium in a Bargaining Model. *Econometrica*, Vol. 52, No. 6, November, pp. 1351-1364.
- [178] Shakhnovich R., Yudashkina G. (2001) Wage-Setting and Employment Behavior of Enterprises during the Period of Economic Transition. EERC-Russia Working Paper No. 01/04. Economics Education and Research Consortium, Moscow, Russia.
- [179] Shanno D.F. (1970) Conditioning of Quasi-Newton Methods for Function Minimization. *Mathematics of Computation*, Vol. 24, No. 111, July, pp. 647-656.
- [180] Shapiro C., Stiglitz J. (1984) Equilibrium Unemployment as a Worker Discipline Device. *American Economic Review*, Vol. 74, June, 433-444.

- [181] Shimer R. (1997) The Optimal Speed of Transition. Princeton University. Mimeo.
- [182] Shleifer A., Vishny R.W. (1998) The Grabbing Hand. Government Pathologies and Their Cures. Harvard University Press, Cambridge, MA, - London, UK. 278p.
- [183] Slemrod J., Yitzhaki S. (2002) Tax Avoidance, Evasion, and Administration. In: A.J.Auerbach and M.Feldstein, eds. *Handbook of Public Economics*, Vol. 3, pp. 1423-1470. Elsevier Science B.V.
- [184] Smirnova N..(2003a) Re-employment Probabilities and Wage Offer Function for Russian Labor Market. William Davidson Working Paper No. 547, July. William Davidson Institute, University of Michigan Business School, Ann-Arbor, USA.
- [185] Smirnova N. (2003b) Job Search Behavior of Unemployed in Russia. William Davidson Working Paper No. 629, November. William Davidson Institute, University of Michigan Business School, Ann-Arbor, USA.
- [186] Sorm V., Terrell K. (1999) A Comparative Look at Labor Mobility in the Czech Republic: Where Have All the Workers Gone? William Davidson Working Paper No. 140, May. William Davidson Institute, University of Michigan Business School, Ann-Arbor, USA.
- [187] Sorm V., Terrell K. (2000) Sectoral Restructuring and Labor Mobility: A Comparative Look at the Czech Republic. *Journal of Comparative Economics*, Vol. 28, pp. 431-455.
- [188] Standing G. (1998) Reviving Dead Souls: Russian Unemployment and Enterprise Restructuring. In: S.Clarke, ed. *Structural Adjustment without Mass Unemployment? Lessons from Russia*. Edward Elgar, Cheltenham, UK - Northampton, MA, USA. 355p.
- [189] Stevens M. (2004) New Microfoundations for the Aggregate Matching Function. University of Oxford, Department of Economics Series.
- [190] Stiglitz J. (2002) Globalization and Its Discontents. Penguin Books, London, 288p.
- [191] Sutton J. (1986) Non-Cooperative Bargaining Theory: An Introduction. *The Review of Economic Studies*, Vol. 53, No. 5, October, pp. 709-724.
- [192] Svejnar J. (2002) Transition Economies: Performance and Challenges. *Journal of Economic Perspectives*, Vol. 16, No. 1, February, pp. 3-28.
- [193] Thomas J.J. (1992) Informal Economic Activity. LSE Handbooks in Economics. London. 371p.
- [194] Timofeev L. (2001) Tenevaya ekonomika i nalogovye poteri v selskom hozyaistve (Shadow Economy and Tax Losses in Agriculture). *Voprosy Ekonomiki*, No. 2. In Russian.
- [195] Trussell J., Richards T. (1985) Correcting for Unmeasured Heterogeneity in Hazard Models Using the Heckman - Singer Procedure. *Sociological Methodology*, Vol. 15, pp. 242-276.

- [196] Turunen J. (2000) Leaving State Jobs in Russia? Third IZA Summer School in Labor Economics, 29 May – 4 June.
- [197] Turunen J. (2004) Leaving State Sector Employment in Russia. *Economics of Transition*. Vol. 12, No. 1, pp. 129-152.
- [198] UNESCO (1998) World Statistical Outlook on Higher Education: 1980-1995. World Conference on Higher Education Higher Education in the Twenty-first Century: Vision and Action, UNESCO, Paris, 5-9 October 1998.
- [199] Van den Berg G.J. (2001) Duration Models: Specification, Identification and Multiple Durations. In: J. J. Heckman, and E. Leamer, eds. *Handbook of Econometrics*. Vol. 5. Elsevier Science B.V
- [200] Vaupel J., Manton K., Stallard E. (1979) The Impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality. *Demography*, Vol. 16, No. 3, pp. 439-454.
- [201] Vodopivec M. (2002) Worker Reallocation During Estonia's Transition to Market. *International Journal of Manpower*. Vol. 23, No. 1, pp. 77-98.
- [202] Voicu A. (2002) Labour Force Participation Dynamics in the Romanian Labor Market. William Davidson Working Paper No. 481, July. William Davidson Institute. University of Michigan Business School, Ann-Arbor, USA.
- [203] Vostroknutova E. (2003) Shadow Economy, Rent-Seeking Activities and the Perils of Reinforcement of the Rule of Law. William Davidson Working Paper No. 578, March. William Davidson Institute. University of Michigan Business School, Ann-Arbor, USA.
- [204] Wane W. (2000) Tax Evasion, Corruption, and the Remuneration of Heterogeneous Inspectors. World Bank Policy Research Working Paper No. 2394, July.
- [205] Wei S.-J. (1999) Corruption in Economic Development: Beneficial Grease, Minor Annoyance, or Major Obstacle? World Bank Policy Research Working Paper No. 2048, February.
- [206] Williamson O.E. (1975) Markets and Hierarchies: Analysis and Antitrust Implications. New York: the Free Press, London: Collier Macmillan Publishers, 286p.
- [207] Williamson O.E. (1985) The Economic Institutions of Capitalism. New York: the Free Press, 450p.
- [208] Yaari M.E. (1965) Uncertain Lifetime, Life Insurance, and the Theory of the Consumer. *The Review of Economic Studies*, Vol. 32, No. 2, April, pp. 137-150.
- [209] Yoon Y.-R., Reilly B., Krstic G., Bernabè S. (2003) A Study of Informal Labour Market Activity in the CIS-7. A Paper Prepared for the Lucerne Conference of the CIS-7 Initiative, January 20-22.