

Unity in diversity : studies on micro and macro panel data sets

Citation for published version (APA):

de Crombrugghe, D. P. I. (2010). Unity in diversity : studies on micro and macro panel data sets. Maastricht: Universiteit Maastricht.

Document status and date:

Published: 01/01/2010

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

UNITY IN DIVERSITY:

Studies on Micro and Macro Panel Data Sets

Denis de Crombrughe

ISBN: 978-90-5278-977-4

Copyright: Denis de Crombrughe

Cover credit: Martine Biebuyck

UNITY IN DIVERSITY:
Studies on Micro and Macro Panel Data Sets

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit Maastricht
op gezag van de Rector Magnificus,
Prof. mr. G.P.M.F. Mols,
volgens het besluit van het College van Decanen,
in het openbaar te verdedigen
op donderdag 30 september 2010 om 16:00 uur

door

Denis de Crombrugghe



Promotor:

Prof. dr. Franz C. Palm

Copromotor:

Prof. em. Anton P. Barten (Katholieke Universiteit te Leuven)

Beoordelingscommissie:

Prof. dr. Martin A. Carree (voorzitter)

Prof. dr. Lex Borghans

Prof. dr. Marno Verbeek (Rotterdam School of Management, Erasmus University)

Contents

Acknowledgements	5
1 Introduction	9
1.1 Pools and panels	9
1.2 Motivation	11
1.3 Outline	12
1.3.1 Two micro-panels, two macro-panels	12
1.3.2 Linked Estimation for multicountry data	13
1.3.3 Estimation of Matrix Variance Components	13
1.3.4 Maternal wage penalty	14
1.3.5 Consumption smoothing in Russia	14
1.3.6 Rule of Law and economic growth	15
1.3.7 The unity in the diversity	16
1.4 Setting the formal stage	16
1.4.1 High-dimensional data	16
1.4.2 Generic linear regression model	17
1.4.3 Specialisations	19
1.4.4 Simple regression models for multi-unit data	20
1.4.5 Classifying the applications	24
References	24
2 Linked Estimation	27
2.1 Introduction	27
2.2 Shrinkage, Pooling and Panels	29
2.3 A bilateral imports model	31
2.4 The parameter distribution	36
2.5 Linked Estimation: Theory	38
2.6 Linked Estimation: Practice	41
2.7 Concluding Remarks	46
References	47
3 Matrix Variance Components	51
3.1 Introduction	51
3.2 The model	52
3.3 Quasi-unbiased estimation of Φ	53
3.4 A Monte Carlo experiment	59
3.5 Concluding remarks	66
References	66

4	Maternal leave penalty	69
4.1	Introduction	69
4.2	The econometric model	72
4.3	The German SOEP data	75
4.4	Estimation results	78
4.5	Conclusion	84
	References	85
	Appendix 4.A: Alternative estimates of the wage equation	87
5	Consumption smoothing	91
5.1	Introduction	91
5.2	The adjustment process	95
5.2.1	Background	95
5.2.2	Static specification	97
5.2.3	Risk-sharing interpretation	97
5.2.4	Life-cycle interpretation	99
5.2.5	Equilibrium Correction	99
5.2.6	The GMM solution to the estimation problem	101
5.3	Data and measurement issues	102
5.3.1	Data	102
5.3.2	Composition of income and expenditures	103
5.3.3	Measurement errors	106
5.4	Estimation results	109
5.4.1	The incomplete panel 1994-2004	109
5.4.2	The annual panel 2000-2004	113
5.4.3	Further interpretations	115
5.5	Heterogeneity in smoothing behaviour	116
5.6	Conclusion	121
	References	121
6	Rule of law and growth	125
6.1	Introduction	125
6.2	Model formulation	130
6.3	Data	133
6.3.1	Data periodicity	133
6.3.2	Choice of indicators	135
6.3.3	Scaling of indicators	137
6.3.4	Other variables	137
6.4	Specification issues and Estimation Methods	138
6.5	Empirical Results	140
6.5.1	Baseline growth equations	140
6.5.2	Introducing the rule of law indicators	143
6.5.3	Fragility of effects of other regressors	147
6.5.4	Rule of law and the investment rate	147
6.6	Conclusions	148
	References	149
	Appendix 6.A: Tables of growth equations	154
	Appendix 6.B: Tables of investment equations	163

7 Conclusion	169
7.1 Looking back	169
7.2 Pondering on the assumptions	170
7.3 Looking ahead	172
Samenvatting in het Nederlands	175
Curriculum vitae	183

Acknowledgements

Life begins at 55!

Economists, in the words of Aunt Deirdre McCloskey (1996), are like boys playing in a sandbox: They build sand castles, and make them ever bigger. The greatest fun of all is smashing each other's. Here is my sand castle, ready for smashing at last, and I imagine Aunt Deirdre's surprise if she could see the loads of friendship and support received in the sandbox while it was in the making.¹ I wish to express thanks to my playmates who pushed, pulled, spurred or urged — so numerous that I am unable to name them all.

Thanks first to Professor Anton (Ton) Barten, my long-suffering master. He taught me the elements and symbols of our esoteric subject, aroused my curiosity, took me on as an assistant, and patiently dispensed lucid advice and thoughtful humour, even when wisdom and subtlety seemed lost on me. (I found notes of our first talk about writing a dissertation — his early cautioning now looks prophetic.)

Thanks to Franz Palm, my long-suffering boss. He employed me in his team, and made me feel fortunate that he did. Characteristic of his leadership is a calm, 'soft power' quality, which has determined the sound cooperative spirit of the entire department. I am grateful that he was willing to take on the role of supervisor along with Ton Barten.

Thanks to Martin Carree, Marno Verbeek, and Lex Borghans, who generously accepted to form my assessment committee. I value the input in time, attention and energy that they gave away. They took the task at heart and provided me with pertinent and helpful comments. Needless to say, although I did my best to implement each suggestion for improvement, they cannot be held responsible for what I made of it.

Thanks to Jean-Pierre Urbain for stimulating common interests, initiating joint work, and swapping palate-thrilling bottles. Although not a native, he is an ideal Maastricht character: keen and prolific at work, culturally hyperactive, jovial, and thoroughly Burgundian. Special appreciation to Franz and Jean-Pierre jointly for their vote of confidence, and their willingness to make things possible that hadn't been for a long time.

Thanks to all colleagues within the QE department for constituting such a friendly team. Sybrand Schim van der Loeff lavished his dry and roguish wit, eager encouragements, and insistent reminders; I benefited from his big-brotherly solicitude (even if he vehemently denies it). Dirk Tempelaar suffers from a rare form of generosity, including a stubborn and unlimited readiness to take over work from others, whatever the pressures

¹With apologies to Aunt Deirdre. See Chapter 1 of McCloskey D. N., *The Vices of Economists, the Virtues of the Bourgeoisie* (Amsterdam University Press, 1996).

on himself. I have been working with Franz, Sybrand, Jean-Pierre, Dirk, *et alii* for many years, and feel privileged. Thanks to Michael Eichler and Stephan Smeekes for taking over probability and statistics courses; to Yolanda Paulissen, Karin van den Boorn and Haydéé Hallmans for helping out on frequent occasions; and to non-QE colleagues in the recently rebaptised School of Business and Economics for very cordial working relationships.

Thanks to my coauthors, each of whom I first knew as a student: I learned a tremendous lot with you! Geert Dhaene is a long-time, brilliant friend with a variety of ‘encompassing’ talents (including things like saxophone, juggling and astronomy, apart from encompassing models); we share a taste for assiduous debates over sips of good wine. Christiane Arndt made her appearance more recently, in search of a Master thesis supervisor, with sweet, mostly harmless looks, but a fiercely resolute mindset, and guts to spare! She would never take an unclear explanation for an answer, and I think she could pep up a sloth over the telephone if she cared to. Geranda Notten was my sparring partner on the subject of poverty risk and made me discover the richness of the Russian Longitudinal Monitoring Survey; her enthusiasm was stimulating and her sense of good organisation proved precious. I enjoyed my late evenings at work with Raymond Montizaan, who was our time beater and proud to call himself a ‘slave driver’; as well as two of his smart ‘slaves’, Bianca Buligescu and Gülçin Menteşoğlu. I am grateful for the interaction with all three and, additionally, for Raymond’s editorial advice (which generally boils down to two words: “cut out”, applied generously).²

Thanks next to all my hosts in Maastricht, not only for the convenience, but even more for the warm-hearted friendship and the quality time. Maybe a hundred and one nights, I stayed with Franziska Gassmann and Willem Rijks for dinner, bed and breakfast, with a view on the Maas from the Kanaaldijk. It feels like going on a weekend in midweek. (I hope they will never move house any more.) Dirk and Concha Tempelaar sometimes put me up at a few minutes’ notice, and always had food, wine and a good book ready. Franziska, Willem, Concha and Dirk gave me and my family ample opportunity to experience that a treat in Maastricht is no Dutch treat.

Similarly, Tom and Thea van Veen opened their home to me and offered me the month-long hospitality of their family (Max, Jolien, Simon) in Newcastle, Australia, not to mention their unceasing moral support. Talking of which, I am itching to give thanks to Bill Mitchell, who invited me to stay at his department in Newcastle, and during his yearly visits to Maastricht spent entire afternoons persuading me to come down to earth and improve my ways.

For kind questioning and prodding I am grateful to many friends who shared their thoughts, among whom Michel Lubrano (my *compadre* in Marseilles, whose home is permanently open to me, and who gave me some penetrating comments too); Anca Roxana Varlan (Cinderella turned queen without the fairy help); and Bob Bartels (What is good research?). Early on, Jacques Drèze gave me a decisive nod. Recently, Ton van Attekum and Anton Schmidt provided effective encouragement. I spent a lot of hours in Hans Heijke’s office lately, and he supplied me just in time with patient and discerning editorial advice.

Moreover, I would like to mention the outstanding students in economics, business, public policy or its variants, and especially econometrics — for example, I fondly remember a particular quartet of German girls — who often forced me to clarify my thoughts or gave me a renewed sense of purpose.

Even though they are not getting the due proportion of space here, faithful friends from

²P.S. Surely this preface, for example, would have been briefer and more to the point, if only I had given him a chance.

outside academia have played a fundamental role in sustaining the balance, the drive and the confidence. Among many other things, I praise them for their companionship along all sorts of discovery trails. I delight in our inventive combinations of hiking, climbing, tasting, and light-hearted pondering.

In spite of the fact that I started off with long-suffering people there are plenty of them left for the end. To begin with, I take pleasure and almost pride in expressing gratitude to my parents whose patience I have been testing, I am afraid, in more ways than one. (You see, dear father and mother, I didn't stop improving!)

Together, Martine and I have been blessed with four lovely daughters who never took the slightest interest in economics, let alone econometrics. Nonetheless, they made sure that even a lengthy gestation never turned tedious, and to say the least they spiced up our lives from the day they were born. They might have inspired my dissertation subject by being all at once so strikingly resembling and so incredibly different. Talking of you, Delphine, Emilie, Flore and Anouck, brings me to one final thought.

Doing a PhD is like marrying. At first you desperately want it, and you swear fidelity. After the honeymoon you realise it's not *so* simple. A few years on, as the offspring grows wild, you begin to wonder how wise it all was. One day, with no-one watching, a bloom forms. Whether the ripe fruit is equally sweet I hope to find out soon. The next thing you expect to read at this point is a word of credit to one's partner for three or four years of unflinching support. It is my embarrassing luck to acknowledge *more than thirty*. Thank you, dear Martine, for bearing my whim and staying staunch (though not stony) like a Caryatid. My loving tribute to our better half is meant to come, as they say on such occasions, 'last but definitely not least'.



Chapter 1

Introduction

To Pool Or Not To Pool: That Is The Question.

(Maddala, 1991)

*True insight lies in recognizing the resemblance among things which differ,
and the difference between things which are alike.*

(Mme de Staël, 1810)¹

1.1 Pools and panels

In many areas of economic interest, the possibility to collect and exploit panel or panel-like data sets is now the rule rather than the exception. Thanks to the progress of information technology, keeping track records of a multitude of units has even become a matter of routine. For instance, marketeers keep files of potential customers, employers of employees, insurers of insured, investors of companies, governments of tax payers, and international organisations of member countries. The main chapters of this dissertation present empirical analyses of four different data sets with a panel or panel-like structure. The reason for bringing them together is that they require a similar methodology, essentially designed to combine cross-sectional and temporal dimensions of variation in economic data. By discussing a range of applications I hope to clarify this connection, contribute some solutions to theoretical or practical issues that arise, and maybe even help open up the field for more potential users of multi-unit data sets.

Originally, a panel referred to a group of individual respondents, and panel data would contain a number of individual track records over a relatively brief period of time. However, models and methods of analysis conceived with individuals in mind have a general structure that can also fit groups or entities like households, firms, governments, financial assets, etc. Hence the term ‘panel’ has become generic and should be understood now in a broad sense. When we speak of a panel or panel-like structure, we basically mean that sampling occurs in several dimensions, one of which is time.

Stimulated by the technological evolution, researchers have been developing models and methods to analyse panel data sets under various conditions and assumptions. The articles

¹Anne Louise Germaine Necker Baronne de Staël-Holstein, *De l’Allemagne* (1810), 3ème Partie, Chap. VIII: “Montesquieu dit que *l’esprit consiste à connoître la ressemblance des choses diverses et la différence des choses semblables.*” (Emphasis in the original.)

of Mundlak (1961), Hoch (1962), Zellner (1962), and Balestra & Nerlove (1966) may be said to have fulfilled the role of signposts. Since their publication a rich literature has developed (one is tempted to say ‘exploded’) which I like to arrange in three overlapping waves covering about five decades so far.

The first wave mainly dealt with linear panel data models in a context of strictly exogenous explanatory variables, and focussed on the question how best to exploit the cross-sectional variation in the panel – if at all. We inherited from that early phase many ideas and much historical terminology, in particular the dichotomy between ‘fixed effects’ and ‘random effects’ models — a choice of terms which doesn’t sound very fortunate nowadays but has become ingrained in the jargon of econometrics. Mundlak (1978) put this dichotomy in a new light, showing that what is at stake is less the random nature of the effects than how they do (or do not) relate to explanatory variables. Chamberlain (1982, 1984) confirmed Mundlak’s argument and furthermore pulled it to a much higher level of generality.

The use of random effects models was introduced in the panel data literature early on by Balestra & Nerlove (1966). At the same time, this seminal paper was explicitly concerned with the complications brought about by model dynamics. By so doing it precluded a second wave of panel data research, that would relax the strict exogeneity assumption and tackle several challenges: nonlinearity, errors in variables, model dynamics, and sample selectivity. As a far-off consequence, the acronym GMM would become almost as commonplace in econometrics as ‘old hat’ OLS, and more fashionable than former generalisations like GLS, 2SLS or LIML. The GMM principle was due to Hansen (1982), but the article that demonstrated its use in dynamic panel models was that of Arellano & Bond (1991). A number of extensions and improvements were proposed and qualifications made in the following decade by the same and several additional authors.

Out of theoretical necessity as well as for good practical reasons, this literature focused on the wide and short panel configuration naturally found in micro-economic data and often described as ‘large n , small T .’ The theoretical attractiveness of this design is largely due to the fact that it preserves standard limit theorems from being subverted by temporal issues of (non-)stationarity and long-memory dependence. In time series analysis, the standard limit theorems (laws of large numbers and central limit theorems for $T \rightarrow \infty$) break down when sequences are nonstationary or have too long memory, which explains the importance acquired by unit root and cointegration tests. In a ‘large n , small T ’ design, the accrual of information in a dimension other than time makes it possible for estimators and their sampling distributions to converge, however capricious their behaviour would be if prolonged over time.

The third wave of panel data research was triggered off by the invention of double-index asymptotics able to deal with wide *and* long panels, *i.e.*, the ‘large n , large T ’ design. A pair of articles by Phillips & Moon (1999, 2000) almost instantly became classic references. Phillips and Moon designed new tools of asymptotic analysis in order to investigate the consequences of nonstationarity over time and distinguish different types of sequential and joint limits. Their objective is a synthesis of convergence theories in the cross-section and over the time dimension. Such a synthesis will terminate the preponderance of micro-economic applications, since the participation of a person in a micro-panel tends (even more than the person herself) to be rather short-lived. We are more likely to see long track records for companies, organisations and countries than for individual people. Typical sources of data in this context may be macro-economic growth histories and on-line financial markets.

This third wave of panel data research is very much ongoing and growing, even as the

second wave is far from finished and the first one lingers on. An important new current in it is the progress being made on models with specific structures (dynamic common factors, spatial interactions) whose aim it is to deal with cross-sectional dependence in the panel; some recent references are Bai & Ng (2004), Pesaran (2006), and Pedroni (2007).

Other combinations of small and large sample sizes have attracted less specific attention. The asymptotic theory of time series can deal with cross-sections of time series of the ‘small n , large T ’ type without additional complications, at the cost of ignoring the cross-sectional dimension from which no continual accrual of information can be expected. As to the ‘small n , small T ’ case, asymptotic accrual of information is denied in either dimension, and attention is confined to (mostly) case-specific finite-sample considerations. Such data configurations are the privileged domain of (generalised) shrinkage estimators, which are deliberately biased in order to gain precision, and ultimately stem from James & Stein (1961). An alternative line of attack to short and narrow samples is Bayesian inference (with an informative prior).

What is it that makes the dichotomy between cross-section and time series variation so pervasive in statistics? It is relatively easy to draw practically random measurements from populations by slicing through space and sampling things, beings or groupings; this is because any one arrangement of observations will be essentially as good as (or, informationally equivalent to) any other. By contrast, the meaning of a sequence of measurements taken over time is strictly subject to its chronological ordering. There is a cleavage between what might be called (borrowing jargon from linguistics) the *synchronic* and the *diachronic* dimensions of data. They require different statistical concepts, models and methods, and it proves difficult to reconcile and integrate both types of information. As a consequence the analysis of panel and pseudo-panel data (the latter being the result of a timed sequence of independent cross-sections) is methodologically different from that of other complex data structures involving only a hierarchy of cross-sectional levels.

An essential feature of panel and panel-like data is that the cross-sectional units of observation (generically called ‘individuals’) may differ from one another. Transcending the three waves of panel data research imagined above, there has been a progression in the degree to which it was possible to tackle heterogeneity among cross-sectional units. Starting from the simple pooling of multi-unit data, the progression went on through models accounting for various shifts in fixed coefficients, heteroskedasticity, common error components or other forms of stochastic dependence, to models including general random effects (possibly correlated with regressors) and random coefficients (possibly depending on individual characteristics). Whereas the differences between models with increasing levels of sophistication may sometimes seem rather theoretical and even philosophical, they have practical implications and frequently lead to different inferences. It is therefore important to be aware of the various conditions under which the distinct methods will be reliable and, of course, to be able to assess in any particular application which conditions are more likely to approximate reality.

1.2 Motivation

The successive chapters of this dissertation present analyses of panel data sets of different types, bearing some light on common as well as distinguishing features of micro and macro applications, discussing methodological issues as they arise, and proposing operational solutions to a number of problems. The overarching purpose is to identify and exploit the potential benefits of combining cross-sectional and temporal dimensions of variation in economic data.

Two applications concern micro-economic panels involving a large cross-section of households or individuals, and two concern macro-economic panels involving a small or less small number of countries. A limited number of years of data is available in each case. For the purpose of long-run growth analysis, years are aggregated further into quinquennia (five-year periods) or even decades, so that the sample size in the time dimension becomes frankly small.

In each application, we pay care to resolve or mitigate the consequences of some problem(s) arising in panel and multi-unit data as a consequence of combining several dimensions of variation. In a nutshell, the issues addressed are: the estimation of matrix variance components; sample selectivity, unobserved heterogeneity and endogeneity in the earnings of working mothers; persistence and dynamics in the consumption expenditures of households exposed to income shocks; fragility of growth equations and of the effects of public institutions or good governance on the long-term growth of a country. In each case, the substantive conclusions are affected or qualified in significant ways. In some cases, we have to learn to live with the fact that, as we broaden the basis for inference, important substantive conclusions drawn previously do not necessarily become sharper and stronger but instead relative and fragile.

The methodological issues faced in the panel data field would be sufficient on their own to justify a dissertation (or much more ambitious projects). Nonetheless, the different areas of application treated here are of substantive interest as well. The topics of these applications will be outlined in the next section. Several of the chapters have been originally written together with graduate students turned colleagues in the meantime. In those cases, typically the specific topic was the initial motivation of my coauthors, and the methodology was what brought us together. The excitement of being able to contribute some relevant answers to substantive problems by the design and implementation of an appropriate statistical model provided the necessary stamina for developing the joint project further.

The common denominator between the various topics of empirical research was the opportunity to combine several levels of information or dimensions of variation in the data. Combining levels of information or dimensions of variation creates possibilities to study and model the behaviour of economic agents from new angles and in greater detail. Furthermore, it widens the scope for inference and stimulates the application of sound econometric methodology. One result, for beginners and more seasoned researchers alike, is self-protection against our natural inclination to stretch short data and to accommodate individual outcomes too well. Reconciling all the evidence is the ultimate goal towards which I hope, in a modest fashion, to make some strides.

1.3 Outline

1.3.1 Two micro-panels, two macro-panels

This section briefly introduces each of the following chapters, sketching the relevant contexts, and then briefly pointing out the methodological connections between them. Chapters 2 and 3 consider aspects of multi-country (macro-economic) model building where the time series dimension is not too short. Chapters 4 and 5 deal with micro-economic applications presenting various methodological complications. Chapter 6 returns to the macro-economy with its world-wide cross-sectional dimension, but with a view on long-term growth and development, implying that the period of observation is short in relation to the dynamic process of interest. A more formal exposition of the panel or multi-unit

data model used is proposed in a separate section which will conclude this introductory chapter.

1.3.2 Linked Estimation for multicountry data

Chapter 2 is of a macro-economic nature. The purpose is to present a practical method for binding together and ‘streamlining’ scattered country estimates of comparable parameters. The streamlining may improve the precision of the estimates and is important in the construction of useful simulation models. A random parameter framework with exchangeable distributions for multi-unit data is adopted, a ‘linked estimation’ procedure is developed, its sampling properties are examined, and it is shown how the parameter distribution hypothesis can be tested. The approach is akin to what has sometimes been referred to as ‘Best Linear Unbiased Prediction’ (BLUP) of random coefficients. The novelty is no longer in the model assumptions or even in the way the estimates are combined, but rather in the extension of the purpose and the scope of application. Whenever we are concerned with observation units which can be characterised by differences in degree rather than in kind, it is worthwhile to consider the adoption of a common theoretical framework and the effort of a joint analysis. In the words of Efron and Morris (1973), we are “combining possibly related estimation problems” and thus widening the basis for inference.

International trade being a natural component of multicountry models, a bilateral imports model serves as an important empirical illustration in this chapter. Simple bilateral import equations with income and price effects are estimated for three small open economies, namely, Belgium, the Netherlands and Denmark. Only a rather short sample of yearly data was available with the right level of detail and continuity. It is typical in such circumstances to be confronted with a proportion of counter-intuitive, stray and sometimes outlying estimates. This happens even with simple and straightforward models, like our import equations. Our exercise proves successful as it is seen how stray estimates with large standard errors are brought into line with the more precisely estimated ones. The estimates are used to generate out-of-sample predictions and the cross-sectional linkage is found to improve the forecast performance substantially (roughly a quarter of root mean square prediction errors; over 40% of mean square prediction errors).

This chapter appeared in the *Review of Business and Economics* (then called *Tijdschrift voor Economie en Management*) in 1995, under the title “Linked Estimation for Units of the Same Kind, or: How They Are All Alike and Yet So Different.”

1.3.3 Estimation of Matrix Variance Components

One practical need that had to be dealt with in the preceding chapter turns out to be a familiar problem in the implementation of random components and random parameter models more generally. In such models one requires estimates of the variance components (or variance component matrices) that are not only statistically but also logically consistent with the structure of the model. This is the issue addressed in the third chapter, coauthored with Geert Dhaene. Even in a well-specified model, sampling variation will cause standard unbiased variance component estimators to turn out negative in a proportion of cases. More generally, estimates of additive components of variance-covariance matrices will often fail to be positive semidefinite. In the case of a scalar variance component (*e.g.*, for a single random coefficient), it is straightforward to replace a negative variance estimate with zero. The case of a matrix variance component (*e.g.*, for a random

coefficient vector) is more complex. We address this issue and propose a new solution to the particular problem of estimating matrix variance components.²

The basic idea is to generalise the standard treatment of the scalar case by imposing upon the matrix estimator the restriction that it must be positive semidefinite. Unbiasedness is imposed only insofar as it is compatible with this restriction. The estimator solving this dilemma turns out to be at least as close to the true parameter – in an appropriate metric – as the usual unbiased estimator. It is obtained by substituting zeroes for any negative roots in a generalised eigenvalue-eigenvector decomposition of the unbiased estimator. Alternatively, it can also be found by a simple iterative procedure. We present some experimental evidence on its performance. It is surprising that more use has not been made in recent years of this sort of approach to the estimation of matrix variance components and other constrained estimation problems in econometrics.

1.3.4 Maternal wage penalty

With this chapter we turn to a micro-econometric panel analysis based on the German Socio-Economic Panel (GSOEP). The focus of this analysis is the size of the wage penalty that working mothers suffer on the German labour market, in particular when they take maternal leave. We realised that, to measure the penalty with any confidence, it is necessary to control for possible self-selection in and out of work, and for the choice of working hours. Existing estimates mostly disregard the variation in working hours and suggest two-digit penalties of up to 30 percent, with little rebound over time. Given this high figure (and besides also the continued findings of apparent gender discrimination in the wages) we decided to find out what a careful analysis of our data would have to say.

Hence, we applied recent panel data methods designed to address problems of sample selectivity, unobserved heterogeneity and endogeneity to the wages and working hours of women participating in GSOEP. The selectivity issue arises mainly because no wage is observed for employees who are on leave. Heterogeneity takes the form of unobserved individual effects likely to be correlated with explanatory variables. Endogeneity is due to the simultaneity of the wage and participation outcomes. As shown in a working paper by Semykina & Wooldridge (2006), Heckman's classic treatment of selectivity can be extended to deal with both heterogeneity and simultaneity. We present an extension for the case of a censored tobit participation model and use it to exploit the actual working hours data available in GSOEP. We also investigate the sensitivity of the results to the choice of method. Our estimates imply a wage penalty due to maternal leave which although substantial remains below previous estimates. Furthermore, we find that this penalty is less persistent than other studies suggest. Five years after their career interruptions German mothers who took a break to rear their children seem to have caught up entirely. They are not condemned to lower wages forever.

This chapter grew from a joint paper with three graduate students: Bianca Buligescu, Gülçin Menteşoğlu, and Raymond Montizaan. It was published in *Oxford Economic Papers* as Buligescu *et al.* (2009). Since our findings were clearly more optimistic than those of many previous studies, we expect continued interest in this topic.

1.3.5 Consumption smoothing in Russia

With the next data set we travel further East, and concern ourselves with households as consumers of goods and services rather than as suppliers of labour. At a time of political

²For all I know this solution is still original at the time of writing.

upheaval and economic transition, Russian households after 1990 had to deal with the new systemic uncertainties in addition to an extended range of idiosyncratic risks: illness, disability or death of a household member, crime, job loss or crop failure. Poverty statistics show that many Russian households were unable to maintain their previous consumption levels, and that privation was a common experience. How did they manage in such a volatile economic environment? The coping strategy that was implicit in Friedman's (1957) permanent income hypothesis will be our focus in this chapter: that of intertemporal consumption smoothing. We wish to investigate to what extent Russian households in the transition period, when faced with an income shock, were able to temporarily maintain their consumption standards; how they would adjust their expenditures over a longer period; and what specific factors may have affected their ability to smooth their consumption expenditures.

An adjustment mechanism typical of macro-econometric empirical models is the Error Correction Model (ECM), rebaptised more recently to Equilibrium Correction Model (Clements & Hendry 1995). This specification offers flexible dynamics and makes an explicit distinction possible between short-term reactions and long-term equilibrium adjustments. There is no obvious reason why the same mechanism could not serve to model dynamic adjustment in a micro-econometric context. In joint work with Geranda Notten this idea is tried out on a large (and very unsettled) series of household surveys, the Russian Longitudinal Monitoring Survey (RLMS). The error correction mechanism allows us to disentangle the smoothing dynamics. We also pay attention to problems of heterogeneity among households and imputation errors in the income data.

Our findings confirm that during the transition, households found ways to smooth their food expenditures to a considerable extent, especially in the short run. We can draw no firm conclusions in the analysis of non-food consumption because of unfavourable specification tests. We obtain only limited evidence of systematic differences in smoothing ability across households, and in particular we conclude that lower smoothing abilities are not necessarily associated with increased poverty risk.

1.3.6 Rule of Law and economic growth

We now turn to macro-economic concerns of a global and long-run nature. There seems to be an emerging consensus in international policy circles, not only that the quality of state governance matters for economic growth and development, but also that it can be observed and recorded quasi-objectively; see for instance the influential series of World Bank reports by Daniel Kaufmann, Aart Kraay and co-workers, started in 1999 and ongoing. How robust is the evidence with respect to the choices of quality of governance indicators, sources of data, aggregation methodology, specifications and assumptions used? Does the accumulation of a comprehensive country panel including competing indicators and decades of data constitute a solution for concerns of fragility? Can time recursivity help resolve the endogeneity problem? The aim of the sixth chapter, co-authored with Christiane Arndt, is to measure the interaction between state governance and economic growth in a comprehensive, historical 'panel' of countries. We constructed a data base including the well-known Heston-Summers Penn World Tables (PWT 6.1), the Quality of Governance collection (QoG institute, University of Gothenburg), and a number of other sources. In particular we obtained a choice of 'Rule of Law' indicators and considered a variety of ways of modelling their effects in growth regressions. We assessed the sensitivity and the fragility of results depending on the choice of the proxy indicators for Rule of Law, the time period covered by and the periodicity of the data, and the details of the

specification used. We modelled heterogeneity in the form of unobserved random effects allowing correlation with the regressors in the spirit of Mundlak. We relied on time recursivity to identify the effect of governance on economic growth over the last decades. We found results to be extremely fragile. In particular, the choice of proxy for the Rule of Law determines the significance of the Rule of Law variable; additionally, it affects the significance of the control variables.

1.3.7 The unity in the diversity

It is hoped that the thematic diversity in these chapters will not mask their unity of purpose. The overarching objective is the reconciliation of the cross-sectional and temporal dimensions of variation in the data. This pursuit motivates and stimulates the search for and application of sound methodology in the analysis of panel or multi-unit data sets.

For instance, in the case of the micro-panel in Chapter 4 (maternal leave penalty in GSOEP), we are led to use what is known as random effects assumptions, albeit with some provision for the presence of correlation between the individual effects and initial or average levels of observed variables. In the case of other chapters, random effects assumptions are not tenable for the individual effects and so are avoided, although random coefficient assumptions for other effects may remain plausible and exploitable.

The realism of specific assumptions, not the general principles or the algebra, is what makes macro-panels and micro-panels different from each other. Although this is an obvious point, there is some tendency in applied work to implement available formulas without due consideration of the assumptions they implicitly rely on. Available options and computational power are combined at little marginal cost to produce a plethora of computer output and it becomes hard to make a convincing choice among the resulting variety of estimates and tests.

The following section proposes a bird's-eye view of the available methodology. Apart from introducing some formal notation and terminology, this is used as an opportunity to create some prior order among the types of structural models and methods to be referred to later.

1.4 Setting the formal stage³

The aim of this section is to provide an orderly characterisation of the models that we will be using in the sequel. The presentation is meant to culminate in the last subsection with the beginnings of a systematic (although still partial) classification of a variety of models, including ours, with randomly distributed parameters. I will also use the opportunity to introduce some of the formal notation and terminology needed later. Although this material is undoubtedly familiar, its introduction here is not superfluous in view of the lack of standardisation and the variety of notions and notations that are in circulation.

1.4.1 High-dimensional data

High-dimensional data are sets of observation points itemised along more than one dimension of variation. Simply put, this means that data points are identified by (at least) double or triple indexes. For instance, observations are recorded over time (indexed $t = 1, \dots, T$) on a number of cross-sectional units or individuals (indexed $i = 1, \dots, N$), for a number of variables (indexed $k = 1, \dots, K$), and one speaks of a cross-section of time series or, in

³This section may be omitted without great loss of continuity.

a micro-economic context, of a panel. Thus panel data combine statistical information on both inter-individual variation (at fixed points in time t) and inter-temporal variation (for specific individuals i). Ideally, panel data sets lend themselves to modelling characteristic differences between individuals as well as behavioural adjustments of individuals over time.

Alternatively, a periodical survey in a population will constitute a time series of cross-sections, sometimes called a pseudo-panel. The important difference with a true panel is that the successive cross-sections are drawn independently and it is not possible to track individual dynamics. Even more different is a cross-section of cross-sections: a set of surveys of customers in different department stores, or of stores in competing supermarket chains, or of households in different metropolitan areas, or of firms in different countries. Any clustered sampling survey is really a cross-section of cross-sections. It is easy to think of more complicated observational arrangements, and in practice the hierarchy of cross-sections often involves three stages or more. However, to fix the ideas and to simplify the vocabulary as well as notation, it is customary to ignore this variety of situations when the subject of study is not survey design.

1.4.2 Generic linear regression model

In a panel data context, one frequently encounters intrinsically nonlinear models and in particular categorical response models (with a nominal or ordinal dependent variable), count models (with a discrete cardinal dependent variable), and censoring models (with a continuous but limited dependent variable). Although part of the remarks that will follow should be generalisable to nonlinear models, they will only be formulated for the simpler case of a linear conditional expectation function. For now, we will also keep aloof from sampling issues that may induce nonlinearity, like selectivity and censoring.

We require notation and a generic linear regression model for two-dimensional data. Keeping in line with convention, let us call one dimension the *cross-sectional* dimension, with observation points referred to as ‘individuals’ and indexed by $i = 1, \dots, N$; and the other the *time series* dimension, with observation points indexed by $t = 1, \dots, T$. Actually, the time series length T may (and in general does) vary between cross-sectional units, and one speaks then of *unbalanced* data. For the present purposes we avoid the additional notational complexity. We adopt temporarily the following notations: Observable variables (‘data’) are denoted y_{it} (‘dependent variable’) and \mathbf{x}_{it} (an exhaustive vector of ‘explanatory variables’); the corresponding coefficients (unknown parameters) are denoted $\underline{\xi}_{it}$. We consider the linear regression functions

$$E(y_{it} \mid \mathbf{x}_{it}) = \underline{\mathbf{x}}'_{it} \underline{\xi}_{it}, \quad (1.1)$$

where $\underline{\mathbf{x}}_{it}$ and $\underline{\xi}_{it}$ are column vectors of order \underline{K} , and it is understood that $i = 1, \dots, N$ and $t = 1, \dots, T$. These regression functions are tautological in the sense that they are not falsifiable; every new observation implies its own set of parameters. They are said to be unidentified. Without constraints on the parameter vectors $\underline{\xi}_{it}$ the ‘model’ lacks structure; it consists of only NT stochastic equations in $NT\underline{K}$ unknowns. Different ways of adding identifying structure will, naturally, result in different models, and there is a bewildering variety of possibilities.

In daily practice it feels almost natural, if only as a matter of habit, to resolve the identification issue by dropping the time and individual indexes from the parameters, hence keeping $\underline{\xi}_{it}$ constant in both dimensions:

$$E(y_{it} \mid \mathbf{x}_{it}) = \underline{\mathbf{x}}'_{it} \underline{\xi}_{..}, \quad (1.2)$$

assuming $\xi_{it} = \underline{\xi}_{..} (\forall i, t)$. Given this simplification it is tempting to view the complete data set as one big homogeneous ‘pool’ and to estimate the parameter vector $\underline{\xi}_{..}$ by a single, pooled least-squares regression.

Alternatively, it is possible to average the data in one dimension and run a regression over the averages in the other. If the averages are taken over time this is called a ‘between’ regression; if the averages are taken over the cross-section we have a regression in the aggregates. There is only one dimension of variation left, and cases that can be seen as examples of this abound in the empirical literature. Many different strategies are possible, though. Why not first run regressions in one dimension, then average in the other? This might make sense if the ξ_{it} vary in the latter dimension only, and we are interested in estimating a central value around which such variation occurs. That is the main intuition behind random coefficient regression models, to be discussed in more detail in the sequel.

More often than not some scrutiny of the results obtained with such pooling and averaging methods will suggest the presence of heterogeneity in the residual variances and in some regression coefficients, especially intercepts. Depending on the subject matter, the heterogeneity may be an important feature or just a nuisance.

Whether or not heterogeneity in the variances (heteroskedasticity or varying volatility) is of interest *per se*, it always affects the sampling distribution of the estimators and hence should not be ignored. Furthermore, taking proper account of heteroskedasticity is usually a matter of efficiency, and in theory there are neat ways to do so (namely, ideal and feasible forms of GLS). In practice, one is often content to use a ‘robust’ variance formula that offers protection against misleading inference without going all the way towards achieving parametric efficiency.

Heterogeneity in the regression coefficients (slopes or intercepts or both) has more radical consequences than heteroskedasticity; it affects the correlation structure of the data. The best-known way to allow for arbitrary variation in intercepts is to ‘de-mean’ the data in one dimension (time or cross section) and then run a regression in the other. If de-meaning is done over time this is a ‘within’ or ‘fixed-effects’ regression; if de-meaning is done over the cross-section we have a regression in ‘deviations from the aggregate’. The second-best-known way to allow for arbitrary variation in intercepts is to first-difference the data over time, and run a regression in the first differences. Note, though, that there exist different ways of differencing data (see Verbeek, 1995). Allowing for arbitrary variation in slope coefficients is, of course, much more demanding.

The number of specialisations of (1.1) that will come up and compete for attention is mind-numbing, and unavoidably choices will have to be made. Before proceeding, a remark on terminology is in order. Following Zellner (1962), systems of the type (1.1) and (1.2) are known as ‘Seemingly Unrelated Regression Equations’ (SUR or SURE). The former, (1.1), is a very general (not estimable) version of SURE; the latter, (1.2), is the special case in which aggregation over the cross-section will cause no aggregation bias.

In this context, the term ‘multivariate regression model’ might be used to refer to an extended configuration, where the dependent variable y_{it} is actually a *vector* of dependent variables sharing the explanatory variables \underline{x}_{it} . This arises, for example, when the regression equations are obtained as the reduced form of a genuinely simultaneous system applying to each individual; or when several traits are explained in the same individuals. The same term ‘multivariate regression model’ might also be used to refer to a restrictive configuration where y_{it} remains scalar but \underline{x}_{it} does not vary at all in one of the two dimensions. This would arise if $\underline{x}_{it} = \underline{x}_i$. or $\underline{x}_{it} = \underline{x}_{..t}$, $\forall i, t$. In the sequel it will be assumed that at least some component of \underline{x}_{it} varies over both i and t .

1.4.3 Specialisations

If we are to make any analytical progress, we will need to come up with some compromise between (1.1) and (1.2). Fortunately, the generality of the equations in (1.1) is somewhat misleading. The regressor vector \underline{x}_{it} contains data on, say, explanatory attributes of individual i and external circumstances at time t , measuring how they differ across individuals and how they evolve over time — insofar as they do differ and evolve. Not all individual attributes change over time, and the distinction matters between those which do and those which don't. Similarly, not all circumstances differ between individuals, and the distinction matters between individual circumstances and common circumstances. It will therefore be interesting to split up the conditional expectation of y_{it} so as to isolate individual attributes that remain constant over time, and temporal circumstances that are common to all individuals.

To achieve this, we (re)define a vector x_{it} containing *only* those regressors that vary *both* across individuals and over time; we isolate time-constant regressors (observed individual characteristics) in a separate vector, say z_i ; and similarly we place time-varying circumstances common to all individuals (observed 'macro' variables) in a separate vector, say w_t . In addition to these, many models would contain binary indicator variables ('dummies') specific to each individual and every time period.

Distinguishing between all these elements, we can respecify the conditional expectation of y_{it} as follows:

$$E(y_{it} \mid i, t, z_i, w_t, x_{it}) = \alpha_i + z_i' \gamma + \eta_t + w_t' \delta + x_{it}' \beta_{it}. \quad (1.3)$$

Observed attributes that vary only between individuals are collected into the vectors $\{z_i, i = 1, \dots, N\}$; circumstances that vary only over time are collected into the vectors $\{w_t, t = 1, \dots, T\}$; and the vectors $\{x_{it}, i = 1, \dots, N; t = 1, \dots, T\}$ contain only data that vary in both dimensions, including any interaction terms involving both individual characteristics and time variables. The order of the vectors x_{it} and β_{it} is denoted by K . In addition, there are N individual-specific intercepts (or intercept shifts) α_i , called (*unobserved*) *individual effects*; and there are T time-specific intercepts (or intercept shifts) η_t , called *time effects*. Specification (1.3) is to be seen here as a restricted version of (1.1), with a substantially smaller set of variables left in the x_{it} term; conversely, it is a generalisation of (1.2).

Specification (1.3) has been drastically simplified by assuming that the coefficients (γ) of the time-constant regressors z_i , as well as the coefficients (δ) of the macro-regressors w_t , vary *neither* across individuals *nor* over time. Even so, there are still (far) too many parameters in relation to the data, and more structure is called for. The model can be restricted further by assuming

$$\delta = 0, \quad (1.4)$$

and even

$$\eta_t = 0, \quad \forall t. \quad (1.5)$$

Disposing in this way of time effects and any purely temporal factors is in principle unnecessary for identification, yet is sometimes done in order to facilitate notation, exposition or computation.

A restriction that is more useful for identification in many two-dimensional data sets, and that is also natural in a 'pooling' context, is that the coefficients of x_{it} vary along only one of the dimensions, say the cross-section, and remain constant in the other, say time:

$$\beta_{it} = \beta_i \quad \forall t, \forall i. \quad (1.6)$$

The classic references on models that do *not* assume (1.6) but let coefficients vary randomly in both data dimensions are Hsiao (1974, 1975).

Under (1.3), (1.4) and (1.6), the conditional expectation function for y_{it} becomes

$$E(y_{it} \mid i, t, z_i, x_{it}) = \alpha_i + z_i' \gamma + \eta_t + x_{it}' \beta_i. \quad (1.7)$$

The corresponding equation in stochastic realisation or ‘error’ formulation is

$$y_{it} = \alpha_i + z_i' \gamma + \eta_t + x_{it}' \beta_i + u_{it}, \quad (1.8)$$

where an ‘everything else’ term u_{it} (error term, disturbance) makes its appearance; u_{it} is assumed to have useful stochastic properties (such as being drawn independently of the other right-hand side terms, from a continuous and symmetric probability distribution remaining stationary along at least one of the data dimensions). Equivalently, stacking the T data points for individual i in vector form, we can write

$$y_i = (\alpha_i + z_i' \gamma) \mathbf{1}_T + \eta + X_i \beta_i + u_i, \quad (1.9)$$

where the vectors $y_i = [y_{i1}, \dots, y_{iT}]'$, $\mathbf{1}_T = [1, \dots, 1]'$, $\eta = [\eta_1, \dots, \eta_T]'$, and $u_i = [u_{i1}, \dots, u_{iT}]'$ are all of order $T \times 1$; and the matrix $X_i = [x_{i1}, \dots, x_{iT}]'$ is of order $T \times K$.

In much of the panel data literature it is customary to go one step further, and set the coefficients of x_{it} equal for all individuals as well:

$$\beta_i = \beta \quad \forall i. \quad (1.10)$$

The last equality is also a condition for exact aggregation and for strict pooling; tests for its validity have been developed, beginning with Zellner (1962). In our overview we will mention some more possibilities.

1.4.4 Simple regression models for multi-unit data

In order to concentrate on the main issues, I propose to focus on the simple regression with a single continuous regressor and cross-sectionally varying coefficients. This is a special case of the modelling equations (1.7)-(1.8)-(1.9); generalisations may be reintroduced later with relative ease. As indicated earlier we wish to distinguish two broad kinds of models:

- (i) wide-sense *fixed effects* models, where all coefficients are classically fixed parameters, which may differ arbitrarily across units; and
- (ii) wide-sense *random effects* models, where the parameters include randomly distributed coefficients, which can only vary independently of the conditioning variables.

The ‘wide-sense’ qualifier is necessary because the terminology of ‘fixed effects’ and ‘random effects’ usually refers to the treatment of the intercept terms (α_i) exclusively. There is a consensus (or at least a revealed preference) among modellers not to be more restrictive about intercept terms than about the slope coefficient(s). In other words, it is considered rather more likely that slopes would be equal or similar across units, than that intercepts would. The term ‘unit’ may refer to individual observation units or, more generally, to any relevant grouping of observations.

On the basis of how the main choices sketched above are made we can categorise simple regression models for multi-unit data into five main classes, starting from

$$E(y_{it} \mid i, t, x_{it}) = \alpha_i + x_{it}' \beta_i. \quad (1.11)$$

Note that for the sake of simplicity and clarity we ignore in this discussion any interest there might be in variance parameters. Making the treatment of variance parameters explicit requires another layer of structural choices to be made.

1. All coefficients are treated as fixed in the classical sense and identical across units:

$$\alpha_i = \alpha, \quad \beta_i = \beta \quad \forall i \in \{1, \dots, N\}. \quad (1.12)$$

This is the strictly homogeneous case. It justifies exact aggregation or alternatively pooling in the strictest sense. Since all units have the same parameters of interest, the pooled data basically constitute a repeated rather than a multi-unit sample. Estimation of the common coefficients (α, β) may proceed by a relatively straightforward ‘Pooled Least Squares’ procedure (Pooled LS). In this class as in the others, ‘Least Squares’ means Generalised Least Squares (GLS) rather than Ordinary Least Squares (OLS), in view of the likelihood of cross-sectional heteroskedasticity, cross-sectional correlation, serial autocorrelation, or a combination of these, in the disturbance terms.

2. All coefficients are treated as fixed in the classical sense, but only slopes are assumed identical:

$$\beta_i = \beta \quad \forall i \in \{1, \dots, N\}. \quad (1.13)$$

The intercepts $\alpha_1, \dots, \alpha_N$ are left free to differ arbitrarily across units — without restriction of mutual similarity, exchangeability, or independence of conditioning variables. With all slopes identical, this is the narrow-sense ‘Fixed Effects’ (FE) model. It is typically estimated by so-called ‘Least Squares with Dummy Variables’ (LSDV) or ‘Analysis of Covariance’ methods, where each observation unit (but one) is represented by a specific dummy variable. If only some slopes are assumed identical and others are not, we have a model marking the transition to the next class:

$$\beta_i = \beta \quad \text{some } i \in \{1, \dots, N\}. \quad (1.14)$$

3. All coefficients are treated as fixed in the classical sense and unrelated across units. They may all differ arbitrarily, unrestricted by any form of mutual similarity, exchangeability, or independence of conditioning variables. All that can be said in general is that their combined parameter space is the full Cartesian product of the individual ones, which is written

$$\alpha_1, \dots, \alpha_N \in \Theta_{\alpha_1} \times \dots \times \Theta_{\alpha_N}, \quad (1.15)$$

$$\beta_1, \dots, \beta_N \in \Theta_{\beta_1} \times \dots \times \Theta_{\beta_N}, \quad (1.16)$$

where Θ_{α_i} and Θ_{β_i} represent the parameter spaces of α_i and β_i , respectively. This is the typical ‘Seemingly Unrelated Regression Equations’ (SURE) case. Joint estimation may be more efficient than separate estimation inasmuch as units are related by correlation between unobserved disturbance terms, but this last remark also holds for the preceding classes and, indeed, for those to follow.

4. The slope coefficients are treated as fixed in the classical sense and equal across units, but the intercepts are allowed to vary only *at random* across units — where ‘at random’ essentially means *independently* of the x_{it} (or, more generally, of all the conditioning variables in the model). This is the narrow-sense ‘Random Effects’

(RE) model, also called ‘error components model’ or ‘variance components model’. We write:

$$\alpha_i \sim f_\alpha(\cdot), \quad \beta_i = \beta, \quad \forall i \in \{1, \dots, N\}, \quad (1.17)$$

where $f_\alpha(\cdot)$ represents a probability (mass or) density function describing the cross-sectional distribution of $\alpha_1, \dots, \alpha_N$ and thus, to some extent, controlling their spread. It should be understood that $f_\alpha(\cdot)$ is unaffected by conditioning on the x_{it} , i.e., $f_\alpha(\cdot) = f_\alpha(\cdot | X_1, \dots, X_N)$. Surveys where a (large) number of individuals are observed repeatedly are usually modelled in this way and the models are then referred to as ‘panel data models’. They have been the subject of an explosive number of methodological studies in the last few decades.

5. If both intercepts and slope coefficients are allowed to vary *at random* across units, we obtain the Random Coefficient Regression (RCR) model:

$$\alpha_i \sim f_\alpha(\cdot), \quad \beta_i \sim f_\beta(\cdot) \quad \forall i \in \{1, \dots, N\}. \quad (1.18)$$

Again, the essential characteristic of the coefficient distributions is that they are assumed *independent* of the x_{it} (or of all the conditioning variables in the model). For simplicity, (1.18) also supposes that α_i and β_i are distributed independently of *each other*, but this is not at all essential. Usually, $f_\alpha(\cdot)$ and $f_\beta(\cdot)$ are probability density functions (pdf’s) depending on some collective ‘hyperparameters’. In Bayesian terms, they would represent exchangeable (not necessarily independent) prior distributions. The randomness of the coefficients is supposed to reflect the actual, physical diversity of the units, but at the same time the nature of the pdf is to restrain the differences. Sometimes, it may be found desirable to free the intercepts from such restraint. The specification (1.18) can then be relaxed to a ‘mixed’ form,

$$\alpha_1, \dots, \alpha_N \in \Theta_{\alpha_1} \times \dots \times \Theta_{\alpha_N}, \quad (1.19)$$

$$\beta_i \sim f_\beta(\cdot) \quad \forall i \in \{1, \dots, N\}, \quad (1.20)$$

where we have arbitrarily differing intercepts next to randomly varying slopes. This case might also be categorised as a class of its own. Generally, if some coefficients vary at random across units while others remain fixed, we have a mixture of Classes 4 and 5 that is referred to as a ‘Linear Mixed Model’ (LMM). In social and biological sciences, RCR models for multi-stage surveys are more often called ‘multilevel models’ or ‘Hierarchical Linear Models’ (HLM).

Informally, the Classes 1 to 3 form a progression from parameter restraint to parameter freedom. Class 1 calls for the strictest kind of pooling, whereas Class 3 is the least restrictive one, implying the weakest kind of pooling. Similarly, Class 4 is more restrictive than Class 5. Moreover the models in those last two classes can be seen as compromises between the models in the first three. While it is true that randomly varying coefficients are almost sure to differ across units, they also tend to differ in a more orderly fashion than if they were varying wholly arbitrarily. The probability density functions governing the parameter distribution impose a stochastic restriction on the coefficients, leaving them *less freedom of variation* than models where the same coefficients are assumed different and fixed. Thus in the simple regression model where there is only one slope coefficient we identify a partial ordering between more and less restrictive classes of models.

Table 1.1 proposes a schematic summary, considering only one slope coefficient and an intercept, and furthermore ignoring variances. The degree of heterogeneity taken into

account increases as we move in the table from left to right and from top to bottom. Classes 1, 2 and 3 are the extremes, as suggested by their corner positions. Class 4 (RE models) is in between Classes 1 and 2, whereas Class 5 is in between Classes 1 and 3 or Classes 2 and 3, depending on how the intercepts are treated. A model from Class 2 and another from Class 5 cannot always be ordered. Only the lower triangle of the square is filled because of the consensus (or revealed preference) amongst modellers that intercepts should not be constrained more severely than slope coefficients.

The five classes of this typology have not been defined in an exhaustive way. As Table 1.1 and the preceding discussion suggest, it is not hard to think of excluded cases. Nevertheless, a large proportion of the situations meaningful in practice is covered conceptually. Furthermore, each class can be seen as including its Bayesian versions. Even though fixed in a classical sense, parameters may still be treated as random in a Bayesian sense, so as to reflect the subjective uncertainty about them. In Classes 1 to 3 this leads to differently constrained versions of the Bayesian multivariate model, and in Classes 4 and 5 to various Bayesian hierarchical models (with exchangeable prior distributions).

Table 1.1: Typology of multi-unit models

		<i>Slope coefficients:</i>	
		Identical	Entirely unrestricted ^b
		Differing at random^a	
<i>Intercepts:</i>			
Identical	Class 1 (Pooled LS)	—	—
Differing at random^a	Class 4 (RE models)	Class 5 (RCR, HLM)	—
Entirely unrestricted ^b	Class 2 (FE models)	Class 5 (RCR, HLM)	Class 3 (SURE)

Notes: (a) “At random” implies independently of conditioning variables.

(b) The nature and extent of (cor)relation with conditioning variables is unrestricted.

In models involving random slope coefficients, an important additional choice is whether to study the micro-relations holding for each individual observation unit, or only a common or representative relationship holding overall. In much of the literature, random parameters or, more accurately, the random components of parameters tend to be treated as nuisances (as ‘error components’) rather than as parameters of interest. Sometimes this is for lack of repeated data on the same units, in which case the problem is one of incidental parameters. Even when repeated data is available, there appears to be some reluctance, in a sampling theory context, to allow for a parameter of interest having been drawn from a probability distribution. One then prefers to assign the role of parameter of interest to a fixed hyperparameter, even though such an escape is unnecessary.

In principle the typology in Table 1.1 is easy to generalise, although the details can get messy. Its purpose at this time is to provide a raster for surveying the field at the conceptual level.

1.4.5 Classifying the applications

To conclude this introduction I will attempt to clarify the organisation of this thesis by reference to the typology of Table 1.1.

Chapters 2 to 6 use models belonging to different cells in the table. In order to situate these in the typology, the classes listed need to be generalised conceptually for multiple conditioning variables (with corresponding ‘slope’ coefficients), which may actually include predetermined lags of the dependent variables and even endogenous regressors. Allowing for such generalisations, Chapter 2 belongs to Class 5, with ‘Linked Estimates’ in a multi-country setting being justified by an underlying parameter distribution equivalent to a random coefficient assumption. Chapter 3 develops a solution to one vexing technical aspect of such models, namely the estimation of the variance-covariance matrix of the parameter distribution under the constraint that it must be positive semidefinite. Chapter 4 turns to the working women in the GSOEP panel, and belongs to Class 4 (although it includes a Mundlak-Chamberlain extension). Chapter 5, on the consumption smoothing behaviour of Russian households, and Chapter 6, which estimates growth equations for a world-wide panel of countries, deal with data sets of a very different nature. It is interesting that nevertheless they are both firmly rooted in Class 2, or at least the extension of fixed-effects models for dynamic panels.

Between Chapters 2 and 6, we will have been surveying countries and households, travelling through time, and visiting several topics of economic interest and human concern. A final chapter will then propose a brief recapitulation and attempt some overall conclusions. Some suggestions will be offered for further work in the field of panel-structured data sets, regarding both the methodology and the substantive issues.

References

- Arellano M. & S.R. Bond (1991): “Some Specification Tests for Panel Data: Monte Carlo Evidence and an Application to Employment Equations”. *Journal of Economic Studies* 58, 277-298.
- Arndt C.E. (2009): *Governance Indicators*. Maastricht University doctoral dissertation, Maastricht Graduate School of Governance dissertation series Nr. 3.
- Bai J. & S. Ng (2004): “A PANIC attack on unit roots and cointegration”. *Econometrica* 72, 1127-1177.
- Balestra P. & M. Nerlove (1966): “Pooling Cross-Section and Time-Series Data in the Estimation of a Dynamic Model: The Demand for Natural Gas”. *Econometrica* 34, 585-612.
- Bulgescu B., D. de Crombrughe, G. Menteşoğlu & R. Montizaan (2009): “Panel estimates of the wage penalty for maternal leave”. *Oxford Economic Papers* 61, i35-i55.
- Chamberlain, G. (1982): “Multivariate Regression Models for Panel Data”. *Journal of Econometrics* 18, 5-46.
- Chamberlain, G. (1984): “Panel Data”. Ch. 22 in *Handbook of Econometrics Vol. II*, ed. Z. Griliches & M.D. Intriligator, North-Holland, 1247-1318.
- Efron B. & C. Morris (1973): “Combining Possibly Related Estimation Problems” (with discussion). *Journal of the Royal Statistical Society B* 35, 379-421.
- Friedman, M. (1957): *A theory of the consumption function*. Princeton University Press.

- Hansen L.P. (1982): "Large Sample Properties of Generalized Method of Moments estimators". *Econometrica* 50, 1029-1054.
- Hoch I. (1962): "Estimation of Production Function Parameters Combining Time Series and Cross-Section Data". *Econometrica* 30, 34-53.
- Hsiao C. (1974): "Statistical Inference for a Model with Both Random Cross-Sectional and Time Effects". *International Economic Review* 15, 12-30.
- Hsiao C. (1975): "Some Estimation Methods for a Random Coefficients Model". *Econometrica* 43, 305-325.
- James W. & C. Stein (1961): "Estimation With Quadratic Loss". In *Proceedings of the Fourth Berkeley Symposium, Vol. 1*, University of California Press, 361-379.
- Mundlak Y. (1961): "Empirical Production Functions Free of Management Bias". *Journal of the Farm Economics* 43, 44-56.
- Mundlak Y. (1978): "On the Pooling of Time Series and Cross-Section Data". *Econometrica* 46, 69-85.
- Maddala G.S. (1991): "To Pool or Not To Pool: That is the Question". *Journal of Quantitative Economics* 7, 255-264.
- Necker A.L.G., Baronne de Staël-Holstein (1810): *De l'Allemagne*, 3ème Partie, Chap. VIII.
- Pesaran M.H. (2006): "Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure". *Econometrica* 74, 967-1012.
- Pedroni P. (2007): "Social Capital, Barriers to Production and Capital Shares: Implications for the Importance of Parameter Heterogeneity from a Nonstationary Panel Approach". *Journal of Applied Econometrics* 27, 429-451.
- Phillips P.C.B. & H. Moon (1999): "Linear regression limit theory for nonstationary panel data". *Econometrica* 67, 1057-1111.
- Phillips P.C.B. & H. Moon (2000): "Nonstationary panel data analysis: An overview of some recent developments". *Econometric Reviews* 19, 263-286.
- Verbeek M. (1995): "Alternative Transformations to Eliminate Fixed Effects". *Econometric Reviews* 14, 205-211.
- Zellner A. (1962): "An efficient method of estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias". *Journal of the American Statistical Association* 57, 348-368.

Chapter 2

Linked Estimation for units of the same kind¹

Festschrift article in honour of Professor Anton P. Barten, with thanks to him for being a great inspirer.

2.1 Introduction

Although Professor Barten's name is quoted frequently in reference to consumer theory, that is by no means the only area of economics and econometrics where he will leave his mark. One other area in which he consumed large amounts of energy and a considerable number of assistants is that of macroeconomic model building, in particular the design, construction and running of medium-term multicountry models. Only a fraction of the acquired experience has found its way to the learned press, like the 1976 presentation of the Comet II model in the *European Economic Review*, and the 1981 monograph on *Methodological Aspects of Macroeconomic Model Construction*. By the time of its arguably much belated and shyly publicised publication, the latter had been in use for many years, as lecture notes attaining the highest standards. More than one other book was written in several versions, treasured by students but never quite completed.

In multicountry modelling, as also in other instances of multi-unit modelling, one is studying simultaneously a number of observation units of the same general kind: besides countries, we might talk of industries, firms, market places, brands of products, etc. What makes such an exercise special, apart from the multiple dimensionality of the data, is that very often the same elements of theory apply equally well to the different units under investigation. For example, it would seem rather incoherent to invent a theory of consumption or import demand anew every time a consumption or import function has to be specified for another industrial country. Accordingly, in the numerous multicountry modelling exercises performed by or under the supervision of Professor Barten, common specifications

¹Published as "Linked Estimation for Units of the Same Kind, or: How They Are All Alike and Yet So Different", *Review of Business and Economics* (then *Tijdschrift voor Economie en Management*), Vol.XL, 3-4, 1995. Prepared for the Festschrift issue of this journal in honour of Professor Anton P. Barten. I am grateful to a referee for very useful suggestions concerning the organisation of the article and for several interesting references quoted below. Thanks are also due to Alexander Italianer for providing the illustrative data.

are a recurring characteristic (entailing of course different coefficient values, and allowing for the obvious exception of institutional relations); see, for instance, d'Alcantara and Barten (1984). We might say that, in the manner of individuals, the juxtaposed country models are all alike, and yet all very different.

One can think of several reasons why common specifications may be a desirable feature in multi-unit models. It creates, in principle at least, the possibility to test a general theory jointly rather than separately with individual samples. It makes sample-dependent respecification and overfitting much harder, and thus constitutes a counterweight to the temptations of data mining. Also, it places the odd or outlying point estimate in a broader perspective, offering (under favourable conditions) a range of comparable estimates for guiding action. Last but not least, it improves the scope for joint multi-unit model simulation by simplifying the model solution algorithms.

As I have come to learn it is one of Professor Barten's concerns, when constructing a model, to design the equations and monitor the coefficient estimates in such a way that the model will be operational for simulation purposes. Simulations are not only used to prepare predictive scenarios and to evaluate the effects of specific policy measures or other shocks exogenous to the system. Before such ultimate experiments can usefully be performed, simulation is often a necessary tool for establishing and quantifying the empirical model's properties (insofar as these are state-dependent and result from sets of complex interactions). Queer point estimates can be disastrous for simulation. It is therefore common practice to 'bring them into line' by fixing them to some safer value – preferably within the range of comparable estimates for the companion units. To many a practitioner, this probably seems a perfectly reasonable way to solve an ordinary outlier problem. Indeed, the possible causes of impaired precision are numerous and usually beyond one's control: multicollinearity, inadequate variation, measurement errors, other forms of 'short data', not to mention structural shifts and more fundamental specification problems.

Notwithstanding such educated pragmatism, one would like to avoid the topical and rather arbitrary interventions that 'streamlining' involves in practice. It was indeed one of Professor Barten's ideas early on to develop a more systematic 'streamlining' method. That method should not prevent point estimates to differ across units, but should bind or 'link' them all by a centripetal pull. In the process, it should duly take account of the respective standard errors: thus it should give more weight to the more precise estimates, and impose more 'streamlining' onto the less precise estimates. He coined the phrase *linked estimation* to describe such mutual binding of scattered estimates. That is how he provided both the subject and the title of this article.

The object of the chapter is to found, present and illustrate a practical method to implement Professor Barten's idea of linked estimation. It is closely related to a number of developments in the literature and, in particular, it involves the idea of random coefficients. This means that a coefficient may possess the character of *both* a parameter of interest *and* (the realisation of) a random variable. Although random coefficients do not appear in the usual version of the classical regression model, there is no incompatibility implied.

The novelty here is not so much in the basic model adopted nor in the way to tackle it, but rather in the extension of its purpose and scope of application. Whenever we are concerned with observation units that can be characterised by differences in degree rather than in kind, I suggest that we should consider the adoption of a common theoretical framework and the effort of a joint, or *linked*, analysis. This is not only a way to "combine possibly related estimation problems" (Efron and Morris, 1973) or to widen the basis for inference for some technical benefit. It is also a form of self-protection against our natural

inclination to stretch short data and to accommodate individual data series too well. Hence linking units of the same kind may be conducive to sound econometric methodology.

The chapter is organised as follows. The next section reviews some of the related developments in the statistical literature and in various applied fields. Section 2.3 introduces a simple bilateral imports model, which will serve as the empirical illustration in this chapter, international trade being a natural component of multicountry models. The bilateral import equations are estimated by Ordinary Least Squares (OLS) for three small open economies, viz. Belgium, the Netherlands and Denmark. Section 2.4 then sets the formal stage by outlining a multi-unit, distributed parameter modelling framework in more general notation. The proposed linked estimation procedure is presented in Section 2.5, its sampling properties are examined, and it is shown how the parameter distribution hypothesis can be tested. These notions are put to work in Section 2.6 for the bilateral imports model introduced earlier, and their worth in terms of predictive power is evaluated. Finally, the closing section outlines a few possible extensions and formulates some concluding remarks.

2.2 Related developments: Shrinkage, Pooling and Panels

Remedying excessive disparity between scattered estimates of directly comparable parameters is a theme with, already, a respectable history in statistics. For decades, a variety of estimation methods pursuing similar aims, often developed independently for different fields of application, have been proposed in the literature. Their purpose is described with phrases like ‘eliminating excess variation’, ‘exploiting cross-equation similarity’, ‘borrowing strength from collateral evidence’, ‘combining possibly related estimation problems’ (Efron and Morris, 1973), etc. Whereas the notion of ‘linkage’ has not been used in this context, the related notions of ‘shrinkage’ and ‘pooling’ are well established.

Shrinkage has its origins in the quadratic loss (or equivalently, scalar mean square error) arguments of James and Stein (1961). The shrinkage resulting from the so-called ‘Stein-rules’ may be towards the origin or towards some other value, like an overall (‘grand’) mean. However, being based on a scalar loss criterion, Stein-type shrinkage is accordingly scalar in nature. This is an unattractive restriction in a multicountry context since one would be reluctant in general to agglomerate the estimation errors for several countries into one single-valued loss function. Shrinkage estimators are applied in empirical finance, e.g., to estimate expected returns on a number of alternative assets, as in Jorion (1986), and to predict earnings per share for a number of firms, as in Landsman and Damodaran (1989).

As soon as they were proposed Stein-rules were also given various Bayesian rationales. Zellner and Vandaele (1974) reviewed such interpretations, and in their framework were able to extend the approach to the multivariate regression context. In doing so, they connected Stein’s shrinkage principle to the literature on *Bayesian hierarchical models* initiated mainly by Lindley and Smith (1972). Lindley and Smith propose the following hierarchy of distributions: the first stage is the distribution of the multi-unit sample data conditional on the individual parameters of all the units; the second stage is the prior distribution of all these individual parameters conditional on a (smaller) set of common, so-called hyperparameters; the third stage is the prior distribution of the latter hyperparameters conditional on another layer of prior parameters; etc. In practice the hierarchy is usually limited to three levels at most.

A full Bayesian analysis is demanding, in that it requires the complete specification and processing of all distributions involved. Still, depending on how the prior distributions are chosen and how point estimates are obtained from the posterior distribution, the hierarchical-Bayes scheme can be made to produce a form of matrix-weighted shrinkage that comes rather close (in practice, if not in spirit) to this chapter's 'linkage' procedure. Reference is made here to the so-called 'parametric empirical Bayes' sequential methods of, among others, Rosenberg (1973), Rao (1975), and Morris (1983). The combination 'empirical hierarchical Bayes' is popular in fields like biostatistics and educational psychology. Recent discussions and illustrations can be found, e.g., in Breslow (1990) and in Bryk and Raudenbush (1992).

Closer to home, we mention a couple of hierarchical-Bayes applications in marketing. Lenk and Rao (1990) use sales rates on seven very dissimilar innovative products to estimate seven not too dissimilar life cycles or diffusion models. These in turn provide a means of forecasting the adoption process of another new product early in its life-cycle, even before actual sales data are available for it. Blattberg and George (1991) analyse a vast data set comprising weekly sales (as registered by supermarket scanners), prices, and promotional actions on four brands of bathroom tissue from three grocery store chains – a total of twelve chain-brand combinations. For each of the twelve chain-brand data subsets, OLS estimation of a sales model provides an excellent fit, but this results in considerable intermodel variability and most haplessly in contradictory and implausible elasticity estimates. Blattberg and George demonstrate nicely how pooling the data by chain or brand into groups with identical (or exchangeable) prior distributions can exploit cross-equation similarities and reduce the variability of the elasticity estimates. They argue that their streamlined estimates are 'far more believable' and hence more likely to be useful to managers.

A third approach to multi-unit data sets, remote from both Stein-type shrinkage and the hierarchical-Bayes philosophy, is provided by the classical 'Error Components' regression model. This approach originated in econometrics during the nineteen-sixties with applications like those of Mundlak (1961), Hoch (1962), and Balestra and Nerlove (1966); they initiated an extensive literature on the problems of combining or *pooling* time-series and cross-section data. More recently, thanks to massive increases in electronic data handling and processing capacities, the emphasis has shifted towards large sets of micro-economic *panel data*. A comprehensive collection of articles on this subject is assembled in two volumes edited by Maddala (1993); for some recent treatments, see Hsiao (1986), Baltagi (1995), and the papers in Mátyás and Sevestre (1992).

Conventional error components regression and more elaborate panel methods are very suitable for the analysis of extensive multi-individual data sets. However, they are not ideal in other multi-unit settings where interest centres on the units themselves, like for instance in multicountry modelling. The reason is that they restrict the inference problem to one involving a reduced number of overall means and variance components common to all observation units (the 'hyperparameters' in the terminology of Lindley and Smith). The individual error components, called 'random effects', are then treated as nuisances merely modifying the error term and, hence, complicating the residual stochastic structure. This reflects the fact that micro-panels usually comprise a relatively huge number of individuals (say thousands) as compared to the number of observation points over time (perhaps less than 10). With such data sets one is almost by necessity interested in average or representative behaviour and not in individual detail. In the case of multicountry or other macro-panels, and in business applications like the marketing examples quoted above, one is primarily interested in modelling the individual units, not their representative average.

Typically in these cases the data involves comparable numbers of time periods and of cross-section units.

An interesting extension of the error components model is the Random Coefficients Regression (RCR) model of Swamy (1970), (1971). The extension essentially consists of allowing *all* regression coefficients to vary randomly across units, and not only the coefficients of what are in effect dummy variables. Swamy's interest concentrates on applying Generalised Least Squares (GLS) or Maximum Likelihood (ML) methods to estimate the overall ('hyper-')parameters common to all units, rather than the individual random coefficients; in the same manner as the error components literature concentrates on estimating overall means and variance components, rather than the individual random effects. Nevertheless, Swamy's model will provide us with an adequate statistical framework for the analysis of the multi-unit situation, albeit with different parameters of interest.

The problem of drawing inferences on the random effects or the random coefficients themselves has, indeed, received less attention in this branch of the literature. Yet it is an important problem in other fields, such as biostatistics and in particular statistical genetics. There it is associated with the topic of animal breeding and with the name of C. R. Henderson ((1975), to quote but one generally accessible article). An important contribution for our purposes is Harville (1976), and a recent discussion of the state of the art is provided by Robinson (1991). In econometrics, there is a rare Monte Carlo study by Kelejian and Stephan (1983). Arguably the subject deserves more interest from economists. One can see no fundamental reason why a parameter should lose its importance on account of being a draw from a probability distribution – irrespective of whether or not (s)he adopts a Bayesian standpoint. For instance, as I noted already, in the typical multicountry or multi-brand situation it is the individual units that naturally call for analysis. That is why this chapter pursues this last line, viz., the estimation of the randomly distributed parameters of comparable units.

Before proceeding, it may be useful to point out the distinction between the multi-unit approaches reviewed in this section, and the joint estimation of sets of 'Seemingly Unrelated Regressions' (SUR), as Zellner (1962) christened them. The regressions are 'seemingly unrelated' in the sense that they do not constitute a 'simultaneous equations' system: none of the 'dependent' variables appears as an 'independent' variable in another equation. In the multi-unit context a SUR system arises naturally when an equation is duplicated for each of a number of units of the same kind (e.g., an investment function for G firms, a consumption function for G countries, or a sales equation for G brands). Zellner's method consists of applying GLS to the whole system of equations in order to exploit any correlation that might exist between their respective disturbance terms. If such correlation is present this may lead to a substantial gain in efficiency. GLS, however, does not seek to exploit any similarity relationship between the parameters of the companion equations – which is the more or less explicit aim of the shrinkage, hierarchical Bayes and pooling methods referred to above. Either or both avenues to improve inference may be explored, and indeed in favourable cases it will pay to combine them.

2.3 A bilateral imports model for small open economies

International trade is one of the building blocks of multicountry macroeconomic models. Somehow, total imports, total exports and their breakdown into bilateral flows need to be determined for the different countries modelled, at whatever degree of product aggregation consistent with the domestic models. Italianer (1986) provides not only a general treatment

of this topic, but also a description of the very interesting data base he constructed for the purposes of a European multicountry, macro-sectoral model under the auspices of the (then) E.C. Commission. This data base contains trade data for five broad categories of goods, thirteen declaring countries (all OECD-members), and eighteen trade partners (the thirteen declaring countries plus five world zones). It is based itself on detailed OECD trade statistics, and covers the years 1963-1980. (Unfortunately, the prolongation of the data base seems to have been discontinued thereafter.) The distinguishing feature of this data set is that for each bilateral trade flow, it contains a value (in U.S. dollars), a price index (with base 1975=1), and a volume (in constant 1975 dollars); consistent series of volumes and prices at appropriate levels of aggregation still seem to be a rarity in this area. See Italianer (1986) for further details and references.

Our limited purpose here is to illustrate the typical multi-unit modelling situation. We will therefore consider only bilateral imports of consumer goods (one of the five categories available) by a small, open economy (for which price-takership, in the sense that it faces an infinitely price-elastic supply, is an acceptable assumption). Furthermore, following Barten and d'Alcantara (1976, 1977), we assume that total imports of consumer goods are determined at some previous stage in the decision-making process. The importer country has seventeen trade partners, to be identified by a subscript letter g ($g = 1, \dots, 17$); a subscript t will index the years. Adopting the basic approach of Barten and d'Alcantara (1976, equation (1); 1977, equation (3.1)), we specify the following simple bilateral import equation:

$$\ln m_{gt} = \alpha_g + \vartheta_g \ln m_t + \gamma_g \ln (pm_{gt}/pm_t) + \lambda_g \ln m_{g,t-1} + u_{gt} \quad (2.1)$$

where

- m_{gt} = imports of consumer goods from partner g in year t , in 1975 US \$
- m_t = $\sum_{g=1}^{17} m_{gt}$ = total imports of consumer goods in year t , in 1975 US \$
- pm_{gt} = the (bilateral) price index corresponding to m_{gt} (1975=1)
- pm_t = the (multilateral) price index corresponding to m_t (1975=1)
- u_{gt} = a serially uncorrelated, normally distributed random disturbance with mean 0 and variance σ_g^2 .

Equation (2.1) involves international trade variables only. It explains what part of the total volume of imports of consumer goods is purchased from partner g . The scale coefficient ϑ_g is therefore called the *allocation elasticity* by Barten and d'Alcantara. It is expected to be (at least predominantly) positive and ideally to average about 1 (for reasons of internal consistency). The price coefficient γ_g measures how the allocation is affected by changes in bilateral relative prices; it is called the *substitution elasticity*. Expressing flows as volumes and prices as relatives is a common way to impose homogeneity of degree zero in prices, a useful property in a medium-term simulation model. The substitution elasticities are expected to be negative, and it would be most reasonable if they were scattered around -1 . Both ϑ_g and γ_g actually measure short-term effects in equation (2.1). The lagged dependent variable allows for some delay in the adjustment due to various sources of inertia. For the lag distribution to make sense, the coefficient λ_g must be between zero (inclusive) and 1 (not inclusive). Then, in general, allowing for delayed adjustment will result in long-term effects larger than the short-term effects (see, e.g., Barten 1989, Sections 11.4-11.6).

Three importer countries seem able to satisfy the small, open economy assumption: Belgium, the Netherlands and Denmark. Only the data from 1963 to 1974 are used for

estimation; the remainder is saved for prediction tests. In the time dimension there are therefore only $T = 11$ observations (the first year being absorbed by the lagged variable in (2.1)). Zellner's SUR method is not feasible, since the number of observation units in the cross-section dimension is $G = 17$, which is greater than $T = 11$. Hence for each importer the 17 equations were estimated separately by OLS. The results were qualitatively very similar for the three importers considered. In order to save space, they will be presented summarily in general, with some more detail for Belgium only.

A preliminary F -test is applied, to test the extreme homogeneity assumption that all four coefficients in (2.1) and the residual variances are identical across the 17 trade partners. If this assumption were satisfied, this would justify plainly pooling the data for $g = 1, \dots, 17$ and estimating the coefficients only once with this extended sample; the upshot being a huge saving in degrees of freedom. The results of the test are reported in Table 2.1, Column (a). Not surprisingly, the null hypothesis of equality was very firmly rejected: all three F -statistics exceed the 0.0001 critical value. The chi-square test in Column (b) is to be discussed later (Section 2.6).

**Table 2.1: Tests of
(a) parameter equality and (b) the parameter distribution hypothesis**

Importing country	(a) $F(64, 119)$	(b) $\chi^2(64)$
Belgium	3.319	26.9
The Netherlands	5.932	30.6
Denmark	2.259	30.0
1% critical value	1.643	93.2

Notes: (a) Under the null hypothesis of parameter equality, $F(64, 119)$ has an F -distribution with degrees of freedom (64,119). The 0.001 and 0.0001 critical values are, respectively, 1.931 and 2.205.

(b) Under the parameter distribution hypothesis, $\chi^2(64)$ has a chi-square distribution with 64 degrees of freedom. The (left-tail) .001 and .0001 quantiles of this distribution are, respectively, 34.6 and 30.2.

The 17 OLS regressions for Belgium are reported in Table 2.2, and summary statistics on the OLS regressions for Belgium, the Netherlands and Denmark are collected in Table 2.3, panel (a). The goodness of fit and residual diagnostics are quite typical for this kind of regressions. For instance, the standard error of regression (s_g) is widely variable, which can be explained by the small number of degrees of freedom; the determination coefficient (R^2) is high, as usual when the dependent variable and one of the regressors are trending variables expressed in levels; the residual autocorrelation (r_1) is mostly insignificant, in part thanks to the presence of the lagged dependent variable as a regressor. The coefficient estimates themselves are broadly acceptable, but rather imprecise and quite variable too.

Table 2.2: The 17 separate OLS regressions for Belgium

Trade partner	Estimated parameters						Diagnostics		
	α_g	ϑ_g	γ_g	λ_g	s_g	R^2	r_1		
Denmark	-4.08 (3.90)	.66 (0.48)	.97 (2.05)	.50 (0.37)	0.26	0.74	-0.14		
West-Germany	-2.99 (.96)	.88 (.21)	-1.43 (.37)	.32 (.14)	.04	>.99	-.06		
Greece	-11.12 (2.39)	1.62 (.34)	1.48 (.92)	-.37 (.30)	.16	.93	-.32		
France	-2.80 (.81)	1.01 (.25)	-.60 (.27)	.15 (.21)	.05	.99	.28		
Ireland	-9.12 (3.43)	1.29 (.48)	-.44 (.88)	.17 (.33)	.17	.95	-.24		
Italy	-4.40 (1.14)	1.32 (.31)	-.48 (.39)	-.24 (.29)	.08	.98	.00		
The Netherlands	-.44 (.33)	.76 (.08)	-.70 (.25)	.13 (.11)	.02	>.99	.08		
Great-Britain	-3.32 (1.84)	1.14 (.32)	-.70 (.22)	-.03 (.20)	.09	.98	-.05		
Portugal	-1.01 (1.61)	.45 (.26)	-1.79 (.31)	.18 (.19)	.09	.98	-.61		
Spain	-4.29 (2.48)	.70 (.39)	.37 (1.56)	.55 (.27)	.19	.92	.29		
U.S.A.	-1.40 (.77)	.62 (.21)	-1.02 (.55)	.31 (.30)	.10	.92	.01		
Japan	-.14 (.93)	.45 (.25)	-1.07 (.52)	.04 (.48)	.11	.73	-.13		
Other OECD countries	-.25 (1.20)	.55 (.25)	-.22 (.67)	.28 (.36)	.09	.92	.24		
Socialist countries	-.30 (.99)	.41 (.25)	-.65 (.77)	.35 (.35)	.13	.84	.21		
Oil producing countries	-.06 (2.27)	.32 (.29)	.36 (.63)	.16 (.34)	.19	.59	.12		
Fast developing countries	-8.35 (1.81)	1.49 (.28)	-1.65 (.62)	-.02 (.16)	.11	.97	-.22		
Other developing countries	1.46 (2.00)	.50 (.18)	-1.08 (1.04)	.04 (.31)	.10	.72	-.08		

Notes: Standard errors are given in parentheses. The number of observations is 11 for all regressions (1964-1974).

s_g = residual standard error (trade partner g)

R^2 = coefficient of determination

r_1 = first-order residual autocorrelation coefficient.

Table 2.3: Summary of
(a) separate OLS and (b) linked estimates for model (1)

Importing country	Summary Statistic	(a) OLS estimates of			(b) Linked estimates of		
		ϑ_g	γ_g	λ_g	ϑ_g	γ_g	λ_g
Belgium	Smallest estimate	.315	-1.789	-.367	.430	-1.650	.118
	Average estimate	.834	-.509	.149	.734	-.778	.230
	Largest estimate	1.617	1.483	.547	1.042	-.124	.426
	'Wrong signs'	(0)	(4)	(4)	(0)	(0)	(0)
The Netherlands	Smallest estimate	-.329	-1.604	-.413	.065	-1.117	-.067
	Average estimate	.782	-.563	.192	.733	-.674	.291
	Largest estimate	1.914	.160	.693	1.242	-.391	.631
	'Wrong signs'	(2)	(4)	(6)	(0)	(0)	(1)
Denmark	Smallest estimate	-.056	-2.661	-.283	.112	-1.632	.131
	Average estimate	.771	-.678	.263	.629	-.838	.365
	Largest estimate	2.104	2.126	.494	1.306	.098	.535
	'Wrong signs'	(1)	(4)	(2)	(0)	(2)	(0)

The lack of precision of the estimates has of course to do with the small number of degrees of freedom. Inasmuch as they are of the expected order of magnitude, some scattering of the coefficients is neither unexpected nor undesirable in itself. Here, however, they extend beyond the expected ranges since, in each case, several estimates have the ‘wrong sign’. Note however that the estimates with a wrong sign are statistically insignificant (that is, indistinguishable from zero). We will see later (Section 2.6) that there is a sense in which the dispersion of the coefficients is systematically exaggerated. Panel (b) of Table 2.4 is also to be discussed later on.

2.4 The parameter distribution

To fix the ideas, let us talk of a cross-section of countries for which comparable time-series are available. (Generalisations for others types of units, and other types of data per unit, are straightforward.) There are G countries, indexed by $g = 1, \dots, G$ (mnemonic for group, the data being grouped by country). Let the model of interest be, for all g :

$$y_g = X_g \beta_g + u_g \quad g = 1, \dots, G \quad (2.2)$$

where, under ideal but not unusual conditions,

y_g	is a $T_g \times 1$	column of observations on a ‘dependent’ variable;
X_g	is a $T_g \times K$	matrix of observations on K ($< T_g$) nonstochastic explanatory variables, possessing full column rank K ;
β_g	is a $K \times 1$	column of unknown parameters of interest; and
u_g	is a $T_g \times 1$	column of unobservable random ‘disturbances’ with mean vector $E(u_g) = 0$ and covariance matrix $E(u_g u_g') = \Sigma_{gg}$.

The data in y_g and in the columns of X_g should represent analogous concepts for each of the G countries. More precisely, they should be defined so that the respective β_g measure directly comparable effects. The easiest way to ensure this is to have the specification (2.2) contain dimensionless parameters, like elasticities, propensities or shares. Parameters that are not dimensionless will generally involve units of measurement that are country-specific and if so they will not be directly comparable. It is possible to exclude some of the coefficients from the comparability requirement. For instance, typically the intercept may depend on measurement units even when the slope coefficients do not. Here, however, for the sake of simplicity, we assume that all components of β_g are dimensionless and directly comparable across countries. If necessary, the intercept is dealt with by transforming all variables into deviations from their means.

The importance of specifying directly comparable country models is to make provision for a formal *homogeneity* assumption about the coefficients: viz., that they emanate from a common, random *parameter generation mechanism*. Specifically, we assimilate β_1, \dots, β_G to random drawings from one and the same K -variate distribution with $K \times 1$ mean vector μ and covariance matrix Φ :

$$\beta_g \sim i.i.d.(\mu, \Phi) \quad g = 1, \dots, G. \quad (2.3)$$

This is known in the literature as a *random coefficients* assumption à la Swamy (1970, 1971). In Bayesian terms, (2.3) would be seen as implying *exchangeability* of the prior distributions for β_1, \dots, β_G . It should be emphasised that (2.3) by no means implies a

prior belief or an anticipation that β_1, \dots, β_G are *equal*. On the contrary, unless Φ were identically a zero matrix, β_1, \dots, β_G are expected duly to spread, with the extent of their spread restrained by (2.3). The smaller the variances in Φ , the tighter the parameter generation mechanism, and the less broad the spread will be; and conversely. In this way intrinsic, structural differences between countries can be reflected in the differences between their respective coefficients.

It may be noted that, although assumption (2.3) implies no anticipation of equal coefficient vectors, it does suggest that we do not possess, or do not wish to incorporate, any nonsample information about the respective coefficient vectors that is country-specific, or differentiated countrywise. However, (2.3) can also be generalised by letting μ be, for instance, a linear function of cross-sectional variables such as country characteristics.

Equations (2.2) and (2.3) are assumed to hold for all G countries. It will prove convenient to rewrite the whole system in more compact notation. Let \mathbf{y} , $\boldsymbol{\beta}$ and \mathbf{u} be the ‘grand’ vectors obtained by stacking the y_g , β_g and u_g , respectively; and $\dot{\mathbf{X}}$ be a rectangular block-diagonal matrix with the X_g matrices arranged along its main diagonal. Define the grand total of sample sizes:

$$G\bar{T} = \sum_{g=1}^G T_g. \quad (2.4)$$

Thus \mathbf{y} and \mathbf{u} have order $G\bar{T} \times 1$, while $\dot{\mathbf{X}}$ is $G\bar{T} \times GK$, and $\boldsymbol{\beta}$ is $GK \times 1$. Then (2.2) for all $g = 1, \dots, G$ collectively can be written as

$$\mathbf{y} = \dot{\mathbf{X}}\boldsymbol{\beta} + \mathbf{u} \quad \text{subject to} \quad \begin{cases} \mathbf{E}(\mathbf{u}) = \mathbf{0} \\ \mathbf{E}(\mathbf{u}\mathbf{u}') = \boldsymbol{\Sigma}. \end{cases} \quad (2.5)$$

The ‘grand’ covariance matrix $\boldsymbol{\Sigma}$ may reflect disturbance correlation across countries, or it may be block-diagonal. The blocks on the diagonal are the $\Sigma_{gg} = E(u_g u_g')$. For an analogous transcription of (2.3) let us define the following ‘stutter’ matrix consisting of ones and zeroes:

$$\mathbf{J} = \iota_G \otimes I_K \quad (2.6)$$

where

- ι_G is the $G \times 1$ vector of ones;
- \otimes denotes the Kronecker product;
- I_K is the identity matrix of order K .

In particular, $\mathbf{J}\boldsymbol{\mu} = \iota_G \otimes \boldsymbol{\mu}$, a $GK \times 1$ vector in which $\boldsymbol{\mu}$ is stacked G times. The grand coefficient vector $\boldsymbol{\beta}$ has the following structure:

$$\boldsymbol{\beta} = \mathbf{J}\boldsymbol{\mu} + \boldsymbol{\delta} \quad \text{subject to} \quad \begin{cases} \mathbf{E}(\boldsymbol{\delta}) = \mathbf{0} \\ \mathbf{E}(\boldsymbol{\delta}\boldsymbol{\delta}') = I_G \otimes \Phi \equiv \dot{\Phi}. \end{cases} \quad (2.7)$$

By virtue of (2.3), $\boldsymbol{\delta}$ is a $GK \times 1$ random vector with mean zero and block-diagonal covariance matrix:

$$\dot{\Phi} \equiv I_G \otimes \Phi, \quad (2.8)$$

where I_G is the identity matrix of order G .

The Random Coefficient Regression (RCR) approach of Swamy (1970, 1971), consists in eliminating $\boldsymbol{\beta}$ by substituting (2.7) into (2.5). This substitution produces the *marginal model*

$$\begin{aligned} \mathbf{y} &= \dot{\mathbf{X}}\mathbf{J}\boldsymbol{\mu} + \dot{\mathbf{X}}\boldsymbol{\delta} + \mathbf{u} \\ &= \dot{\mathbf{X}}\mathbf{J}\boldsymbol{\mu} + \mathbf{v} \end{aligned} \quad (2.9)$$

where the composite error term \mathbf{v} has the following moments:

$$\begin{cases} E(\mathbf{v}) &= \mathbf{0} \\ E(\mathbf{v}\mathbf{v}') &= \dot{\mathbf{X}}\dot{\Phi}\dot{\mathbf{X}}' + \Sigma \equiv \Omega. \end{cases}$$

In model (2.9), the mean vector μ (a ‘hyperparameter’ in the terms of Section 2.2) can be estimated by (some feasible version of) GLS.

Here we are, however, interested in estimating β rather than μ . Straightforwardly, GLS may be applied directly to model (2.5), disregarding (2.7). For given covariance matrices, this produces the best linear unbiased estimator

$$\mathbf{b} = \left(\dot{\mathbf{X}}'\Sigma^{-1}\dot{\mathbf{X}} \right)^{-1} \dot{\mathbf{X}}'\Sigma^{-1}\mathbf{y} \quad (2.10)$$

for the grand vector β . An instance of this is Zellner’s (1962) SUR estimator mentioned in Section 2.2, for cases in which Σ , though not block-diagonal, still possesses the structure of a Kronecker product. We now turn to the question how to utilise the information assumed in (2.3) (or, alternatively, in (2.7) and (2.8)) for the estimation of β .

2.5 Linked Estimation: Theory

A special feature of the multi-unit, distributed parameter framework set out in the last section is that we may contemplate two different levels of ‘repeated sampling’: in the conventional way, drawing from the population of \mathbf{u} ’s; and additionally, drawing from the population of β ’s (or, equivalently, δ ’s). Under the former, as is well-known, GLS is Best Linear Unbiased (BLU) for a fixed parameter. More precisely, in the present framework, the GLS estimator (2.10) is BLU for β *conditionally on the realised value of β* . Alternatively, taking both levels of repeated sampling into consideration, an estimator for β , say $\hat{\beta}$, may be evaluated in terms of its *unconditional*, or *marginal*, sampling properties. One then investigates the marginal sampling distribution of the estimation errors $(\hat{\beta} - \beta)$.

It follows that a natural generalisation of the BLU-criterion in a random coefficient framework is BLU-ness *with respect to unconditional repeated sampling*.

Let $E_{\delta}(\cdot)$ denote an expectation taken with respect to the distribution of δ , and $E_{\delta, \mathbf{u}}(\cdot)$ an expectation taken with respect to the distributions of both δ and \mathbf{u} . Then an estimator is *marginally unbiased* for β if it has the same expectation as β :

$$E_{\delta, \mathbf{u}}(\hat{\beta}) = E_{\delta}(\beta). \quad (2.11)$$

Equivalently, the estimation error must be, marginally, unbiased for zero:

$$E_{\delta, \mathbf{u}}(\hat{\beta} - \beta) = \mathbf{0}. \quad (2.11')$$

The marginal covariance matrix of the estimation error is

$$\begin{aligned} \text{ESM}(\hat{\beta}) &\equiv E_{\delta, \mathbf{u}} \left[(\hat{\beta} - \beta) (\hat{\beta} - \beta)' \right] \\ &= E_{\delta} \left\{ E_{\mathbf{u}} \left[(\hat{\beta} - \beta) (\hat{\beta} - \beta)' \mid \beta \right] \right\}, \end{aligned} \quad (2.12)$$

where $E_{\mathbf{u}}[\cdot \mid \beta]$ denotes an expectation taken with respect to the conditional distribution of \mathbf{u} given β (or, equivalently, given δ). The matrix in (2.12) may properly be called the

expected or marginal *Error Second Moment* (ESM) matrix. Minimising (2.12), in a matrix sense, under the constraint (2.11), over all linear estimation functions $\hat{\beta} = \mathbf{c} + \mathbf{C}\mathbf{y}$, will produce an estimator for β that is BLU with respect to unconditional repeated sampling, or BLU *Marginally* (BLUM). Because of the formal analogy with Goldberger's (1962) best linear unbiased prediction problem, this procedure has been designated Best Linear Unbiased *Prediction* (BLUP) in the biostatistical literature. (I prefer to reserve the term *prediction* for the act of forecasting future values of *observable* variables, like here those of \mathbf{y}).

Harville (1976) develops an extension of the Gauss-Markov theorem for a very general linear model containing both fixed and random coefficient vectors. His extension of the Gauss-Markov BLU criterion is equivalent to the BLUM or BLUP criterion above. The results to follow can be derived as specialisations of his general solution.

The expressions we present involve the joint GLS estimator \mathbf{b} for β , defined in (2.10), and its (conditional) sampling covariance matrix which we will denote

$$\mathbf{V} \equiv \left(\dot{\mathbf{X}}' \Sigma^{-1} \dot{\mathbf{X}} \right)^{-1}. \quad (2.13)$$

The BLUM estimator of β , say $\hat{\beta}$, satisfies the equation system

$$\begin{aligned} \hat{\beta} &= \left(\mathbf{V}^{-1} + \dot{\Phi}^{-1} \right)^{-1} \left(\mathbf{V}^{-1} \mathbf{b} + \dot{\Phi}^{-1} \mathbf{J} \hat{\mu} \right), \\ \hat{\mu} &= \frac{1}{G} \mathbf{J}' \hat{\beta}. \end{aligned} \quad (2.14)$$

Here $\hat{\beta}$ is presented as a matrix-weighted average of \mathbf{b} and $\mathbf{J} \hat{\mu} = \iota_G \otimes \hat{\mu}$. The importance of the GLS estimates \mathbf{b} will increase with their precision, as measured by \mathbf{V}^{-1} , and with the looseness of the parameter generation mechanism, as measured by Φ (since $\dot{\Phi} \equiv I_G \otimes \Phi$). Conversely, the importance of the overall mean $\hat{\mu}$ will increase with the lack of precision of and with the tightness of the parameter generation mechanism. The mean vector $\hat{\mu}$ itself is the (equally weighted) average of the segments $\hat{\beta}_1, \dots, \hat{\beta}_G$ of $\hat{\beta}$. Hence the tendency of the estimates to be pulled towards a common mean is a form of mutual attraction, explaining the term 'linked estimation'.

Eliminating $\hat{\mu}$ from (2.14) produces explicit expressions for $\hat{\beta}$, such as:

$$\begin{aligned} \hat{\beta} &= \left(\mathbf{V}^{-1} + \mathbf{A} \right)^{-1} \mathbf{V}^{-1} \mathbf{b} \\ &= \left(I_{GK} + \mathbf{V} \mathbf{A} \right)^{-1} \mathbf{b}, \end{aligned} \quad (2.15)$$

where we have introduced the $GK \times GK$ matrix

$$\mathbf{A} \equiv \left(I_G - \frac{1}{G} \iota_G \iota_G' \right) \otimes \Phi^{-1}. \quad (2.16)$$

Formula (2.15) expresses the 'linked estimators' $\hat{\beta}_1, \dots, \hat{\beta}_G$ (the segments of $\hat{\beta}$) as linear combinations of the complete set of GLS estimators b_1, \dots, b_G (the corresponding segments of \mathbf{b}). The weights depend on the relative precision of the GLS estimators and on the tightness of the parameter generation mechanism. Since (2.15) may be substituted back into (14b), the same can be done for $\hat{\mu}$. As a matter of fact, it turns out that $\hat{\mu}$ coincides with the GLS estimator of μ in the marginal RCR model (2.9).

The expressions in (2.15) are to a degree reminiscent of shrinkage estimators. This suggests that with $\hat{\beta}$ we trade off the (conditional) unbiasedness of \mathbf{b} in exchange for a smaller variability. Note, however, that the matrix product $\mathbf{V} \mathbf{A}$ in the last line of (2.15) is not necessarily positive semi-definite. That distinguishes $\hat{\beta}$ from all shrinkage estimators,

including generalised variants of ridge regression and Stein rules; see, for an overview, Vinod and Ullah (1981).

In view of the way $\hat{\beta}$ modifies \mathbf{b} it must be conditionally biased for β . From (2.15) the conditional bias is:

$$\begin{aligned} E_{\mathbf{u}} \left(\hat{\beta} - \beta \right) &= - \left(\mathbf{V}^{-1} + \mathbf{A} \right)^{-1} \mathbf{A} \beta \\ &= - \left(\mathbf{V}^{-1} + \mathbf{A} \right)^{-1} \mathbf{A} \delta. \end{aligned} \quad (2.17)$$

The last equality follows from the identity $\mathbf{A}\mathbf{J} = \mathbf{0}$. Marginally, as in (2.11), the bias vanishes:

$$E_{\delta, \mathbf{u}} \left(\hat{\beta} \right) = E_{\delta} \left(\beta \right) = \mathbf{J}\mu. \quad (2.18)$$

As a compensation for the conditional bias in $\hat{\beta}$, one would expect less variance than in \mathbf{b} . The conditional covariance matrix of \mathbf{b} is \mathbf{V} , defined in (2.13). The conditional covariance matrix of $\hat{\beta}$ can be found too and is the difference between \mathbf{V} and a positive semi-definite matrix. In this sense $\hat{\beta}$ has a smaller variance than \mathbf{b} , indeed.

Is the decrease in variance sufficient to make up for the bias? In view of (2.17), the answer is contingent on the realisation of β . The question can be examined further by comparing the conditional Error Second Moment (ESM) matrices of \mathbf{b} and $\hat{\beta}$. The ESM matrix combines both the bias vector (in the form of an outer product) and the covariance matrix of a vector estimator. It is the generalisation for multivariate estimators of the Mean Square Error (MSE) concept for univariate estimators. The conditional ESM matrix of \mathbf{b} is equal to \mathbf{V} since \mathbf{b} is conditionally unbiased. The conditional ESM matrix of $\hat{\beta}$ is the expression within the braces $\{\cdot\}$ on the right-hand side of (2.12). It turns out that the latter will most often be smaller in some matrix sense than the former, especially when the parameter generation mechanism is relatively tight (in the sense that $\hat{\Phi}$ is not large as compared to \mathbf{V}). There will be exceptions for very extreme realisations of β .

Considering that the conditional ESM matrix of $\hat{\beta}$, since it depends on β , is random and unknown, it is natural to ask what we may expect it to be 'on the average'. The expected or marginal ESM matrix of $\hat{\beta}$ is

$$\begin{aligned} E_{\delta, \mathbf{u}} \left[\left(\hat{\beta} - \beta \right) \left(\hat{\beta} - \beta \right)' \right] &= \left(\mathbf{V}^{-1} + \mathbf{A} \right)^{-1} \\ &= \mathbf{V} - \left(\mathbf{V}^{-1} + \mathbf{A} \right)^{-1} \mathbf{A} \left(\mathbf{V} + \hat{\Phi} \right) \mathbf{A} \left(\mathbf{V}^{-1} + \mathbf{A} \right)^{-1}. \end{aligned} \quad (2.19)$$

As to \mathbf{b} , its marginal ESM matrix remains equal to \mathbf{V} (which is also its conditional ESM and covariance matrix). The last expression in (2.19) shows that the marginal ESM matrix of $\hat{\beta}$ is necessarily smaller than \mathbf{V} by a positive semi-definite matrix. This is, indeed, a direct implication of the fact that $\hat{\beta}$ was derived as the BLUM estimator of β .

The practical side of this discussion is that most elements of $\hat{\beta}$, though biased towards a common average, will have smaller MSE than the corresponding elements of \mathbf{b} . Sometimes, for some country, an extreme realisation of β_g will coincide with a moderate variance in \mathbf{V} to provide the statistical exception to the rule. For the remaining countries, small-sample deficiencies of the GLS estimates will be alleviated by the linkage with comparable countries.

Since the linkage alleviates small-sample deficiencies, it appears that countries with abundant data have no need for it. How will they be affected if they are nevertheless included? This is one way in which asymptotic questions arise in this context. Actually, it is not hard to establish that, under the standard assumptions, GLS and linked estimators

are asymptotically equivalent. The asymptotic distributions of $\hat{\beta}_g$ and b_g are the same for $T_g \rightarrow \infty$. So, clearly, linking a large-sample country g to other countries entails no loss of efficiency for the estimation of β_g . What happens as data on country g accumulate can be analysed as follows: The weight in $\hat{\beta}_g$ of information extraneous to country g tends to zero, while the weight of the ‘indigenous’ information becomes predominant; soon, $\hat{\beta}_g$ and b_g become indistinguishable. Meanwhile, the linkage is more beneficial than ever to those other countries with small sample sizes and where the extraneous information is accordingly not dominated. Actually, the large-sample country will contribute most to the collateral information incorporated in the linked estimators: the weight of b_g in $\hat{\mu}$ will be maximal, and the convergence of $\hat{\beta}_g$ will benefit the estimates for the countries with relatively poor data.

A different type of asymptotic question is concerned with G , the number of countries available for linkage. As a matter of course, the number of unknown parameters increases with G . Yet at the same time, as will be intuitively clear, the potential benefit of linking the countries also increases with their number. This intuition can without difficulty be shown to hold in terms of marginal ESM. The conclusion is that it pays to estimate more parameters at once, by analysing jointly more units of the same kind.

This section should not be ended without presenting a test for the parameter distribution hypothesis (2.3). Under this hypothesis, and adding assumptions of normality, the statistic

$$\begin{aligned} q &= \mathbf{b}'\mathbf{A}(\mathbf{V}^{-1} + \mathbf{A})^{-1}\mathbf{V}^{-1}\mathbf{b} \\ &= \mathbf{b}'\mathbf{A}\hat{\beta} \end{aligned} \tag{2.20}$$

obeys a chi-square distribution with $(G - 1)K$ degrees of freedom. A variety of analogous tests for subsets of countries can also be devised.

2.6 Linked Estimation: Practice

In the previous section, we have swayed under the carpet the fact that in practice not only the coefficient vectors but also the covariance matrices are unknown. Before we can apply linked estimation on the bilateral imports model introduced earlier, we must deal with this fact in some way. In our random coefficients framework, there are two unknown covariance matrices: the disturbance variance-covariance matrix, Σ , and the cross-country variance-covariance matrix of the parameter generation mechanism, Φ . The customary way to deal with such unknowns in estimation functions is to replace them with consistent estimates. The resulting estimators are called *feasible*, in contrast with the ‘ideal’ formulae which contain unknowns and are therefore *unfeasible*. Feasible counterparts cannot attain the same optimal finite-sample properties as the unfeasible originals. Still, the use of feasible estimators is unavoidable by necessity; and it can be argued that those optimal sampling properties, even though lost strictly speaking, remain valid approximately.

In the case of Σ the GLS residuals coming with \mathbf{b} are available to construct an estimator. Some structure must then be imposed on Σ . For instance, the SUR estimator of Zellner (1962) requires Σ to be the Kronecker product of a $G \times G$ (cross-country) covariance matrix and a $T \times T$ identity matrix, assuming $T = T_g$, all g . This would be a natural specification in our case, were it not that $G > T$ and hence that such a covariance matrix cannot be estimated for want of degrees of freedom. In view of this, a natural and also

simpler choice is to let Σ be diagonal, with homoskedastic diagonal blocks:

$$\Sigma_{gg} = \sigma_g^2 I_{T_g}, \quad g = 1, \dots, G. \quad (2.21)$$

The only unknowns left in Σ are then the country variances $\sigma_1^2, \dots, \sigma_G^2$. Familiar estimators for these are the OLS residual variances, say s_1^2, \dots, s_G^2 . For each g , supposing $T_g \rightarrow \infty$, then s_g^2 is consistent for σ_g^2 . Furthermore, thanks to the diagonal structure of Σ , the formulae for the coefficient estimators simplify. The GLS estimator \mathbf{b} boils down to separate OLS estimators, and the calculation of $\hat{\beta}$ requires no matrix inversions of an order higher than K (the same order as in each single OLS regression; incidentally, *block*-diagonality of Σ is what matters to avoid high-order matrix inversions.)

The second covariance matrix to be dealt with is Φ . Different attitudes are possible. On the one hand, Φ (and chiefly its diagonal) can serve as a control on the degree of linkage to be applied. Its elements are then set at values reflecting the cross-country coefficient disparity one is willing to tolerate. The off-diagonal elements (the covariances) will often be set equal to zero. It is worthwhile to exploit even vague ideas one has about Φ , albeit in a conservative way.

On the other hand, Φ can also be estimated statistically on the basis of the cross-country sample variance of the respective GLS estimates $\mathbf{b}_1, \dots, \mathbf{b}_G$:

$$S_{\mathbf{b}} = \frac{1}{G-1} \sum_{g=1}^G (\mathbf{b}_g - \frac{1}{G} \sum_{h=1}^G \mathbf{b}_h) (\mathbf{b}_g - \frac{1}{G} \sum_{h=1}^G \mathbf{b}_h)'. \quad (2.22)$$

With respect to synchronous increases in (almost) all sample sizes T_g and G , this simple estimator will be consistent for Φ . Given finite sample sizes, however, $S_{\mathbf{b}}$ is biased upwards for Φ . Swamy (1970) proposed to subtract (an estimate of) the bias so as to obtain an unbiased estimator of Φ (his equation (2.12)). Further refinements are also possible. Nevertheless, $S_{\mathbf{b}}$ has the advantages of being a conservative estimator and, more importantly, of being always positive semidefinite, which is not the case with unbiased estimators. An even simpler estimator retaining both advantages is a matrix with the same diagonal as $S_{\mathbf{b}}$ but with all off-diagonal elements (the covariances) set equal to zero.

We are now ready to discuss some linked estimation results for the bilateral imports model (2.1). Starting from the separate OLS regressions of Section 2.3, diagonal estimates for Σ based on (2.21) and for Φ based on (2.22) were obtained. These estimates were substituted for their population analogues in (2.13) and (2.16), in order to calculate the feasible counterpart of the coefficient $\hat{\beta}$ in (2.15); the same was done for the chi-square statistic q in (2.20).

The resulting coefficient estimates for Belgium as the importer country are reported in Table 2.4, the ‘linked’ counterpart of Table 2.2. Next to the coefficients, in parentheses, their conditional standard errors are shown. It should however be kept in mind that these take no account of the bias in the coefficients, and consequently exaggerate the gain in accuracy. For ease of comparison, the separate (OLS) and linked estimates for the slope coefficients of the Belgian import equations are contrasted graphically in Figures 2.1, 2.2 and 2.3.

As it turns out, the linkage has neatly brought several conspicuous coefficients back into line and reversed the ‘wrong signs’. However, it should be acknowledged that the sign reversals are a virtuous by-product of the linkage, and not an algebraic necessity. As will be seen in the cases of The Netherlands and Denmark, desirable sign reversals will not always come about.

Table 2.4: The 17 linked regressions for Belgium

Trade partner	Parameter estimates							
	α_g		ϑ_g		γ_g		λ_g	
Denmark	-2.83	(1.76)	.62	(.23)	-.51	(.77)	.26	(.20)
West-Germany	-2.98	(.79)	.87	(.17)	-1.35	(.33)	.34	(.11)
Greece	-6.21	(1.64)	.91	(.22)	-.12	(.71)	.27	(.18)
France	-2.53	(.59)	.93	(.18)	-.61	(.26)	.22	(.14)
Ireland	-6.21	(1.70)	.90	(.23)	-.66	(.64)	.37	(.16)
Italy	-3.09	(.80)	.94	(.19)	-.51	(.39)	.12	(.18)
The Netherlands	-.48	(.32)	.75	(.08)	-.73	(.24)	.15	(.10)
Great-Britain	-2.20	(1.37)	.91	(.22)	-.70	(.21)	.12	(.14)
Portugal	-1.55	(1.20)	.52	(.18)	-1.65	(.29)	.17	(.14)
Spain	-4.91	(1.50)	.82	(.22)	-.48	(.79)	.43	(.16)
U.S.A.	-1.43	(.73)	.66	(.14)	-.97	(.47)	.25	(.19)
Japan	-.38	(.85)	.43	(.14)	-1.11	(.44)	.15	(.21)
Other OECD	-.73	(.97)	.63	(.16)	-.48	(.51)	.25	(.19)
Socialist	-.69	(.88)	.52	(.15)	-.84	(.53)	.23	(.19)
Oil producing	-2.09	(1.63)	.54	(.20)	-.22	(.48)	.19	(.20)
Fast developing	-5.49	(1.40)	1.04	(.21)	-1.10	(.55)	.21	(.12)
Other developing	.62	(1.24)	.50	(.13)	-1.19	(.65)	.18	(.19)

Notes: Conditional standard errors are given in parentheses.

The number of observations is 11 for all regressions (1964-1974).

From Figures 2.1 to 2.3 and Tables 2.2 and 2.4, it may be observed that the estimates with the largest standard errors have changed most when they were far astray. For instance, the substitution elasticities for Denmark and Spain, whose OLS estimates were positive and had the largest standard errors, have taken on average values (with much decreased standard errors). Conversely, very negative price elasticities that had a relatively small OLS standard error, those for Germany and Portugal, have hardly budged at all. After some scrutiny, it will also be noticed that the effect of the linkage is not uniform like that of an ordinary shrinkage. For instance, the price elasticities for Japan and for the zone of 'Other developing countries' move *away from*, rather than towards the average. The algebraic explanation is that matrix-weighted averages are much more flexible and less predictable than ordinary scalar-weighted averages.

Summary statistics for the three importer countries, Belgium, The Netherlands and Denmark, were already presented earlier in Panel (b) of Table 2.3. They exhibited a substantial decrease in the ranges of the estimates and in the number of 'wrong signs'. Note that the 'average estimates' in Panel (b) of Table 2.3 correspond to $\hat{\mu}$ in (2.14). In Figures 2.1 to 2.3, the average estimates for Belgium are represented by bold gridlines.

FIGURE 2.1: Allocation elasticities for Belgian imports, per trade partner

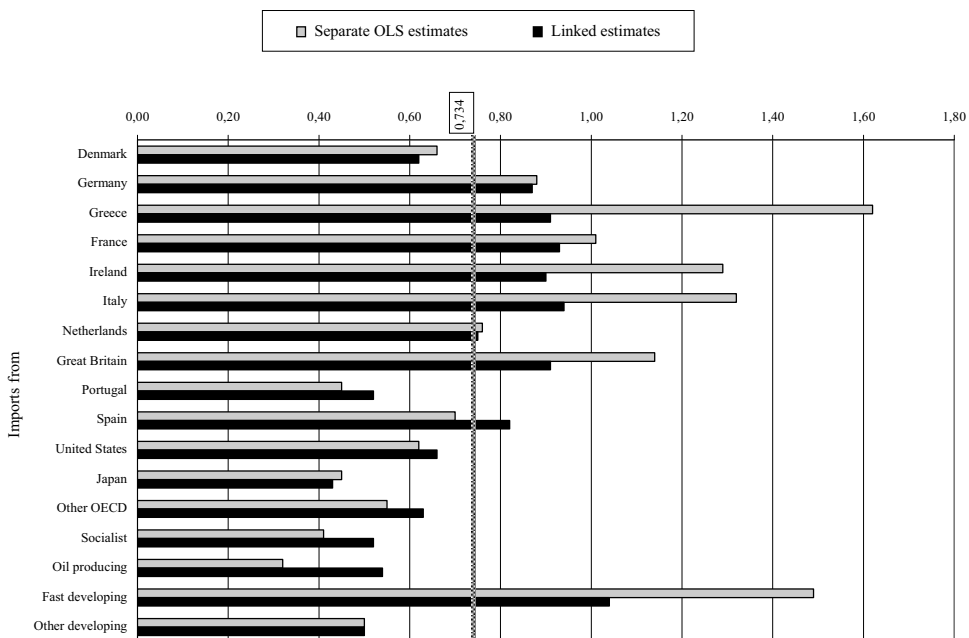


FIGURE 2.2 : Substitution elasticities for Belgian imports, per trade partner

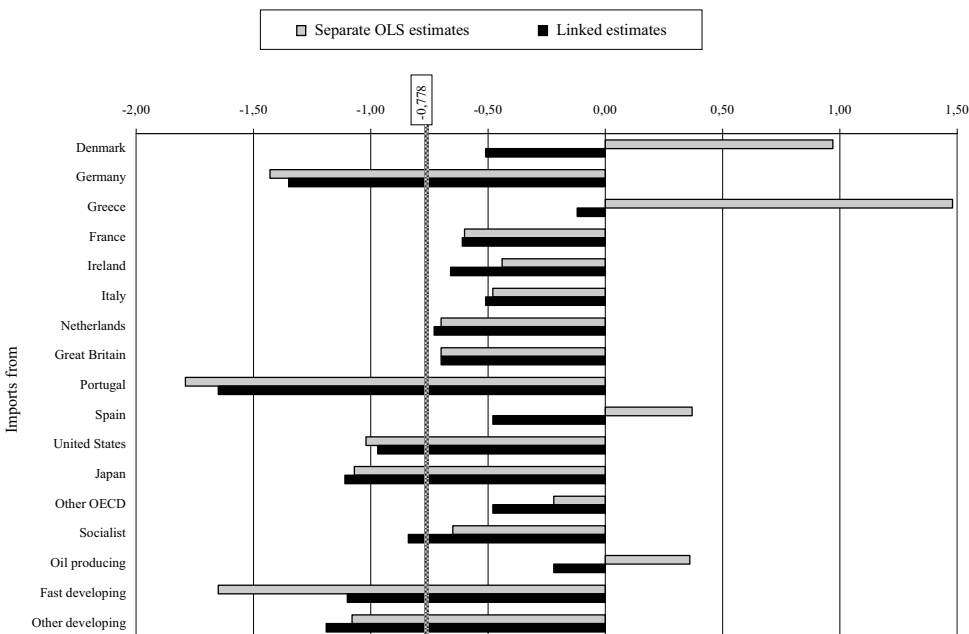
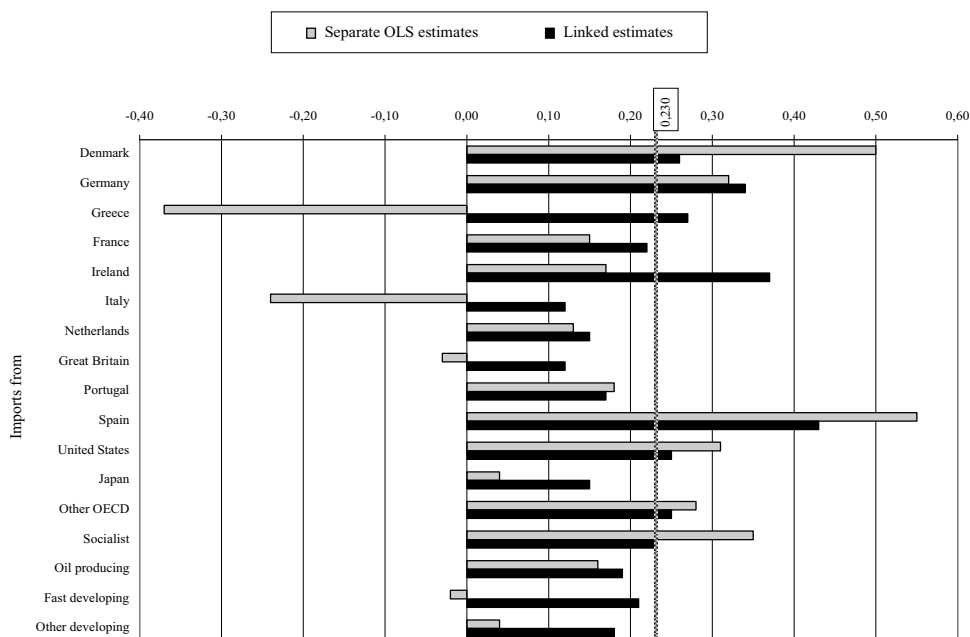


FIGURE 2.3 : Lag coefficients for Belgian imports, per trade partner



At the end of previous section, a chi-square test of the parameter distribution hypothesis was proposed. The statistic (2.20) was made ‘feasible’ in the same way as the coefficient estimators, so that the test belongs to the asymptotic variety. The calculated statistics have been reported earlier in Column (b) of Table 2.1. The outcomes do not give a hint of evidence against the hypothesis, with values well within the *first* (rather than the 99th) percentile of the null distribution (chi-square with 64 degrees of freedom; first percentile 40.6). If anything, this indicates the cross-country parameter distribution has been made too widespread by too cautious an estimate of Φ . The stochastic restriction imposed on the coefficients is too loose rather than too strict.

Finally, prediction tests were run for the period 1975-1980. They revealed structural breaks, possibly due to misspecification. Nevertheless, linkage reduced Root Mean Square (RMS) prediction errors by roughly a quarter. Table 2.5 reports RMS residuals and RMS prediction errors, both combined over the 17 partner countries. The large differences between RMS residuals and prediction errors reflect the lack of structural stability. The relatively slight increase in the RMS residual brought about by linkage is due to the rather weak stochastic restriction imposed on the coefficients. More spectacular is the decrease in the RMS prediction error, by almost a quarter, achieved by the linkage. This suggests an overall efficiency gain in both estimation and prediction.

Table 2.5: RMS residuals and RMS prediction errors for model (2.1)

Importing country		RMS estimation residual (1964-1974)	RMS prediction error (1975-1980)
Belgium	Separate OLS	.1041	.3263
	Linked estimation	.1130	.2468
The Netherlands	Separate OLS	.1112	.4157
	Linked estimation	.1189	.3177
Denmark	Separate OLS	.1039	.3328
	Linked estimation	.1097	.2626

Note : The entries are combined statistics for the 17 trade partners.

Because there is evidence of structural instability these outcomes should not be taken at face value. There are some indications of dynamic misspecification that call for further investigation. It was already noted in Section 2.3 that the dependent variable and one regressor, both measuring trade flows in levels, are to some extent trending. Furthermore it appears that, if we use the diagonal elements of the matrix Φ as controls and make the cross-country parameter distribution tighter, the RMS prediction error decreases further as the RMS residual increases. In the limit, assuming $\Phi = \mathbf{0}$ and running a single OLS regression on the pooled data, the RMS residual and prediction error attain comparable values, while the regression practically reduces to a random walk. This suggests, with equal strength for all three importer countries, that the 17 bilateral imports series may have one particularity in common: a unit root; and that the dynamic specification should be reconsidered accordingly. That endeavour, I regret to say, is beyond the scope of the present contribution.

2.7 Concluding Remarks

We have presented in this chapter a random coefficients framework for multi-unit data sets that can be extended in many different ways. Some have been alluded to along the way; a few more are mentioned now. In several cases the added complication is mainly notational. In other cases the extensions are substantive and will be the subject of further research.

The grand disturbance covariance matrix Σ may have more complex structures than considered here, allowing, e.g., for SUR with unequal sample sizes. The number of regressors may differ across units; e.g., there may be country-specific dummy variables. The parameter distribution hypothesis may concern only part of the coefficient vector; e.g., coefficients of country-specific dummy variables must obviously not be linked. The matrix Φ may have less than full rank. This will occur if exact restrictions on β are implemented,

in addition to the stochastic ones. Both types of restrictions may even connect coefficients within units as well as across units. The common ‘hypermean’ μ is easily replaced by a linear structure involving relevant unit characteristics. This creates the possibility to specify and estimate cross-sectional effects on the level of the coefficients, and so can be said to generate a multi-level model. Other challenges are stochastic regressors (like lagged dependent variables), simultaneous equations, non-linear equations and non-linear restrictions, etc.

An important domain for refinement that has already been mentioned is that of dynamics. In a model like (2.1), a careful distinction must be made between short-run and long-run effects. Above we have only discussed the estimated short-run elasticities. In a more thorough dynamic model it would be important to examine the long-run implications, and also to ponder different adjustment processes for allocation and price effects. It may even be more natural to specify the parameter generation mechanism in terms of long-run rather than short-run elasticities. Countries may be expected to be more similar in their long-run reactions. Furthermore, to treat such aspects properly, it will first of all be necessary to analyse and take due account of the time-series properties of the variables involved.

Having made these qualifications, and acknowledging that many more can be made, I hope to be forgiven for treating the empirical application as illustrative rather than substantive. Professor Barten has contributed to setting high standards for the econometrician’s profession, and his disciples may at times feel like knights of the Round Table in pursuit of some elusive Holy Grail. Yet the master himself possesses a sound sense of expediency too, and he would be sorry not to have passed some of it on.

References

- Balestra, P. and M. Nerlove (1966): “Pooling Cross-Section and Time-Series Data in the Estimation of a Dynamic Model: The Demand for Natural Gas”, *Econometrica* 34, 585-612.
- Baltagi, B.H. (1995): *Econometric Analysis of Panel Data*, (Wiley, Chichester).
- Barten, A.P. (1981): *Methodological Aspects of Macroeconomic Model Construction*, (Cabay, Louvain-la-Neuve); 2nd edition: 1987, (Academia, Louvain-la-Neuve).
- Barten, A.P. (1989): *Econometrische Lessen*, (Academic Service, Schoonhoven).
- Barten, A.P. and G. d’Alcantara (1976): “The linkage of models of the EEC countries”, in Glejser, H., ed., *Quantitative Studies of International Economic Relations*, (North-Holland, Amsterdam), 25-71.
- Barten, A.P. and G. d’Alcantara (1977): “Models of Bilateral Trade Flows”, in: Albach, H., E. Helmstädter and R. Henn, eds., *Quantitative Wirtschaftsforschung: Wilhelm Krelle zum 60. Geburtstag*, (Mohr, Tübingen), 43-57.
- Barten, A.P., G. d’Alcantara and G.J. Carrin (1976): “COMET: A medium-term macroeconomic model for the European Economic Community”, *European Economic Review* 7, 63-115.
- Blattberg, R.C. and E.I. George (1991): “Shrinkage Estimation of Price and Promotional Elasticities: Seemingly Unrelated Equations”, *Journal of the American Statistical Association* 86, 304-315.
- Breslow, N. (1990): “Biostatistics and Bayes” (with discussion), *Statistical Science* 5, 269-298.

- Bryk, A.S. and S.W. Raudenbush (1992): *Hierarchical Linear Models: Applications and Data Analysis Methods*, (Sage Publications, Newbury Park).
- d'Alcantara, G. and A.P. Barten (1984): "Factor demand explanation in the COMET model", *Economic Modelling* 1, 233-251.
- Efron, B. and C. Morris (1973): "Combining Possibly Related Estimation Problems" (with discussion), *Journal of the Royal Statistical Society, Series B*, 35, 379-421.
- Goldberger, A.S. (1962): "Best Linear Unbiased Prediction in the Generalized Linear Regression Model", *Journal of the American Statistical Association* 57, 369-375.
- Harville, D.A. (1976): "Extension of the Gauss-Markov Theorem to include the estimation of Random Effects", *The Annals of Statistics* 4, 384-395.
- Henderson, C.R. (1975): "Best Linear Unbiased Estimation and Prediction under a Selection Model", *Biometrics* 31, 423-447.
- Hoch, I. (1962): "Estimation of production function parameters combining time-series and cross-section data", *Econometrica* 30, 34-53.
- Italianer, A. (1986): *Theory and Practice of Trade Linkage Models*, (Kluwer, Dordrecht).
- James, W. and C. Stein (1961): "Estimation with Quadratic Loss", in: *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Vol. I*, (University of California Press, Berkeley), 361-379.
- Jorion, Ph. (1986): "Bayes-Stein estimation for portfolio analysis", *Journal of Financial and Quantitative Analysis* 21, 279-292.
- Kelejian, H.H. and S.W. Stephan (1983): *Efficient Prediction in Random Coefficient Panel Data Models: Theoretical and Monte Carlo Results Concerning Alternatives to Pooling and Seemingly Unrelated Regression*, Working Paper 1983-12, University of Maryland.
- Landsman, W.R. and A. Damodaran (1989): "Using Shrinkage Estimators to Improve upon Time-Series Model Proxies for the Security Market's Expectation of Earnings", *Journal of Accounting Research* 27, 97-115.
- Lenk, P.J. and A.G. Rao (1990): "New Models from Old: Forecasting product adoption by Hierarchical Bayes procedures", *Marketing Science* 9, 42-53.
- Lindley, D.V. and A.F.M. Smith (1972): "Bayes estimates for the linear model" (with discussion), *Journal of the Royal Statistical Society, Series B*, 34, 1-41.
- Maddala, G.S., ed. (1993): *The Econometrics of Panel Data, Vol. I and II*, (Edward Elgar, Cheltenham).
- Mátyás, L. and P. Sevestre, eds. (1992): *The Econometrics of Panel Data: Handbook of theory and applications*, (Kluwer, Dordrecht).
- Morris, C.N. (1983): "Parametric Empirical Bayes Inference: Theory and Applications", (with discussion), *Journal of the American Statistical Association* 78, 47-65.
- Mundlak, Y. (1961): "Empirical production function free of management bias", *Journal of Farm Economics* 43, 44-56.
- Rao, C.R. (1975): "Simultaneous Estimation of Parameters in Different Linear Models and Applications to Biometric Problems", *Biometrics* 31, 545-554.
- Robinson, G.K. (1991): "That BLUP is a good thing: The Estimation of Random Effects" (with discussion), *Statistical Science* 6, 15-51.
- Rosenberg, B. (1973): "Linear Regression with Randomly Dispersed Parameters", *Biometrika* 60, 65-72.

- Swamy, P.A.V.B. (1970): "Efficient Inference in a Random Coefficient Regression Model", *Econometrica* 38, 311-323.
- Swamy, P.A.V.B. (1971): *Statistical Inference in Random Coefficient Regression Models*, (Springer, Berlin).
- Vinod, H.D. and A. Ullah (1981): *Recent Advances in Regression Methods*, (Dekker, New York).
- Zellner, A. (1962): "An efficient method of estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias", *Journal of the American Statistical Association* 57, 348-368.
- Zellner, A. and W. Vandaale (1974): "Bayes-Stein estimators for k-means, regression and simultaneous equation models", in: Fienberg, S.E. and A. Zellner, eds., *Studies in Bayesian Econometrics and Statistics*, (North-Holland, Amsterdam), 627-653.

Chapter 3

Constrained Estimation of Matrix Variance Components¹

Abstract

Negative variance estimates are a familiar problem in random coefficient models, among others. For scalar variances, it is simple to replace a negative estimate with zero. This chapter deals with the estimation of the covariance matrix of a random coefficient vector, subject to the restriction that the estimator must be positive semidefinite. Unbiasedness is imposed only insofar as it is compatible with the restriction. The estimator solving this dilemma turns out to be at least as close to the true parameter (in an appropriate metric) as the usual unbiased estimator. It can be found either by replacing negative roots with zero in a generalized eigenvalue-eigenvector decomposition of the unbiased estimator, or by a simple iterative procedure. We present some experimental evidence on its performance.

3.1 Introduction

In the field of statistical models with random effects an important development has been the Random Coefficient Regression (or RCR) model of Swamy (1970, 1971). An awkward parameter to estimate in this model has always been the variance-covariance matrix of the random coefficients. As is well-known, familiar estimation procedures frequently yield estimates which are not positive semidefinite. This is the case with the unbiased estimator proposed by Swamy and also with the (unrestricted) maximum likelihood estimator. Furthermore, it can be shown that the likelihood function may be plurimodal for finite samples. In this chapter, an alternative estimator is proposed to circumvent these problems. The estimator is explicitly constrained to be positive semidefinite, and unbiasedness is imposed only insofar as it is compatible with that constraint. The estimator is therefore dubbed a ‘quasi-unbiased’ estimator. Some of its properties are established analytically. In particular, it is found to dominate Swamy’s unbiased estimator according to a suitable squared error criterion. To investigate its sampling distribution further, a simulation experiment is conducted, where Swamy’s unbiased estimator is used as a benchmark.

¹ Coauthored with Geert Dhaene, Katholieke Universiteit Leuven, Belgium; released as Research Memorandum RM 91-025, School of Business and Economics, Maastricht University (then Faculty of Economics, Universiteit Limburg), under the title *Constrained Estimation of the Covariance Matrix of Random Coefficients*; presented at ESEM 1991 in Cambridge.

The problem of negative variance estimates is also a familiar one in error components and latent variables models. For scalar variances, a straightforward way out is to use standard estimators and to replace negative estimates with zero; see, *e.g.*, Baltagi (1981, p. 41) or Searle (1971, p. 23) and the references therein. The unbiasedness property (assuming it was present) is then lost but this is not a serious objection, since the restricted estimator is either identical to the unbiased estimator (when it is positive) or closer to the true value (when the unbiased estimator is negative), and therefore has lower mean squared error. The quasi-unbiased estimator we propose in this chapter can be seen as a generalization of this approach for the non-scalar case.

The chapter is organized as follows. Section 2 describes the model and briefly discusses the estimation problem. Section 3 presents the quasi-unbiased estimator and Section 4 describes the Monte Carlo study. Some concluding remarks are made in Section 5.

3.2 The model

We start off from Swamy's (1970, 1971) Random Coefficient Regression model. Consider a set of G (≥ 2) independent normal regressions :

$$y_g \mid \beta_g \sim i\mathcal{N}_T(X_g\beta_g, \sigma_g^2 I), \quad g = 1, \dots, G, \quad (3.1)$$

where y_g and β_g are a $T \times 1$ and a $K \times 1$ vector, respectively, the $T \times K$ matrix of regressors X_g is non-stochastic and has full column rank, and $\sigma_g^2 > 0$, for each 'group' of observations g . The set of regressions is in some sense 'homogeneous', which is expressed by the formal assumption that

$$\beta_g \sim i\mathcal{N}_K(\mu, \Phi), \quad g = 1, \dots, G. \quad (3.2)$$

In words, the coefficient vectors β_1, \dots, β_G are assimilated to random drawings from a common normal distribution. Swamy originally assumed its covariance matrix Φ to be non-singular, but we relax this assumption. It follows from (3.1) and (3.2) that, unconditionally with respect to β_g ,

$$y_g \sim i\mathcal{N}_T(X_g\mu, X_g\Phi X_g' + \sigma_g^2 I), \quad g = 1, \dots, G. \quad (3.3)$$

The parameters β_g , σ_g^2 , μ and Φ are unknown and have ultimately to be estimated. Swamy (1970, 1971) focussed on the parameters appearing in the 'marginal' model (3.3).

Defining the ordinary least squares estimator of β_g ,

$$b_g = (X_g'X_g)^{-1}X_g'y_g, \quad (3.4)$$

the following unbiased estimators of μ and σ_g^2 are straightforward :

$$\bar{b} = \sum_g b_g/G, \quad (3.5)$$

$$s_g^2 = y_g' [I - X_g(X_g'X_g)^{-1}X_g'] y_g/(T - K). \quad (3.6)$$

Given the structure of the model, there exist more efficient, though more complicated, estimators for β_g , μ and σ_g^2 than those defined above. They typically depend on Φ , or its estimate. We will not discuss them here, since we concentrate on the estimation of Φ . It may be expected, however, that the more efficient estimators would benefit from

an improved estimation of Φ , although we note that the conclusions of Taylor (1980) and Baltagi (1981) concerning error components models are not encouraging in this respect.

We define

$$W_g = s_g^2(X_g'X_g)^{-1}, \quad (3.7)$$

$$\bar{W} = \sum_g W_g/G, \quad (3.8)$$

$$S_b = \sum_g (b_g - \bar{b})(b_g - \bar{b})'/(G-1). \quad (3.9)$$

Swamy (1970) shows that

$$\hat{\Phi}_U \equiv S_b - \bar{W} \quad (3.10)$$

is an unbiased estimator of Φ . The problem with (3.10), however, is that it can easily generate estimates which are not positive semidefinite. Although such an outcome may cast doubt on the underlying assumptions of the model, it cannot be excluded on theoretical grounds, even if the model is correctly specified. Moreover, the probability of such an outcome increases when Φ tends to be singular, and, loosely speaking, also when the order of Φ increases. (Actually, the probability is 1 when $G < K$.)

A few pragmatic ways to deal with this problem have been suggested in the literature; see, *e.g.*, the discussion in Swamy (1971, pp. 107 ff. and pp. 147 ff.). Judge *et al.* (1985, p. 542) plead for the simplest solution, which is to use S_b as an estimator of Φ in spite of the systematic overestimation this implies. Alternatively, U.L.G. Rao (1982, p. 400) and, in a related model, Havenner & Swamy (1981, p. 185) generate a positive definite estimator by augmenting $\hat{\Phi}_U$ with a scalar matrix, say γI . To determine the scalar γ , the eigenvalues of $\hat{\Phi}_U$ are computed. If all eigenvalues are positive γ is set at zero, but otherwise γ is chosen slightly larger than minus the smallest eigenvalue. This modification may be useful but is more radical than necessary since all eigenvalues are equally augmented. In this vein, one could think of augmenting only the negative (or, if desired, the non-positive) eigenvalues of $\hat{\Phi}_U$. As it turns out, the estimation principle proposed in the next section will lead to a refinement of this idea.

3.3 Quasi-unbiased estimation of Φ

In view of the relationship $E(S_b) = \Phi + E(\bar{W})$, the bias of S_b as an estimator of Φ is estimated unbiasedly by \bar{W} . Hence in (3.10), \bar{W} was subtracted from S_b . Rather than subtracting the estimated bias from S_b , one could alternatively think of premultiplying S_b by a matrix factor so as to shrink it and perhaps reduce its bias, while at the same time preserving positive-semidefiniteness. A theoretical candidate for a corrective factor would be $\Phi(\Phi + \bar{W})^{-1}$, since $\Phi(\Phi + E(\bar{W}))^{-1}S_b$ is unbiased for Φ , and S_b and \bar{W} are independent sample statistics. But, of course, Φ itself is unknown. This motivates our interest in solutions of

$$F = F(F + \bar{W})^{-1}S_b \quad (3.11)$$

for $F \in \mathcal{C}^K$, where \mathcal{C}^K represents the class (convex cone) of positive semidefinite symmetric matrices of order K . Solving (3.11) provides a 'consistent plan' for estimating Φ , in the sense that it requires the estimator to be a fixed point of a matrix function designed to estimate Φ . Note that, provided S_b is non-singular, $\hat{\Phi}_U$ is a solution of (3.11), though not necessarily in \mathcal{C}^K . Since, *e.g.*, also the null matrix is a solution of (3.11), we clearly have to restrict the class of solutions further. Actually, we propose to estimate Φ by

$$\hat{\Phi}_{QU} \equiv \arg \max_{F \in \mathcal{C}^K} \text{rank}(F) \quad \text{subject to} \quad F = F(F + \bar{W})^{-1}S_b, \quad (3.12)$$

where $\text{rank}(F)$ denotes the rank of F . We establish below that $\hat{\Phi}_{QU}$ is well-defined, *i.e.*, the solution to (3.12) exists and is unique, and present it in explicit form. Then we find that it has some attractive interpretations and properties, particularly in relation to the unbiased estimator $\hat{\Phi}_U$. Clearly $\hat{\Phi}_{QU}$ is not unbiased, yet it results from an attempt at correcting the bias in S_b insofar as this is compatible with non-negativity. The expression ‘Quasi-Unbiased’ and the abbreviated identifier ‘QU’ refer to this feature. They are *not* meant to suggest that the bias of $\hat{\Phi}_{QU}$ is necessarily small.

The following asymptotic result is weak, but immediate. Let $G \rightarrow \infty$ and $T \rightarrow \infty$ simultaneously in some way. Then, under standard conditions, $S_b \xrightarrow{a.s.} \Phi$ and, since $\overline{W} \xrightarrow{a.s.} 0$, also $\hat{\Phi}_U \xrightarrow{a.s.} \Phi$, while in the limit (3.11) has the one full-rank solution $F = \Phi$ almost surely. In this sense all three estimators S_b , $\hat{\Phi}_U$ and $\hat{\Phi}_{QU}$ are consistent for Φ .

We have organized the main theoretical results concerning (3.11)–(3.12) into a few lemma’s and theorems. Most will be based on the simultaneous diagonalization of two real symmetric matrices; see, *e.g.*, Magnus & Neudecker (1988, p. 21), C.R. Rao (1971, p. 41) or Strang (1988, p. 343). For future reference, we introduce it as follows. Given that \overline{W} is positive definite symmetric and $S_b - \overline{W}$ is symmetric, there exist a non-singular matrix P and a diagonal matrix Λ satisfying simultaneously

$$\overline{W} = PP' \quad \text{and} \quad S_b - \overline{W} = P\Lambda P'. \quad (3.13)$$

The diagonal elements of Λ , say $\lambda_1, \dots, \lambda_K$, are the roots of

$$|S_b - \overline{W} - \lambda\overline{W}| = 0 \quad (3.14)$$

and are sometimes called generalized eigenvalues of $S_b - \overline{W}$; the columns of P , say p_1, \dots, p_K , are the associated generalized eigenvectors. A transcription of (3.13) is

$$\begin{aligned} \overline{W} &= p_1p_1' + \dots + p_Kp_K', \\ S_b - \overline{W} &= \lambda_1p_1p_1' + \dots + \lambda_Kp_Kp_K'. \end{aligned} \quad (3.15)$$

If all the roots of (3.14) are distinct, then P and Λ are unique up to matching permutations of p_1, \dots, p_K and $\lambda_1, \dots, \lambda_K$. Otherwise, the columns of P associated with a multiple root can be varied in linear combinations so long as the sum of their outer products remains unchanged. The lemma which follows shows that any symmetric solution of (3.11) can be diagonalized simultaneously with \overline{W} and S_b .

Lemma 1. A necessary and sufficient condition for $\tilde{\Phi}$ to be a symmetric solution of (3.11) is that there exist a non-singular matrix P and diagonal matrices $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_K)$ and $\tilde{\Lambda} = \text{diag}(\tilde{\lambda}_1, \dots, \tilde{\lambda}_K)$, such that (3.13) holds and at the same time

$$\tilde{\Phi} = P\tilde{\Lambda}P', \quad (3.16)$$

with, for $k = 1, \dots, K$,

$$\tilde{\lambda}_k = \begin{cases} 0 & \text{if } \lambda_k \in \{0, -1\} \\ 0 \text{ or } \lambda_k & \text{otherwise.} \end{cases} \quad (3.17)$$

Proof. Substitution of (3.13) and (3.16) into (3.11) and pre- and postmultiplication by P^{-1} and $(P')^{-1}$, respectively, yield

$$\tilde{\Lambda} = \tilde{\Lambda}(\tilde{\Lambda} + I)^{-1}(\Lambda + I), \quad (3.18)$$

which holds indeed if (3.17) holds. This proves sufficiency. To prove necessity, let Q be a non-singular and Γ a diagonal matrix satisfying $\bar{W} = QQ'$ and $S_b - \bar{W} = Q\Gamma Q'$. If $\tilde{\Phi}$ is a symmetric solution of (3.11), then

$$\tilde{\Phi}_* = \tilde{\Phi}_*(\tilde{\Phi}_* + I)^{-1}(\Gamma + I), \quad (3.19)$$

where $\tilde{\Phi}_* = Q^{-1}\tilde{\Phi}Q'^{-1}$. Hence, since $\tilde{\Phi}_*$ and $(\tilde{\Phi}_* + I)^{-1}$ commute,

$$\tilde{\Phi}_*(\Gamma + I) = \tilde{\Phi}'_*(\Gamma + I) = (\Gamma + I)\tilde{\Phi}_*(\tilde{\Phi}_* + I)^{-1}(\Gamma + I) = (\Gamma + I)\tilde{\Phi}_*, \quad (3.20)$$

i.e., $\tilde{\Phi}_*$ and $(\Gamma + I)$ commute. Therefore $\tilde{\Phi}_*$ and $(\Gamma + I)$ have the same eigenvectors, and there exists an orthogonal matrix R such that $R\tilde{\Phi}_*R'$ and $R(\Gamma + I)R'$ are both diagonal matrices, say $\tilde{\Lambda}$ and $(\Lambda + I)$, respectively. It follows that (3.19) implies (3.18), which in its turn implies (3.17). Also, with $P = QR'$, the factorizations (3.13) and (3.16) are implied. This proves necessity. (Q.E.D.)

It is useful to note that (3.13) also implies the decomposition $S_b = P(\Lambda + I)P'$ and hence, since $S_b \in \mathcal{C}^K$, that $\lambda_k \geq -1$, all k . If any $\lambda_k = -1$ then S_b is singular, which will be the case if $G < K$. If any $\lambda_k = 0$ then $S_b - \bar{W}$ is singular, which will happen with zero probability.

The second lemma establishes under what condition the number of symmetric solutions to (3.11) is finite or infinite.

Lemma 2. A necessary and sufficient condition for the number of different symmetric solutions to (3.11) to be finite, is that those roots of (3.14) which are neither 0 nor -1 are all distinct.

Proof. If the roots of (3.14) which are neither 0 nor -1 are all distinct, the columns of P associated with those roots are uniquely determined. The number of symmetric solutions to (3.11) is then equal to the number of different choices for $\tilde{\Lambda}$ in (3.17). This proves sufficiency. If the condition is not met at least two roots of (3.14) which are neither 0 nor -1 coincide, or $\lambda_l = \lambda_m \notin \{0, -1\}$ for some $l \neq m$. The corresponding column vectors p_l and p_m are not unique and can be varied in an infinite number of ways. If in (3.17) we let $\tilde{\lambda}_l = 0$ and $\tilde{\lambda}_k = \lambda_k$ for all $k \neq l$, then by varying (p_l) and p_m a continuum of different symmetric solutions to (3.11) can be generated, which proves necessity. (Q.E.D.)

It is worthwhile to note that the condition of Lemma 2 is satisfied with probability 1, so that in practice we need not worry about the possibility of a continuum of solutions $\tilde{\Phi}$. Nevertheless, even a finite number of different solutions may be too many. Not all solutions, however, will in general be positive semidefinite. In fact, by Sylvester's law of inertia, the restriction $\tilde{\Phi} \in \mathcal{C}^K$ will be satisfied if and only if $\tilde{\lambda}_k \geq 0$, $k = 1, \dots, K$. The following theorem derives an explicit expression for the estimator defined in (3.12) and establishes its uniqueness. For conciseness we define an operator, denoted by a subscript $+$ sign, to replace any negative elements of a diagonal matrix with zeroes, as in $\Lambda_+ = \text{diag}(\max(0, \lambda_1), \dots, \max(0, \lambda_K))$.

Theorem 1. Let the non-singular matrix P and the diagonal matrix $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_K)$ satisfy (3.13), or equivalently (3.15). Then, the solution to the problem

$$\max_{F \in \mathcal{C}^K} \text{rank}(F) \quad \text{s.t.} \quad F = F(F + \bar{W})^{-1}S_b \quad (3.21)$$

is unique and is given by

$$\hat{\Phi}_{QU} = P\Lambda_+P'. \quad (3.22)$$

Proof. From Lemma 1 it follows that every positive semidefinite solution of $F = F(F + \bar{W})^{-1}S_b$ takes the form $P\tilde{\Lambda}P'$ for some (P, Λ) with the diagonal elements of $\tilde{\Lambda}$ now restricted as follows :

$$\tilde{\lambda}_k = \begin{cases} 0 & \text{if } \lambda_k \leq 0 \\ 0 \text{ or } \lambda_k & \text{if } \lambda_k > 0. \end{cases} \quad (3.23)$$

Clearly under this restriction the rank of $P\tilde{\Lambda}P'$ is maximal for $\tilde{\Lambda} = \Lambda_+$, given (P, Λ) . The uniqueness of $P\Lambda_+P'$ can be verified as follows. Define the index sets $\mathcal{S}_k = \{j : \lambda_j = \lambda_k\}$, $k = 1, \dots, K$. Then, collecting terms with equal values of λ_k , the second line of (3.15) can be rewritten as

$$S_b - \bar{W} = \sum_{k \in \mathcal{D}} \lambda_k \left(\sum_{j \in \mathcal{S}_k} p_j p_j' \right) \quad (3.24)$$

for some index set \mathcal{D} . Note that, although the p_j need not be unique in the presence of repeated roots λ_k , each term in the outer sum in (3.24) is unique. Since $P\Lambda_+P'$ can be written as

$$P\Lambda_+P' = \sum_{k \in \mathcal{D}'} \lambda_k \left(\sum_{j \in \mathcal{S}_k} p_j p_j' \right) \quad (3.25)$$

for some index set $\mathcal{D}' \subseteq \mathcal{D}$, the uniqueness of $P\Lambda_+P'$ follows. (Q.E.D.)

Clearly, $\hat{\Phi}_{QU} = \hat{\Phi}_U = P\Lambda P'$ if and only if $\hat{\Phi}_U \in \mathcal{C}^K$. The following lemma, stated in slightly more general terms than strictly necessary, provides a rather unexpected interpretation of $\hat{\Phi}_{QU}$.

Lemma 3. Let A and B be given $K \times K$ symmetric matrices, with A positive definite. Let Ω be a $K \times K$ diagonal matrix with on its diagonal the roots of

$$|B - \omega A| = 0 \quad (3.26)$$

in some order, and let H be an associated non-singular matrix satisfying both

$$A = HH' \quad \text{and} \quad B = H\Omega H'. \quad (3.27)$$

Consider the real-valued function $f(\cdot)$ defined by

$$f(F) = \text{tr}(F - B)'^{-1}(F - B)A^{-1}, \quad F \in \mathcal{C}^K. \quad (3.28)$$

Then $f(\cdot)$ has a unique minimum at the point

$$F = H\Omega_+H'. \quad (3.29)$$

Proof. Write F as HZH' , with $Z \in \mathcal{C}^K$. Then, given H , $f(F) = \text{tr}(Z - \Omega)^2$. This attains a unique minimum, restricted by $Z \in \mathcal{C}^K$, when $Z = \Omega_+$. The proof of the uniqueness of $H\Omega_+H'$, with H and Ω satisfying (3.27), is the same as that for $P\Lambda_+P'$ in Theorem 1. (Q.E.D.)

The function $f(F)$ in the lemma can be viewed as the square of the following function measuring the distance between F and B :

$$d_A(F, B) = [\text{diag}(F - B)'^{-1}(F - B)A^{-1}]^{\frac{1}{2}}. \quad (3.30)$$

This provides the quasi-unbiased estimator with a minimum distance interpretation : Setting $A = \overline{W}$ and $B = S_b - \overline{W} = \hat{\Phi}_U$, it turns out that $\hat{\Phi}_{QU} = P\Lambda_+P'$ is the (unique) point in \mathcal{C}^K nearest to $\hat{\Phi}_U$ in terms of $d_{\overline{W}}$, i.e.,

$$d_{\overline{W}}(\hat{\Phi}_{QU}, \hat{\Phi}_U) = \min_{\hat{\Phi} \in \mathcal{C}^K} d_{\overline{W}}(\hat{\Phi}, \hat{\Phi}_U). \quad (3.31)$$

One noteworthy property of $d_{\overline{W}}$ is its invariance under non-singular linear transformations $X_g \rightarrow X_g C^{-1}$ and $\beta_g \rightarrow C\beta_g$, $g = 1, \dots, G$ (implying $\overline{W} \rightarrow C\overline{W}C'$, $S_b \rightarrow CS_bC'$ and $\Phi \rightarrow C\Phi C'$).

Not surprisingly, $d_{\overline{W}}$ is not the only invariant distance function under which $\hat{\Phi}_{QU}$ is the point in \mathcal{C}^K nearest to $\hat{\Phi}_U$. The following theorem characterizes the class of invariant distance functions d_A which exhibit this property.

Theorem 2. Let the non-singular matrix P and the diagonal matrix Λ satisfy (3.13) and let d_A be defined as in (3.30) for some positive definite symmetric matrix A . Then, $P\Lambda_+P'$ is the unique solution to the problem

$$\min_{\hat{\Phi} \in \mathcal{C}^K} d_A(\hat{\Phi}, \hat{\Phi}_U) \quad (3.32)$$

for each possible realization of \overline{W} (positive definite) and S_b if and only if A can be written as $A = P\Delta P'$ where Δ is a positive definite diagonal matrix.

Proof. Let, for a given positive definite symmetric matrix A , the non-singular matrix D and the diagonal matrix Γ satisfy the equations

$$A = PDD'P' \quad \text{and} \quad S_b - \overline{W} = PD\Gamma D'P'. \quad (3.33)$$

Since P is non-singular, it follows that $D\Gamma D' = \Lambda$, and since by Lemma 3 $PD\Gamma_+D'P'$ minimizes $d_A(\cdot, \hat{\Phi}_U)$, $P\Lambda_+P'$ will be the unique solution to (3.32) if and only if

$$D\Gamma_+D' = (D\Gamma D')_+. \quad (3.34)$$

Clearly, (3.34) will hold for any realization of Γ if and only if D can be chosen to be diagonal, or equivalently, A is of the form $A = P\Delta P'$ with Δ diagonal and positive definite. (Q.E.D.)

Note that \overline{W} , S_b (if non-singular), and any positive definite linear combination of \overline{W} and S_b are possible choices for A , each giving $\hat{\Phi}_{QU}$ as the unique solution to (3.32).

This minimizing property of $\hat{\Phi}_{QU}$, apart from allowing one to view it as the projection of $\hat{\Phi}_U$ onto \mathcal{C}^K , has another interesting consequence : $\hat{\Phi}_{QU}$ is in terms of d_A at least as close to Φ as $\hat{\Phi}_U$. The latter fact is the object of the following theorem. Intuitively, in the space of $K \times K$ matrices the vector $\hat{\Phi}_U - \hat{\Phi}_{QU}$ is perpendicular to the set \mathcal{C}^K . Because \mathcal{C}^K is convex, this perpendicular (besides being unique) makes an obtuse angle with any vector $\Phi - \hat{\Phi}_{QU}$ with $\Phi \in \mathcal{C}^K$. Hence $\hat{\Phi}_U - \Phi$ is the longest side (facing the obtuse angle) in the triangle $\hat{\Phi}_U, \hat{\Phi}_{QU}, \Phi$.

Theorem 3. Let the non-singular matrix P and the diagonal matrix Λ satisfy (3.13). Then, for any matrix A of the form $P\Delta P'$, with Δ diagonal and positive definite, and for any $\Phi \in \mathcal{C}^K$,

$$d_A(\hat{\Phi}_{QU}, \Phi) \leq d_A(\hat{\Phi}_U, \Phi), \quad (3.35)$$

with equality if and only if $\hat{\Phi}_{QU} = \hat{\Phi}_U$.

Proof. For any given positive definite symmetric matrix A , an inner product between $K \times K$ matrices X and Y can be defined as

$$(X, Y)_A = \text{tr } X'^{-1} Y A^{-1}. \quad (3.36)$$

Note that, since $d_A^2(X, Y) = (X - Y, X - Y)_A$, the distance function d_A actually derives from the inner product. Furthermore, the space of real $K \times K$ matrices, equipped with the inner product $(X, Y)_A$, is a Hilbert space, and \mathcal{C}^K is a closed convex subset of it. Therefore, it follows from Theorem 2 that

$$(\hat{\Phi}_U - \hat{\Phi}_{QU}, \Phi - \hat{\Phi}_{QU})_A \leq 0 \quad (3.37)$$

for all $\Phi \in \mathcal{C}^K$; see, *e.g.*, Luenberger (1969, p. 69). Now, (3.37) is equivalent with

$$d_A^2(\hat{\Phi}_U, \hat{\Phi}_{QU}) + d_A^2(\hat{\Phi}_{QU}, \Phi) - d_A^2(\hat{\Phi}_U, \Phi) \leq 0 \quad (3.38)$$

and therefore implies (3.35), with equality if and only if $\hat{\Phi}_{QU} = \hat{\Phi}_U$. (Q.E.D.)

In the scalar case ($K = 1$), a variance estimator bounded by zero obviously cannot have a larger (mean) squared error than the corresponding unconstrained, possibly unbiased estimator taking negative values with positive probability. Theorem 3 is as close as we can get to a generalization of this property for the multivariate case ($K > 1$).

The computation of $\hat{\Phi}_{QU}$ when $\hat{\Phi}_U \notin \mathcal{C}^K$ may be done through the simultaneous diagonalization of \bar{W} and S_b defined in (3.13). Alternatively, the following theorem suggests an iterative procedure.

Theorem 4. Let Λ and P satisfy (3.13). The sequence $\{F^{(i)}\}$ defined by

$$\begin{aligned} F^{(0)} &= S_b, \\ F^{(i)} &= F^{(i-1)} [F^{(i-1)} + \bar{W}]^{-1} S_b, \quad i = 1, 2, \dots \end{aligned} \quad (3.39)$$

converges to $P\Lambda_+P'$. Moreover, $F^{(i)} \in \mathcal{C}^K$ and $F^{(i)} - F^{(i+1)} \in \mathcal{C}^K$, $i = 0, 1, \dots$ (In this sense, the sequence is monotonic.)

Proof. It is easily verified by induction that

$$F^{(i)} = P \left\{ (\Lambda + I)^{i+1} \left[\sum_{j=0}^i (\Lambda + I)^j \right]^{-1} \right\} P', \quad i = 0, 1, \dots \quad (3.40)$$

This is obviously symmetric for all i . Write the diagonal matrix between curly braces in (3.40) as $\text{diag}(\delta_1^{(i)}, \dots, \delta_K^{(i)})$. Then, for $i = 0, 1, \dots$ and $k = 1, \dots, K$,

$$\delta_k^{(i)} = \begin{cases} \frac{1}{i+1} & \text{if } \lambda_k = 0 \\ \frac{\lambda_k(1+\lambda_k)^{i+1}}{(1+\lambda_k)^{i+1}-1} & \text{otherwise.} \end{cases} \quad (3.41)$$

Since $\lambda_k \geq -1$, it follows that $\delta_k^{(i)} \geq \delta_k^{(i+1)} \geq 0$ whence $F^{(i)} \in \mathcal{C}^K$ and the monotonicity of the sequence $\{F^{(i)}\}$. Furthermore, it can be seen from (3.41) that $\delta_k^{(i)} \rightarrow \max\{0, \lambda_k\}$ as $i \rightarrow \infty$ and consequently

$$\lim_{i \rightarrow \infty} F^{(i)} = P\Lambda_+P', \quad (3.42)$$

which completes the proof. (Q.E.D.)

Defining the approximation error of $\delta_k^{(i)}$ as $\epsilon_k^{(i)} \equiv \delta_k^{(i)} - \max\{0, \lambda_k\}$, we have

$$\epsilon_k^{(i)} = \begin{cases} 0 & \text{if } \lambda_k = -1 \\ \frac{\lambda_k(1+\lambda_k)^{i+1}}{(1+\lambda_k)^{i+1}-1} & \text{if } -1 < \lambda_k < 0 \\ \frac{1}{i+1} & \text{if } \lambda_k = 0 \\ \frac{\lambda_k}{(1+\lambda_k)^{i+1}-1} & \text{if } 0 < \lambda_k. \end{cases} \quad (3.43)$$

Note the continuity of $\epsilon_k^{(i)}$ on $[-1, +\infty[$. The rate of convergence of $\delta_k^{(i)}$ is given by

$$\lim_{i \rightarrow \infty} \frac{\epsilon_k^{(i)}}{\epsilon_k^{(i-1)}} = \begin{cases} (1 + \lambda_k) & \text{if } -1 < \lambda_k < 0 \\ (1 + \lambda_k)^{-1} & \text{if } 0 \leq \lambda_k. \end{cases} \quad (3.44)$$

This shows that the order of convergence of $\{F^{(i)}\}$ is linear. Clearly, the worst case is provided by $\lambda_k = 0$ for some k .

Rewriting $\epsilon_k^{(i)}$ for $\lambda_k \neq -1$ gives

$$\epsilon_k^{(i)} = \begin{cases} (1 + \lambda_k) \left[\sum_{j=0}^i (1 + \lambda_k)^{-j} \right]^{-1} & \text{if } -1 < \lambda_k < 0 \\ \left[\sum_{j=0}^i (1 + \lambda_k)^{-j} \right]^{-1} & \text{if } 0 \leq \lambda_k. \end{cases} \quad (3.45)$$

from which the following bounds for $\epsilon_k^{(i)}$, independent of λ_k , are readily derived :

$$0 \leq \epsilon_k^{(i)} \leq \frac{1}{i+1}. \quad (3.46)$$

3.4 A Monte Carlo experiment

Apart from the unbiasedness of $\hat{\Phi}_U$, we have no analytical results on the small-sample distributions of $\hat{\Phi}_U$ and $\hat{\Phi}_{QU}$. In this section we present some experimental evidence as a substitute. The experiment was programmed and run on MS-DOS driven personal computers using the GAUSSTM language (Aptech Systems, Inc.). Also the pseudo-random number generator of the GAUSS system was used. The design of the experiment was chosen so as to mimic a situation where one would be estimating three elasticities as coefficients. So $K = 3$, while G and T were both allowed to vary within the set $\{5, 10, 25\}$. The ordinary least squares vectors b_g were generated directly from

$$b_g \sim i\mathcal{N}_K(\mu, \Phi + \sigma_g^2(X_g'X_g)^{-1}), \quad g = 1, \dots, G, \quad (3.47)$$

with the following parameter values:

$$\mu = 0,$$

$$\Phi = \begin{bmatrix} .04 & & & \\ .06 & .16 & & \\ 0 & 0 & .64 & \\ & & & \end{bmatrix},$$

$$\sigma_g^2 = 4^{(g-1)/(G-1)}, \quad g = 1, \dots, G;$$

and the following inverse product matrix, common to all g :

$$(X'_g X_g)^{-1} = T^{-1} \begin{bmatrix} 1 & & \\ -.9 & 1 & \\ -.6 & .3 & 1 \end{bmatrix}, \quad g = 1, \dots, G.$$

(Throughout this section, only the lower triangular part of symmetric matrices is printed.) The implied design product matrix $(X'_g X_g)$ is, for all g ,

$$(X'_g X_g) = T \begin{bmatrix} 14.22 & & \\ 11.25 & 10.00 & \\ 5.15 & 3.75 & 2.97 \end{bmatrix},$$

and the corresponding matrix of correlation coefficients is

$$\begin{bmatrix} 1 & & \\ .94 & 1 & \\ .79 & .69 & 1 \end{bmatrix}.$$

Note that the diagonal elements of $\sigma_g^2(X'_g X_g)^{-1}$ vary from within the range $[.2, .8]$ when $T = 5$ to within the range $[.04, .16]$ when $T = 25$, and so are comparable in magnitude to the diagonal elements of Φ . Table 3.1 shows $E(\bar{W})$ for the selected values of G and T .

Table 3.1 : $E(\bar{W})$ for selected values of G and T

	$T = 5$			$T = 10$			$T = 25$		
	.4497			.2249			.0899		
$G = 5$	-.4047	.4497		-.2024	.2249		-.0809	.0899	
	-.2698	.1349	.4497	-.1349	.0675	.2249	-.0540	.0270	.0899
	.4403			.2201			.0881		
$G = 10$	-.3963	.4403		-.1981	.2201		-.0793	.0881	
	-.2642	.1321	.4403	-.1321	.0660	.2201	-.0528	.0264	.0881
	.4356			.2178			.0871		
$G = 25$	-.3921	.4356		-.1960	.2178		-.0784	.0871	
	-.2614	.1307	.4356	-.1307	.0653	.2178	-.0523	.0261	.0871

The regression variances s_g^2 were generated using

$$s_g^2 = \sigma_g^2 \chi_{T-K}^2 / (T - K), \quad (3.48)$$

where χ_{T-K}^2 is a chi-square variate formed by summing the squares of $T - K$ independent drawings from the standard normal distribution.

All the results reported below are based on 100,000 replications, with almost no use of variance-reducing devices (control variates). We will present the main results in two parts. In the first part, we will compare the element-by-element performance of $\hat{\Phi}_U$ and $\hat{\Phi}_{QU}$ in terms of bias, mean squared error (MSE), and the probability of being closer to (the corresponding element of) Φ . In the second part, we will compare the overall performance of $\hat{\Phi}_U$, $\hat{\Phi}_{QU}$ and a number of alternative estimators of Φ , in terms of a scalar distance function of the type used in the previous section. First of all, however, Table 3.2 reports the proportion of cases (replications) in which $\hat{\Phi}_U$ was not a positive semidefinite matrix, and hence in which $\hat{\Phi}_{QU}$ differed from $\hat{\Phi}_U$. By no means can this proportion be called small or unimportant; as was to be expected it decreases as T and (especially) G increase. In this and following tables, sampling standard errors of the reported estimates are given in parentheses.

Table 3.2 : *Probability of the unbiased estimator not being positive semidefinite*

	$T = 5$	$T = 10$	$T = 25$
$G = 5$.8990 (.0010)	.9052 (.0009)	.8747 (.0010)
$G = 10$.7667 (.0015)	.7545 (.0014)	.7002 (.0014)
$G = 25$.6362 (.0015)	.6193 (.0015)	.5388 (.0016)

Table 3.3 presents the bias of $\hat{\Phi}_{QU}$. The bias is clearly non-zero and may be sizeable, especially for small values of G and (especially) T . Nevertheless, by reference to Table 3.1, the bias of $\hat{\Phi}_{QU}$ is evidently much smaller than $E(\bar{W})$, the bias of S_b as an estimator of Φ . One may notice in Table 3.3 that the bias elements decrease along the diagonal of Φ . The bias is smaller (relatively and absolutely) for larger diagonal elements of Φ . As to the non-diagonal bias elements, they are more or less proportional to the corresponding elements of $E(\bar{W})$, especially for small G and T . The connection between these patterns and the (relative) magnitudes of the elements of Φ and $E(\bar{W})$ remains to be investigated.

Table 3.3 : *Bias of $\hat{\Phi}_{QU}$*

	$T = 5$			$T = 10$			$T = 25$		
$G=5$.2323 (.0007)			.1147 (.0003)			.0461 (.0001)		
	-.2173 (.0006)	.2215 (.0007)		-.1083 (.0002)	0.1079 (0.0002)		-.0439 (.0001)	.0428 (.0001)	
	-.1324 (.0005)	.0911 (.0004)	.1447 (.0006)	-.0635 (.0002)	.0499 (.0002)	.0562 (.0002)	-.0249 (.0001)	.0218 (.0001)	.0175 (.0001)
$G=10$.1467 (.0005)			.0683 (.0002)			.0245 (.0001)		
	-.1391 (.0005)	.1374 (.0005)		-.0652 (.0002)	.0632 (.0002)		-.0235 (.0001)	.0226 (.0000)	
	-.0806 (.0003)	.0658 (.0003)	.0658 (.0003)	-.0366 (.0001)	.0330 (.0001)	.0236 (.0001)	-.0130 (.0000)	.0122 (.0000)	.0074 (.0000)
$G=25$.0814 (.0003)			.0344 (.0001)			.0107 (.0000)		
	-.0779 (.0003)	.0753 (.0002)		-.0329 (.0001)	.0317 (.0001)		-.0103 (.0000)	.0099 (.0000)	
	-.0433 (.0002)	.0398 (.0001)	.0263 (.0001)	-.0182 (.0001)	.0172 (.0001)	.0101 (.0000)	-.0057 (.0000)	.0054 (.0000)	.0031 (.0000)

To save space the mean squared errors (MSE) of the elements of $\hat{\Phi}_{QU}$ and $\hat{\Phi}_U$ are omitted, but their ratios are presented in Table 3.4. In all the cases considered and for all elements of Φ , quasi-unbiased estimation achieves a statistically significant reduction in MSE. In relative terms the reduction varies from less than 1% (for the largest element of Φ and for large T) to more than 30% (for the smallest variance in Φ and for large G). It seems clear from Tables 4 and 2 that the scope for $\hat{\Phi}_{QU}$ improving upon $\hat{\Phi}_U$ decreases as T increases. It would also seem that in most cases (but not all) large diagonal elements and small off-diagonal elements in Φ are relatively little affected by quasi-unbiased estimation, that is, by the positive semidefiniteness requirement. Beyond that, we detect no obvious relationships in Table 3.4 between the design and the relative improvement in MSE. In particular, the effect of increasing G seems ambiguous.

Table 3.4 : Mean Square Error Ratios for $\hat{\Phi}_{QU}$ and $\hat{\Phi}_U$, element by element

	$T = 5$			$T = 10$			$T = 25$		
$G=5$.8001 (.0034)			.9514 (.0029)			.9921 (.0028)		
	.7971 (.0032)	.8208 (.0032)		.9477 (.0029)	.9439 (0.0025)		.9869 (.0027)	.9832 (.0017)	
	.7777 (.0024)	.8216 (.0019)	.8597 (.0018)	.8669 (.0018)	.9042 (.0014)	.9453 (.0009)	.9347 (.0012)	.9688 (.0008)	.9860 (.0003)
$G=10$.7055 (.0031)			.8104 (.0029)			.8469 (.0027)		
	.7118 (.0030)	.7621 (.0030)		.8097 (.0028)	.8833 (.0025)		.8611 (.0024)	.9512 (.0016)	
	.7506 (.0024)	.8403 (.0019)	.8784 (.0015)	.8612 (.0017)	.9262 (.0014)	.9666 (.0006)	.9423 (.0011)	.9781 (.0007)	.9944 (.0002)
$G=25$.6282 (.0029)			.6959 (.0027)			.7743 (.0023)		
	.6259 (.0028)	.7162 (.0028)		.6898 (.0026)	.8476 (.0023)		.7864 (.0020)	.9404 (.0014)	
	.7671 (.0022)	.8677 (.0020)	.9273 (.0010)	.8837 (.0016)	.9414 (.0013)	.9827 (.0005)	.9591 (.0009)	.9829 (.0006)	.9972 (.0002)

Another way to compare the performances of $\hat{\Phi}_U$ and $\hat{\Phi}_{QU}$ is to look at the probabilities that the elements of $\hat{\Phi}_{QU}$ are closer than those of $\hat{\Phi}_U$ to the corresponding elements of Φ . Such probabilities can be defined either conditionally upon the two estimators not being identical, or unconditionally. The conditional probabilities are easier to interpret, since they give the advantage to $\hat{\Phi}_{QU}$ when they are above $\frac{1}{2}$ and to $\hat{\Phi}_U$ when they are below $\frac{1}{2}$. Table 3.5 presents these conditional probabilities, element by element, as approximated by experimental relative frequencies. The proportion of replications where the two estimators differed, given in Table 3.2, can be used to reconstruct the unconditional probabilities if desired. The conditional probabilities reported are in most cases significantly larger than $\frac{1}{2}$. The only exceptions concern element (2, 1) of Φ for $G = 5$ and $T = 10$ or 25. For this element and under this criterion an increase in T from 5 to 10 or 25 seems to benefit $\hat{\Phi}_U$ more than $\hat{\Phi}_{QU}$.

Table 3.5 : Probability that $\hat{\Phi}_{QU}$ is closer to Φ than $\hat{\Phi}_U$, element by element, given that $\hat{\Phi}_{QU} \neq \hat{\Phi}_U$

	$T = 5$		$T = 10$			$T = 25$			
	.5044 (.0017)		.5256 (.0017)			.5610 (.0017)			
$G=5$.5134 (.0017)	.5378 (.0017)	.4863 (.0017)	.5587 (0.0017)	.4353 (.0017)	.5770 (.0017)			
	.6317 (.0017)	.6523 (.0016)	.5894 (.0016)	.5934 (.0016)	.5970 (.0016)	.5972 (.0016)	.5478 (.0017)	.5451 (.0017)	.5960 (.0017)
	.5761 (.0018)		.6016 (.0018)			.6317 (.0018)			
$G=10$.5659 (.0018)	.5841 (.0018)	.5521 (.0018)	.5833 (.0018)	.5274 (.0019)	.5881 (.0019)			
	.6282 (.0017)	.6049 (.0018)	.5879 (.0018)	.5724 (.0018)	.5487 (.0018)	.5758 (.0018)	.5467 (.0019)	.5208 (.0019)	.5669 (.0019)
	.6874 (.0018)		.6927 (.0019)			.7063 (.0020)			
$G=25$.6693 (.0019)	.6394 (.0019)	.6626 (.0019)	.6156 (.0020)	.6414 (.0021)	.5994 (.0021)			
	.6062 (.0019)	.5547 (.0020)	.5820 (.0020)	.5703 (.0020)	.5365 (.0020)	.5568 (.0020)	.5519 (.0021)	.5290 (.0022)	.5461 (.0021)

Having compared the performances of $\hat{\Phi}_{QU}$ and $\hat{\Phi}_U$ on an element-by-element basis, we now turn to a comparison of ‘overall’ performance. A scalar measure of closeness or distance that can meaningfully be applied to competing estimators of Φ will make it easy to compare more than two alternatives. Referring to Section 3, we propose the following distance function $d_A(\hat{\Phi}, \Phi)$ for an estimator of Φ :

$$d_A(\hat{\Phi}, \Phi) = [\text{tr}(\hat{\Phi} - \Phi)^{\prime-1}(\hat{\Phi} - \Phi)A^{-1}]^{\frac{1}{2}}, \tag{3.49}$$

with

$$A = E(\overline{W}) = (1/G) \sum_g \sigma_g^2 (X_g' X_g)^{-1}. \tag{3.50}$$

Whether we set A equal to $E(\overline{W})$, \overline{W} , or even $(X_g' X_g)^{-1}$ for any g , is immaterial here, considering that all $(X_g' X_g)$ matrices are equal and that we will only look at relative values of $d_A(\hat{\Phi}, \Phi)$. The way $\hat{\Phi}_{QU}$ has been derived implies that it should perform better than $\hat{\Phi}_U$ under this criterion. The different estimators we consider here are (i) $\hat{\Phi}_U$, (ii) $\hat{\Phi}_{QU}$, (iii) S_b and (iv) $S_b(S_b + \overline{W})^{-1}S_b$. The last one is the first iterate in the procedure proposed in Theorem 4. It will tend to improve upon S_b by shrinking it and hence reducing its bias. Results are presented in Table 3.6. The distance function is normalized by its value for $\hat{\Phi} = \hat{\Phi}_U$. The entries are $d_A(\hat{\Phi}, \Phi)/d_A(\hat{\Phi}_U, \Phi)$ for the different $\hat{\Phi}$ and the selected values of G and T .

Table 3.6 : Average distance to Φ of alternative estimators, relative to $\hat{\Phi}_U$

		$T = 5$	$T = 10$	$T = 25$
$G = 5$	$\hat{\Phi}_{QU}$	0.8612 (.0018)	0.9011 (.0016)	0.9379 (.0010)
	$S_b(S_b + \overline{W})^{-1}S_b$	1.0779 (.0025)	1.0353 (.0021)	0.9828 (.0013)
	S_b	1.6105 (.0042)	1.4411 (.0033)	1.1719 (.0021)
$G = 10$	$\hat{\Phi}_{QU}$	0.8572 (.0017)	0.9047 (.0014)	0.9422 (.0009)
	$S_b(S_b + \overline{W})^{-1}S_b$	1.3461 (.0033)	1.2378 (.0026)	1.0775 (.0016)
	S_b	2.3423 (.0062)	2.0357 (.0049)	1.5015 (.0031)
$G = 25$	$\hat{\Phi}_{QU}$	0.8506 (.0014)	0.9003 (.0011)	0.9497 (.0007)
	$S_b(S_b + \overline{W})^{-1}S_b$	1.9546 (.0050)	1.7111 (.0039)	1.3196 (.0024)
	S_b	3.7761 (.0100)	3.2427 (.0079)	2.2337 (.0051)

As predicted, $\hat{\Phi}_{QU}$ is always closer to Φ than $\hat{\Phi}_U$. The estimator (iv) is an improvement upon S_b , but not in general upon $\hat{\Phi}_U$, except of course for the fact that it is necessarily positive semidefinite. It is obviously much easier to calculate than $\hat{\Phi}_{QU}$. Except perhaps on this last count, the quasi-unbiased estimator is a clear winner.

3.5 Concluding remarks

In this chapter we were concerned with the estimation of the covariance matrix of a random coefficient vector, under the restriction that the estimator must be positive semidefinite. A so-called quasi-unbiased estimator was proposed, which subordinates unbiasedness to positive semidefiniteness. Theoretically as well as experimentally, this quasi-unbiased estimator seems to offer some improvement on the usual unbiased estimator. Many issues, however, remain to be investigated. One of these is, how quasi-unbiased estimation compares to maximum likelihood estimation and other methods. Another is, what effect the choice of covariance matrix estimator may have on the estimation of the remaining parameters of the RCR model, mainly the random coefficients β_g themselves when these are estimated by a GLS-like method.

Before closing this chapter, we draw attention to the fact that the scope of Theorem 3 is wider than appears at first sight. In fact, an analogue of Theorem 3 holds for any estimator other than $\hat{\Phi}_U$ being ‘projected onto’ \mathcal{C}^K . Thus one might constrain Maximum Likelihood or MINQUE – see C.R. Rao (1972) – in the same way as quasi-unbiased estimation constrains Swamy’s estimator $\hat{\Phi}_U$. More generally, consider the estimation of any parameter, not necessarily a covariance matrix, subject to theoretical constraints. Whenever the constraints to be imposed reduce the relevant parameter space to a closed convex region within a wider space, then the projection of some unconstrained estimator onto that region will produce a (unique) constrained estimator which is at least as close to the true parameter as the unconstrained one. The crucial step in the derivation of the constrained estimator is, however, the choice of an appropriate distance function. This must correspond to an inner product (which it does if and only if it satisfies the Parallelogram Law), and it should satisfy, *e.g.*, (at least) minimal invariance properties. Incidentally, it is not obvious that the distance function d_A used in Theorem 3 is the best invariant distance function that is possible, or that $A = P\Delta P'$ is an optimal weighting matrix in some sense.

Finally, we note that the set-up used in this chapter generalizes immediately to the ‘unbalanced’ case, where there are different numbers of observations (say T_1, \dots, T_G) pertaining to each equation. The only added complication is notational.

References

- Baltagi, B.H. (1981): “Pooling. An Experimental Study of Alternative Testing and Estimation Procedures in a Two-Way Error Component Model”, *Journal of Econometrics*, 17, 21–49.
- Havenner, A. & P.A.V.B. Swamy (1980): “A Random Coefficient Approach to Seasonal Adjustment of Economic Time Series”, *Journal of Econometrics*, 15, 177–209.
- Judge, G.G., W.E. Griffiths, R.C. Hill, H. Lütkepohl & T.-C. Lee (1985): *The Theory and Practice of Econometrics*, 2nd edition, Wiley, New York.
- Luenberger, D.G. (1969): *Optimization by Vector Space Methods*, Wiley, New York.
- Magnus, J.R. & H. Neudecker (1988): *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Wiley, Chichester.
- Rao, C.R. (1972): “Estimation of Variance and Covariance Components in Linear Models”, *Journal of the American Statistical Association*, 67, 112–115.

- Rao, C.R. (1973): *Linear Statistical Inference and Its Applications*, 2nd edition, Wiley, New York.
- Rao, U.L.G. (1982): "A Note on the Unbiasedness of Swamy's Estimator for the Random Coefficient Regression Model", *Journal of Econometrics*, 18, 395–401.
- Strang, G. (1988): *Linear Algebra and its Applications*, 3rd edition, Harcourt Brace Jovanovich, San Diego.
- Swamy, P.A.V.B. (1970): "Efficient Inference in a Random Coefficient Regression Model", *Econometrica*, 38, 311–323.
- Swamy, P.A.V.B. (1971): *Statistical Inference in Random Coefficient Regression Models*, Springer-Verlag, New York.
- Taylor, W.E. (1980): "Small Sample Considerations in Estimation from Panel Data", *Journal of Econometrics*, 13, 203–223.

Chapter 4

Panel estimates of the wage penalty for maternal leave¹

Abstract

The focus of this chapter is the size of the wage penalty due to maternal leave incurred by working mothers in Germany. Existing estimates suggest large penalties with little rebound over time. We apply recent panel data methods designed to address problems of sample selectivity, unobserved heterogeneity and endogeneity to German Socio-Economic Panel (GSOEP) data. Heckman's classic treatment of selectivity is extended to deal with both heterogeneity and simultaneity. In order to exploit the actual working hours data available in GSOEP, we develop the case of a censored tobit participation model. We also investigate the sensitivity of the results to the choice of method. Our estimates imply a maternal leave wage penalty of 10 to 14% which is less persistent over time than other studies suggest. Five years after the return to work maternal wages have caught up.

4.1 Introduction

Most industrialised countries provide maternal leave legislation enabling mothers to temporarily stop working around child birth while keeping the option to return to their previous employers. Legal protection of their job does not guarantee, however, that working mothers taking maternal leave do not suffer wage disadvantages compared to women who do not interrupt their careers or interrupt them only as briefly as legally possible. The question whether this really happens, and if so to what extent, has been investigated in a number of countries with different data sets and methods leading to contrasting answers. Obvious explanations for differing estimates include differences or changes in local conditions such as institutional, legal, economic and cultural circumstances, the exact definition of maternal leave used, and the selection of indirect effects and controls included in the analyses. In addition, though, differing estimates may result from the way econometric problems such as sample selectivity, unobserved heterogeneity and endogeneity have been

¹Published as Bianca Buligescu, Denis de Crombrughe, Gülçin Menteşoğlu, Raymond Montizaan, "Panel estimates of the wage penalty for maternal leave", *Oxford Economic Papers* 61 (2009), i35-i55. The authors are all affiliated with Maastricht University. Thanks to Miriam Beblo, Lex Borghans, Charlotte Büchner, Thomas Dohmen, Kathrin Goeggel, Bart Goldsteyn, Andries de Grip, Ben Kriechel, Jan Sauermann, Sybrand Schim van der Loeff, Maarten Vendrik, other participants at the LOWER conference in Volos, September 2007, the editors, and two anonymous referees for their generous comments and suggestions. JEL classification: J31, J24, C23, C24.

dealt (or not dealt) with.

The aim of this chapter is to assess the wage penalty for maternal leave in Germany. We contribute to the literature by analyzing the labour market participation and wages of women in Germany with recent panel data methods that account for the three econometric problems just mentioned, using data from the German Socio-Economic Panel (GSOEP) covering the period 1994-2005. Many previous studies overlooked the basic fact that the choice faced by mothers is not just whether or not to work, but also whether to take on full-time or part-time employment, and more specifically how many hours to work. GSOEP contains detailed information on working hours and provides an opportunity to take fully into account women's choice of the extent of participation.

Unfortunately, since wages can only be observed for people in work, observed wages are plagued by the usual selectivity problem. A methodology to deal with the sample selectivity problem in panel data under general assumptions has been developed by Wooldridge (1995) and Semykina & Wooldridge (2006). We extend their approach in order to exploit the available information on working hours. Specifically, we replace the usual probit selection by a censored tobit process. Since an outcome limited by a boundary is more informative than a binary outcome, the tobit specification makes more efficient use of the data than a probit or logit selection. So the choice of working hours is explicitly made within the model, and working hours is an endogenous regressor in the wage equation. To assess the robustness of the resulting estimates we investigate their sensitivity to the choice of method.

In this chapter we define 'maternal leave' to include the legally mandatory break around maternity ('maternity protection') as well as additional time off work taken by a mother to look after her child in the framework of German parental leave legislation or beyond.² For the whole period covered by our data set, the law enabled women to freely prolong the duration of maternal leave up to a maximum of 36 months. We define the term 'maternal leave wage penalty' in a limitative sense, consisting in the partial or direct effect of maternal leave on the wage rate as distinct from indirect effects taking place through a reduction in working hours, foregone experience, or lost tenure, which are observable separately. The overall wage disadvantage sustained by mothers, combining direct and indirect effects, is variously known in the literature as motherhood wage penalty, child penalty or family gap (Waldfogel, 1998; Budig & England, 2001; Simonsen & Skipper, 2006).

Two main explanations can be found in the literature for the existence of a maternal leave wage penalty.³ The dominant explanation is that maternal leave, like other career interruptions, is accompanied by a deterioration of human capital (Mincer & Polachek, 1974; Mincer & Ofek, 1982). During prolonged leaves individuals cease to accumulate new experience, and their existing knowledge and skills depreciate due to non-use and new

²The German maternal leave system consists of three parts: maternity protection, protected parental leave and parental benefits. In addition to the compulsory 'maternity protection' period of 14 weeks surrounding child birth, mothers are entitled to a period of job-protected parental leave and maternal benefits. The exact terms are determined by the Federal Law of Parental Leave and Parental Benefit, which was introduced in 1986 and several times extended. Since 1993 the law gives mothers the right to prolong their maternal leave for up to 36 months in total while keeping the option to return to the former or a comparable job.

³Explanations mentioned in the literature for (overall) motherhood wage penalties include discrimination, job matching, human capital theory and signalling. The first two are of little relevance for the specific case of protected maternal leave in Germany. The discrimination approach focuses on the presence of children rather than on career interruptions (Waldfogel, 1998). The job matching argument is not properly applicable since mothers taking maternal leave have a legal right to return to a similar job with the same employer.

developments on the workplace. By standard economic arguments the deterioration of human capital will result in decreases in productivity and related wage cuts.⁴ Such effects may depend on elapsed time, and may to some extent be reversible (see Mincer & Ofek, 1982). In particular, the rate of investment in human capital may accelerate in the period directly following an interruption in order to offset the loss incurred. If this is the case, the wage penalty would be highest immediately after a career interruption and it would diminish or even disappear in later years.

The second possible explanation for the existence of a maternal leave wage penalty is that mothers choosing to take a long maternal leave may be signalling a low career commitment to their present or future employers. As Spence (1973) pointed out, in an environment in which productivity is hard to observe, spells away from work can be important indicators of weak productive commitment. They can cause a scar or stigma, resulting in a lower wage offer or slower wage growth (cf. Albrecht *et al.*, 1999).

Compared to the overall motherhood wage penalty, the component of the penalty due specifically to maternal leave has been less extensively researched. The following papers analysed maternal leave effects in German data. First, Beblo & Wolf (2002), using a sample of social security data compiled by the Institut für Arbeitsmarkt und Berufsforschung (IAB) in Nürnberg, investigated wage penalties resulting from various types of non-employment spells: maternal leave, registered unemployment, training, or other. The intermissions were measured as the fraction of time spent in each activity status during each year of the past career (covering up to twenty years). After correcting for individual heterogeneity in the form of unobserved fixed effects, Beblo & Wolf found little or no significant wage penalties for maternal leave, at least within the first five years after re-entry. On the other hand, they found substantial penalties for other spells out of the labour force. Differences between the effects of various types of career interruptions on wages lead the authors to conclude that, in addition to human capital depreciation, signalling or stigma effects were likely to be at work.

Kunze (2002) also used IAB sample data and differentiated between maternal leave and other types of career breaks. Instead of time shares, she used dummy variables indicating the occurrence of interruptions in the previous years. Her sample was restricted to full-time skilled workers during their early career. Even allowing for unobserved fixed effects, Kunze found a large wage penalty of about 18% for a maternal leave interruption in the previous year (although it varied with the type of occupation). This penalty was quite persistent over time, still reaching 13% six years after the interruption. In contrast, other 'non-work' spells were accompanied with much smaller penalties that faded away after five years. Although she interpreted her results within human capital theory, Kunze also mentioned signalling effects as an additional factor that could influence wage rates.

Two other studies of German data used treatment effects evaluation to control for selection into motherhood. Kunze & Ejrnaes (2004), using data from both the IAB sample and GSOEP, studied the evolution of women's wages in Germany around the birth of their first child and developed a matching technique to construct a comparison group with similar labour market characteristics. They reported that new mothers returning to work after a one-year interruption suffer a drop in wages of 5 to 15%, depending mainly on their qualification, with a small rebound effect due to increased returns to experience afterwards. Beblo *et al.* (2009) applied another matching strategy to IAB data and created a control group of women who had been continuously employed in the same firm. Although they concentrated on measuring the motherhood penalty corrected for

⁴In addition, Mincer & Ofek (1982) argued that wages could also be affected by a diminished investment in training during the pre-interruption periods as workers anticipate the career break.

establishment-specific effects, their results also suggest that the penalty is sensitive to the duration of the maternal leave.

In this chapter we calculate short term and delayed effects of maternal leave on wages, following an approach similar to those of Kunze (2002) and Beblo & Wolf (2002). From the GSOEP data files covering the years 1994 through 2005, we calculate the time share of maternal leave in the five calendar years preceding each wave and include these in the wage equation for working women. The sequence of time shares accounts for both the timing and the duration of interruptions. Zero time shares following an interruption reflect the time elapsed since, and a run of nonzero time shares implies an interruption of corresponding length. In this way we capture both short term and delayed effects of maternity leave on wages. Like Kunze (2002) and Beblo & Wolf (2002) we compare the penalty for maternal leave with the penalty for other career interruptions. Unlike them, however, we apply an econometric methodology that simultaneously accounts for sample selectivity, unobserved heterogeneity and endogeneity.

Our findings provide robust evidence of a wage penalty of 10 to 14% in the first year following maternal leave, which is in between estimates reported in previous studies. In subsequent years, the wages of mothers who took only short leaves catch up rapidly. These results can mostly be explained by human capital depreciation. However, we find that the penalty increases substantially when the interruption lasts into its fourth year, suggesting the presence of an additional effect. A possible explanation for the worsening of the penalty is the signal sent out by late returners. By taking up the maximum time of legally protected maternal leave, they reveal a relatively low career commitment (and a high family attachment) which may negatively affect their wage level. Additionally, our estimation results confirm the existence of selectivity, heterogeneity and endogeneity biases. In particular, not accounting for the endogeneity of actual working hours would approximately halve the estimated wage penalty.

The remainder of the chapter is organized as follows. In the next section we introduce the econometric model. The GSOEP data set used is described in Section 3. Results are presented in Section 4, and concluding remarks in Section 5.

4.2 The econometric model

Our labour market model for women is based on that of Heckman (1976, 1979). The actual wage rate earned by a woman is determined as the ‘offer wage’ that she can potentially earn on the labour market. If she has a higher reservation wage, however, she prefers not to participate and the offer wage remains unobserved. The participation decision is modelled by means of a latent variable measuring the extent of participation, with a reduced-form equation explaining how the extent of participation depends on exogenous variables, and a censoring equation generating corner solution outcomes (for zero participation). Formally we specify the following system of equations determining three variables, the wage rate y_{1it} , the latent participation measure y_{2it}^* , and the actual participation measure y_{2it} :

$$y_{1it} = x'_{1it}\beta_1 + \alpha_{1i} + u_{1it}, \quad (4.1)$$

$$y_{2it}^* = z'_{it}\delta_2 + \gamma_{2i} + u_{2it}, \quad (4.2)$$

$$y_{2it} = \max\{0, y_{2it}^*\}. \quad (4.3)$$

The subscript i refers to a random cross-section of women in their fertile age ($i = 1, \dots, N$) and the subscript t to a fixed set of calendar years ($t = 1994, \dots, 2005$). The primary dependent variable y_{1it} is the logarithm of the hourly wage rate of individual i in year

t and is only selectively observed, namely, when the latent participation decision y_{2it}^* is positive. GSOEP respondents report actual working hours for the week preceding the interview and these are taken as selected observations of y_{2it} in equation (4.3). Nonpositive values of y_{2it}^* are observed as zero values of y_{2it} , indicating no paid activity took place. The relevant population is restricted to women aged 16 to 45 at the time of entry into the panel, and who are neither students nor self-employed since the motives of the latter with respect to labour market participation are specific.

The system comprises two sets of explanatory variables. In the wage equation, is a vector of K variables some of which may be endogenous, with β_1 the corresponding vector of coefficients. The variables in x_{1it} include y_{2it} and other determinants of both the offer and the reservation wage. In the participation equation, z_{it} is a vector of ($L \geq K$) strictly exogenous variables which can be observed whether or not individual i participates in the labour market, with δ_2 the corresponding vector of coefficients. In particular, z_{it} includes any individual or household characteristic possibly affecting the wage rate indirectly and available for non-participants as well as participants.

Equations (4.1) and (4.2) both include unobserved heterogeneity in the form of time-constant individual effects denoted α_{1i} and γ_{2i} . As emphasised originally by Mundlak (1978), such unobserved individual effects are likely to be related to observed characteristics in and rather than independent. The remaining terms u_{1it} and u_{2it} are idiosyncratic, transitory disturbances, with u_{1it} likely to have nonzero conditional expectation due to selective sampling, i.e., wages being observed for labour market participants only. Our model therefore combines problems of panel heterogeneity, endogenous sample selection and endogenous regressors.

To tackle these combined problems we adopt the formal panel data framework of Wooldridge (1995) and Semykina & Wooldridge (2006). This framework presents a number of advantages for our purposes. First and most recognisably, it disposes with the classical independence assumption on random effects. Specifically, it allows for unobserved effects correlated with observed individual characteristics, à la Mundlak (1978) or Chamberlain (1982), and manages to obtain consistent estimates for the wage equation in spite of the incidental parameter problem affecting the participation equation. Second, while in the cross-sectional dimension of the panel the standard assumption of independent sampling is maintained, no assumption of independence over time within panel units is required. Estimates will therefore be robust to serial correlation in the idiosyncratic disturbances u_{1it} and u_{2it} . Third, the assumptions allow for endogenous regressors in the wage equation. To deal with such endogeneity, it is required that z_{it} can serve as a set of instrumental variables, i.e., that the variables in z_{it} are correlated with the endogenous elements of x_{1it} yet strictly exogenous in both equations. The possibility to treat a regressor as endogenous is especially relevant for the number of working hours y_{2it} in the wage equation, which is by design part of the labour market outcome.⁵

In one respect, however, the Semykina-Wooldridge framework does not suffice for our purposes. Whereas in their case the participation decision is observed only in the form of a binary indicator (namely $d_{it} = 1[y_{2it}^* > 0]$), amenable to a probit selection process, here we can partially observe the latent variable y_{2it}^* (in the form $y_{2it} = d_{it}y_{2it}^*$). This feature of

⁵Previous studies have pointed out that the number of hours worked by women is potentially endogenous with respect to the wage as high earning women might decide to work longer hours and interrupt their careers less frequently than women who earn a lower wage (see Lundberg & Rose, 2000; Beblo & Wolf, 2002). Dustmann & van Soest (1998) reject the hypothesis that working hours are exogenous for the male workers in GSOEP; arguably, if working hours are endogenous for men, this is even more likely to be the case for women.

our data motivates us to extend the existing framework by introducing a selection equation of the tobit type involving a corner solution.

The vulnerability of selection models to unobserved heterogeneity in panel data is well known. The presence of unobserved individual effects causes an incidental parameter problem undermining the consistency results that are the usual justification for Fixed Effects estimators.⁶ Semykina & Wooldridge (2006) propose to circumvent the incidental parameters problem in the probit binary choice model by formalising the dependency of unobserved individual effects on appropriate instrumental variables or time averages, following the now classic approaches of Mundlak (1978) and Chamberlain (1982).⁷ We apply this idea in the context of the censored tobit participation model.

Let us denote time-averages, initial levels or similar time-constant values of the exogenous variables in z_{it} by so-called ‘Mundlak terms’ \bar{z}_i , and interpret them as a set of persistent individual features. We can decompose the unobserved individual effects into a systematic part linearly dependent on \bar{z}_i , and an unexplained remainder term. Splitting both α_{1i} and γ_{2i} in this way, and introducing some self-explanatory notation, the model is rewritten as the following system of two equations:

$$y_{1it} = x'_{1it}\beta_1 + \xi_{01} + \bar{z}'_i\xi_1 + \varepsilon_{1i} + u_{1it}, \quad (4.4)$$

$$y_{2it} = \max\{0, z'_{it}\delta_2 + \xi_{02} + \bar{z}'_i\xi_2 + \varepsilon_{2i} + u_{2it}\}. \quad (4.5)$$

The composite disturbances in these two equations are likely to be correlated with each other, inasmuch as correlation is likely between the remaining individual effects and as well as between the idiosyncratic disturbances and ε_{2i} . To deal with the potential correlation between the composite disturbances, Semykina & Wooldridge propose the following linear conditional expectation assumption:⁸

$$E(\varepsilon_{1i} + u_{1it} \mid \varepsilon_{2i} + u_{2it}, z_{i1}, \dots, z_{iT}) = \rho_{1t}(\varepsilon_{2i} + u_{2it}). \quad (4.6)$$

While this is a formal assumption keeping the model manageable, it still provides flexibility by allowing residual correlation to subsist even after the introduction of the Mundlak terms controlling for persistent individual features in (4.4) and (4.5).

Additional parametric assumptions are needed in order to estimate the model with standard methods. The tobit selection model requires normality assumptions on the unobserved terms of the participation equation ε_{2i} and u_{2it} . The participation equation (4.5) then forms a sequence of T cross-sectional censored tobit models, in which all individual effects are absorbed by the disturbances. From each cross-section equation the residuals can be computed, to serve as selectivity correction terms in the wage equation, in the same way that inverse Mills ratios would in a probit-based specification; see, e.g., Wooldridge (1995, 2002). The estimation of the tobit participation equation is made less restrictive and arguably more robust by allowing all its coefficients to be time-varying. When we actually estimated the sequence of censored tobit models we found that the T cross-sectional

⁶This is because every new individual in the panel comes with a new parameter, namely the unobserved effect. In linear models the problem is resolved by a data transformation (either ‘within’ or ‘first differencing’) designed to eliminate the unobserved effects from the model and the corresponding likelihood function. In nonlinear models such as selection models this is not in general possible. The tobit limited choice specification shares this incidental parameters problem with the probit binary choice model.

⁷For the case of probit-based selectivity in panel data, Kyriazidou (1997) and Rochina-Barrachina (1999) propose alternative corrections based on pairwise differencing of observations; for a thorough discussion and comparison of these methods, see Dustmann & Rochina-Barrachina (2007). Different approaches based on the imputation of potential wages for subjects not in work are explored by Olivetti & Petrongolo (2006).

⁸This corresponds to parts (iii) and (iv) of Assumption 5.1 in Semykina & Wooldridge (2006).

equations were very similar and that restricting the coefficients to stay constant over time had very little effect on their estimated values. Although we exploit this finding to simplify our presentation of the tobit model estimates in Table 4.2 below, the selectivity correction terms for the wage equation will be obtained from unrestricted, time-varying estimates.

Turning to the wage equation, Semykina & Wooldridge (2006) formally present a set of conditions under which equation (4.4) can be consistently estimated in spite of the combined problems of unobserved heterogeneity, sample selectivity, and endogeneity of some of the regressors in x_{1it} . The proposed consistent estimation procedure is a pooled 2SLS regression of y_{1it} on x_{1it} , \bar{z}_i , and the selectivity correction terms, using as instruments z_{it} , \bar{z}_i , and calculated inverse Mills ratios.⁹ In order to allow for a more general error structure than required for the consistency of pooled 2SLS, we also estimate the wage equation by Random Effects 2SLS.¹⁰ In addition, we estimate a Fixed Effects 2SLS wage equation including the time-varying regressors only, as a robust basis for comparison and for specification testing.

4.3 The German SOEP data

The empirical analysis is based on data of the 95% version of the German Socio-Economic Panel (SOEP), henceforth GSOEP, for the years 1994-2005. GSOEP is a representative survey of the German population, which includes detailed information on employment, demographic and personal characteristics. We select the female respondents aged between 16 and 45 at the time they entered the panel, who are not students or self-employed. This age interval represents the fertile period in a woman's life. The economic rationale for excluding self-employed and students is that in both groups the choice of participation into the labour market is determined by other motives. Apart from economic considerations these restrictions keep our sample comparable with those used in other studies. We obtain a final sample of 9,967 women, of whom 7,419 earn wages at some time during the 12 year sample period, resulting in 59,406 panel observations overall.

This section describes first the variables selected for the main wage equation and next the (exogenous or predetermined) variables influencing the choice of working hours. Summary statistics of the main variables are presented in Table 4.1. The dependent variable of primary interest is the logarithm of the hourly wage rate. Hourly wages are calculated by dividing gross monthly earnings by the number of contractual working hours in a month; we use contractual rather than actual working hours in the denominator to avoid generating spurious correlation.¹¹ We exclude gross hourly wage rates lower than 3 euros

⁹This corresponds to step (ii) in Semykina & Wooldridge's Procedure 5.1. In the case of tobit selection, the selectivity correction terms are tobit residuals, which are not exogenous in the wage equation. Since the (calculated) inverse Mills ratios are (consistent estimates of) nonlinear functions of they are (asymptotically) exogenous and can be used as instruments. Since no consistent estimator is available for Fixed Effects probit specifications we calculate the inverse Mills ratios from a Fixed Effects logit model for participation.

¹⁰The wage equation (4) exhibits a composite disturbance term implying a nondiagonal variance-covariance matrix. Under the assumptions made this does not affect the consistency of pooled 2SLS, although a GLS method like Random Effects 2SLS is in principle more efficient.

¹¹Actual working hours are observed for only one week and show considerable dispersion. Using this data to construct hourly wage rates from monthly earnings is likely to induce a common measurement error and hence spurious negative correlation between the working hours and the calculated wage rate. We partly avoid the problem by using contractual hours of work, which are also observed on a weekly basis but should not vary from week to week.

Table 4.1 Summary statistics

Variable description	Entire sample	Participants	Others
Participation (1 if working)	0.61	1.00	0.00
Actual work hours in previous week		33.86 (12.0)	0.00 (0.00)
Hourly earnings (in logarithms)		2.36 (0.46)	
Maternal leave in year $t - 1$	0.08 (0.25)	0.03 (0.14)	0.17 (0.35)
Maternal leave in year $t - 2$	0.08 (0.25)	0.04 (0.18)	0.15 (0.32)
Maternal leave in year $t - 3$	0.08 (0.25)	0.05 (0.20)	0.14 (0.31)
Maternal leave in year $t - 4$	0.08 (0.25)	0.06 (0.22)	0.13 (0.31)
Maternal leave in year $t - 5$	0.08 (0.25)	0.06 (0.22)	0.12 (0.30)
Other interruptions in previous year (1 if interruption)	0.31 (0.46)	0.10 (0.30)	0.67 (0.47)
Number of children	1.26 (1.21)	1.12 (1.10)	1.48 (1.31)
Age of youngest child in years	9.58 (7.33)	11.61 (6.86)	6.80 (7.08)
Age in years	33.98 (9.46)	35.28 (9.00)	31.97 (9.83)
Bad health (1 if health 'not so good' or 'bad')	0.10 (0.30)	0.09 (0.29)	0.13 (0.34)
Marital status (1 if married)	0.58 (0.49)	0.56 (0.49)	0.61 (0.48)
Education in institutionalised years	11.78 (2.47)	12.11 (2.46)	11.21 (2.37)
Training (1 if trained)	0.06 (0.24)	0.10 (0.31)	0.00 (0.00)
Full-time work experience in years	7.49 (7.41)	9.06 (7.80)	4.98 (5.95)
Part-time work experience in years	2.29 (4.09)	2.95 (4.61)	1.21 (2.77)
Unemployment experience in years	0.62 (1.46)	0.46 (1.09)	0.90 (1.90)
Temporary contract (1 if temporary contract)		0.16 (0.36)	
Mismatch (1 if mismatch)		0.45 (0.50)	
Firm tenure in years		6.93 (7.05)	
Firm size (1 if 200 workers or more)		0.39 (0.49)	
East (1 if living in Eastern part of Germany)	0.23 (0.42)	0.26 (0.44)	0.23 (0.42)
Other household income, monthly (in logarithm)	7.37 (0.80)	7.21 (0.91)	7.61 (0.52)
House ownership (1 if household own their house)	0.42 (0.49)	0.43 (0.50)	0.41 (0.49)
Number of observations	59,406	33,791	21,352
Number of individuals*	9,967	7,419	6,304

Notes: Main entries are sample means, numbers in parentheses are standard deviations. Maternal leave is expressed as a fraction of the year; other interruptions as a 0-1 dummy. *We observe 3,756 individuals who appear both as a participant and as a non-participant in different waves of the panel.

or higher than 100 euros. Although Germany formally does not have a legal minimum wage, standards are introduced by employers and unions which provide a guideline for labour market negotiations.¹² Therefore we reason that hourly wages below the threshold of 3 euros are most likely due to measurement errors.

Our prime interest is to investigate how maternal leave affects wages. Following the approach of Beblo & Wolf (2002), we calculate the time share (fraction) of maternal leave in the five calendar years (denoted $t-1$, $t-2$, $t-3$, $t-4$, $t-5$) preceding each wave.¹³ The sum of these time shares measures the total duration of maternal leave taken in the last five years. To illustrate, consider a working mother who had her first baby 32 months before date, had 9 months of leave in the year of delivery, 12 months in the next, and 6 months in the last. The respective maternal time shares are 0 for years $t-5$ and $t-4$; 0.75 for year $t-3$; 1 for year $t-2$; and 0.50 for year $t-1$. The fractions add up to 2.25 years (27 months). The wage loss will depend both on the duration of the mother's leave and on the time elapsed since her return to work. The specification in terms of time shares is flexible and allows the regression relationship to reflect various patterns. In addition to the time shares, we take into account the past frequency of interruptions due to motherhood by including the number of children in the wage equation. Since other types of career interruptions may also affect the wage, we construct a dummy variable controlling for these in each of the last five years preceding each wave. The dummy variable combines the following types of interruptions: non-working, civil (or military) service, unemployment, and not working due to training or education.

Labour market participation is accounted for in the wage equation by actual (as distinct from contractual) working hours. Our reason for using actual rather than contractual working hours at this point is that the participation of women with part-time jobs turns out extremely flexible, and contractual hours are unlikely to reflect this.¹⁴ Since participation is endogenous, the actual working hours variable is instrumented with age and age squared, other monthly household income (in logarithms), house ownership, number of children, age of the youngest child (in 9 categories, see below), marital status (1 if married) and bad health (the respondent declaring her health to be either 'not so good' or 'bad').¹⁵

The human capital development indicators appearing in the wage equation are education, training, mismatch between education and the present job, experience and firm tenure. Education is measured in institutionalised years of schooling. Our training variable gives information on training incidence relevant for work in the year previous to the survey and is coded 1 if respondents trained and 0 if they did not train. The mismatch variable is based on the question whether respondents work in a job corresponding to their education. The variable is coded 1 if respondents declared that their job did not match their education. Further, we distinguish between full-time work experience, part-time

¹²See WSI-Tarifarchiv of the Hans Böckler Stiftung.

¹³This information is taken from the ARTKALEN file of GSOEP which indicates the activity status in the previous calendar year.

¹⁴Another potential problem with contractual working hours as a measure for labour force participation is that women who are on maternal leave may still report contractual working hours due to the fact that they keep their employment relationship, although they do not actually participate.

¹⁵All these instruments are variables appearing in the working hours equation that are not under the direct control of the individual, at least in the short run, and that can be assumed independent of the error components in equations (4) and (5). With the exception of 'bad health', practically the same instruments are used by Semykina & Wooldridge (2006). Heuristically, the F-statistic testing the relevance of our list of instruments in the 'first-stage' reduced form regression for actual working hours is 10.58. Using the weak instruments test of Stock & Yogo (2005), this is close to the corresponding 5% critical value for a 10% bias but comfortably above the 5% critical value for a 20% bias, viz. 11.50 and 6.39, respectively, in the case of a single endogenous regressor with 16 instruments.

work experience, and accumulated unemployment spells; for full-time work experience an additional quadratic term is included. The tenure variable may account for firm-specific human capital and is expressed as firm tenure in years.¹⁶

The wage equation further includes the following explanatory variables: dummies for temporary work contract and firm size, dummy variables indicating if the individual lives in former East Germany and if her country of origin is non-Western, year dummies, as well as the selectivity correction terms (tobit residuals) and Mundlak terms. As instrumental variables targeting the selectivity regressors we used inverse Mills ratios calculated from a Fixed Effects logit participation equation.¹⁷ The ‘Mundlak terms’ (\bar{z}_i) included are either time averages or initial levels of the explanatory variables of the participation equation, to be specified shortly.

The participation equation is set up in terms of actual working hours (for the reasons mentioned previously). This is a limited dependent variable, with zero working hours standing for unemployment or inactivity. Among the explanatory variables for participation, the need for child care time is proxied by the number of children, and in addition by a set of dummy variables indicating the age of a mother’s youngest child in nine categories (*age* < 1, 1 through 5, 6-12, 13-18, and *age* > 18; the reference category being women without children). Next we control for ‘Other household income’, meaning household income excluding the panel member’s own wage. This consists of the earnings of other household members as well as asset income. It is calculated as the difference between the net monthly total household income and the current net monthly earnings of the respondent, and expressed in logarithms.

Further explanatory variables in the working hours equation are age and age squared, education (institutionalised years), and dummy variables indicating marital status, bad health, house ownership, whether an individual lives in former East Germany, and whether her country of origin is non-Western.

Time-constant Mundlak terms for both equations (\bar{z}_i) are specified as either time averages or initial levels of time-varying variables that can be assumed exogenous and that are observable for all panel members, whether employed or not: average number of children, the averages of the dummy variables reflecting the age of the youngest child and the mother’s health, average education level, initial experience levels (full-time, part-time and unemployment), average tenure, average logarithm of other household income, and initial house ownership.

4.4 Estimation results

The first step in the estimation procedure is to estimate the working hours equation as a sequence of standard cross-sectional censored tobit regressions. Although we used unrestricted year-by-year estimates to calculate the selectivity correction terms required for the wage equation, we found that the year-by-year estimates did not differ much. To save space, Table 4.2 reports a single set of coefficient estimates for the censored

¹⁶The experience and tenure variables are based on the ‘generated files’ of GSOEP which provide information on yearly labour market states. The mismatch variable was taken as such from the GSOEP generated files. It measures the perception of respondents concerning the (mis)match between their job and education.

¹⁷The justification for using a Fixed Effects logit rather than probit selection equation for the calculation of the inverse Mills ratios is that the incidental parameters can be eliminated in the former and not in the latter (see also footnote 8). A separate instrumental variable containing the inverse Mills ratio is constructed for each of the 12 panel years.

tobit working hours equation, pooled over the whole observation period 1994-2005. The coefficients of the Mundlak terms are omitted from the table.

Inspection of the signs of the estimated coefficients reveals that all variables have effects in the expected direction. Having more children decreases the labour market participation of women, and mothers of young children work considerably shorter hours, especially while their youngest child is younger than 6 years. Mothers hardly participate on the labour market at all while they have a baby younger than one. The results reflect the fact that the need for constant child care decreases as the child grows up and that mothers get more spare time when children go to school. As the youngest child in the household grows older, mothers do again increase their labour force participation. A striking result is that mothers with children older than 12 actually work 4 to 5 more hours per week than women without children (the reference group). Furthermore, working mothers whose children are over 18 spend roughly the equivalent of an additional day on the job. The age-hours profile is concave with the maximum number of hours worked around the age of 26. Among the other variables, we find that married women and women with substantial other household income work less, while more educated women and house owners work more. Women living in former East Germany work a few more hours than those living in former West Germany, which may be related to the better availability of child care in the East (see Kreyenfeld, 2001). Finally, women with a non-Western origin tend to work a little less.

The residuals from the year-by-year tobit estimates are retrieved to serve as selection correction terms in the wage equation, to which we now turn. Three models allowing for different degrees of heterogeneity are estimated: Pooled, Random Effects (RE) and Fixed Effects (FE). The first two include the ‘Mundlak terms’ as specified at the end of previous section, whereas in the latter all time-constant terms are redundant. All three include the selectivity correction terms as additional regressors. The number of hours worked is allowed to be endogenous with respect to the wage rate since it is determined as an outcome of the labour supply decision (cf. footnote 4). In addition to the working hours, also the selection correction terms (tobit residuals) are endogenous. The method of estimation is therefore Two-Stage Least Squares (2SLS).

Table 4.3 presents the estimation results for the three models, referred to as pooled, RE and FE models, respectively. We observe a sizeable and significant negative impact of the fraction of the previous calendar year spent on maternal leave. The coefficient measures the wage penalty that a working mother is subject to if she has taken exactly one year of maternal leave in the previous calendar year; that is, the maternal fraction of year $t - 1$ is 1 and those of earlier years are zero. The size of that penalty ranges from about 10% in the FE estimates to over 19% in the pooled estimates. The pooled estimation, however, exhibits a high degree of residual autocorrelation, for which a likely interpretation is that persistent individual effects remain which are not captured by the Mundlak terms and cause biases in the coefficients. Discounting the pooled estimates, the estimated maternal leave wage penalty ranges from 10% (FE) to 14% (RE) for 12 months of leave in the first year after re-entry. It is important to note that this is a partial effect, keeping participation, experience and tenure variables constant. That is, the maternal leave penalty comes in addition to wage losses due to foregone tenure and experience accumulation.

Table 4.2 Selection equation

Dependent variable: Participation	Tobit model	
Number of children	-2.675	(0.150)
Youngest child aged < 1	-40.518	(0.738)
Youngest child aged 1	-29.357	(0.639)
Youngest child aged 2	-21.355	(0.638)
Youngest child aged 3	-13.697	(0.648)
Youngest child aged 4	-10.531	(0.662)
Youngest child aged 5	-8.786	(0.673)
Youngest child aged between 6 and 12	-1.559	(0.420)
Youngest child aged between 13 and 18	4.857	(0.452)
Youngest child aged > 18	7.592	(0.548)
Marital status (1 if married)	-0.731	(0.291)
Education in institutionalised years	2.811	(0.150)
Bad health (1 if health 'not so good' or 'bad')	-0.375	(0.451)
Age in years	1.597	(0.108)
Age squared	-0.031	(0.002)
Other household income, monthly (in logarithm) (in logarithm)	-6.757	(0.211)
House ownership (1 if house owner)	4.380	(0.326)
East (1 if living in Eastern Germany)	1.859	(0.267)
Country of origin (1 if non-Western)	-1.462	(0.348)
Mundlak terms	Included	
Number of effective observations	47,321	
Number of individuals	8,355	

Notes: Standard errors (in parentheses) should be interpreted as asymptotic lower bounds. Included in the equation but omitted from the table are year dummies and Mundlak terms.

The tobit coefficients shown are pooled estimates over 1994-2005.

Year-by-year estimates (which vary slightly over time) were used to calculate the selectivity correction terms in the wage equation.

Table 4.3 Wage equation with selectivity correction and endogeneity (2SLS)

Dependent var.: log hourly wage	Pooled		RE model		FE model	
Maternal leave in year $t - 1$	-0.193	(0.038)	-0.319	(0.029)	-0.101	(0.032)
Maternal leave in year $t - 2$	-0.025	(0.031)	-0.033	(0.022)	-0.016	(0.022)
Maternal leave in year $t - 3$	0.010	(0.028)	0.006	(0.020)	0.002	(0.020)
Maternal leave in year $t - 4$	-0.074	(0.026)	-0.061	(0.018)	-0.044	(0.018)
Maternal leave in year $t - 5$	-0.021	(0.023)	-0.019	(0.017)	-0.010	(0.017)
Other interruptions, year $t - 1$	-0.095	(0.019)	-0.086	(0.014)	-0.072	(0.014)
Other interruptions, year $t - 2$	-0.064	(0.016)	-0.065	(0.012)	-0.051	(0.012)
Other interruptions, year $t - 3$	-0.071	(0.016)	-0.065	(0.011)	-0.050	(0.011)
Other interruptions, year $t - 4$	-0.028	(0.015)	-0.008	(0.010)	0.008	(0.011)
Other interruptions, year $t - 5$	-0.043	(0.014)	-0.027	(0.010)	-0.015	(0.010)
Number of children	0.025	(0.024)	-0.017	(0.019)	-0.022	(0.021)
Actual working hours (hours/40)	-0.495	(0.095)	-0.612	(0.106)	-0.494	(0.189)
Education (institutionalised years)	0.015	(0.005)	0.016	(0.005)	0.016	(0.005)
Full-time work exper. (years/100)	4.031	(0.309)	4.872	(0.396)	4.108	(1.550)
Idem squared	-1.479	(0.516)	-3.972	(0.730)	-4.659	(1.180)
Part-time work exper. (years/100)	-0.725	(0.415)	0.040	(0.359)	1.566	(1.469)
Unemployment exper. (years/100)	-0.870	(0.588)	-2.740	(0.758)	-9.119	(2.083)
Temporary contract (1 if temp.)	-0.048	(0.016)	-0.060	(0.014)	-0.081	(0.016)
Training (1 if trained)	0.071	(0.013)	0.016	(0.009)	-0.003	(0.009)
Mismatch (1 if mismatch)	-0.155	(0.009)	-0.132	(0.010)	-0.100	(0.012)
Firm tenure (years/100)	0.396	(0.075)	0.208	(0.086)	-0.010	(0.102)
Firm size (1 if ≥ 200 workers)	0.186	(0.008)	0.124	(0.009)	0.052	(0.011)
East (1 if living in Eastern region)	-0.207	(0.012)	-0.166	(0.020)	-0.238	(0.055)
Country of origin (1 if non-Western country)	-0.044	(0.013)	-0.037	(0.021)		
Mundlak terms	Included		Included		Redundant	
Selectivity correction (p -value)	Yes	(0.00)	Yes	(0.000)	Yes	(0.000)
Standard error individual effects			0.275		0.410	
Standard error residual	0.337		0.196		0.195	
Autocorrelation of residuals	0.724	(0.009)	-0.012	(0.013)	-0.035	(0.013)
Hausman test RE vs FE:						
χ^2_{28} (p -value)					118.0	(0.000)
Number of effective observations	9,137		9,137		9,137	
Number of individuals	2,901		2,901		2,901	

Notes: Standard errors (in parentheses) should be interpreted as asymptotic lower bounds. Maternal leave is expressed as a fraction of the year; other interruptions as a 0-1 dummy. Year dummies and selection correction terms (tobit residuals) are included. Furthermore, Mundlak terms are included in the pooled and RE models but redundant in the FE model. Actual working hours and the tobit residuals are treated as endogenous. Instruments used: age and age squared, other household income (log), house ownership, number of children, dummy variables indicating the age of the youngest child, bad health, marital status, and inverse Mills ratios calculated from FE logit estimates of the participation equation. The selection correction terms are statistically significant in the RE and FE models.

Next to the maternal time share of the last calendar year, the equation includes the maternal time shares of the four preceding years. Mothers who took a maternal leave lasting longer than a year and returned to work recently have nonzero time shares in one or more of those preceding years. The corresponding terms allow for increases in the penalty depending on the duration of the leave. We find coefficients close to 0 on the delayed fractions for years $t - 2$, $t - 3$ and $t - 5$, but not $t - 4$. Thus, mothers who took two to three years of leave sustain almost the same wage penalty upon return to work as mothers who took just one year. The coefficient for year $t - 4$ is, on the contrary, sizeable and significant, implying that the penalty jumps up by 4 to 6% when the leave duration reaches or exceeds the legal job protection period of 36 months. It is not very plausible to explain the corresponding jump in the penalty by a sudden acceleration in human capital deterioration. Tellingly, whether or not a fifth year of leave was taken by such ‘late returners’ hardly has an additional effect.

For mothers who returned to work more than a year ago, the fractional terms for the years $t - 2$ to $t - 5$ represent delayed (rather than duration) effects. Focusing on mothers who took a one-year leave, one might expect them to suffer a continued wage disadvantage in the second and third years after re-entry. This turns out not to be the case. Consistently with Mincer & Ofek’s (1982) version of human capital ‘restoration’ theory, mothers’ wages soon start to catch up with those of other women. However, due to the term for year $t - 4$, we observe a temporary resurgence in the fourth year after re-entry. Five years after re-entry, we observe no significant penalty any more.

Table 4.4 Initial and five-year cumulated penalties for maternal leave

Duration of maternal leave:	1 year	2 years	3 years	4 years	5 years
Wage penalty:					
- in first year after leave	-0.101	-0.117	-0.115	-0.159	-0.169
- cumulated over 2 years	-0.117	-0.131	-0.173	-0.227	-0.237
- cumulated over 3 years	-0.115	-0.173	-0.225	-0.279	-0.289
- cumulated over 4 years	-0.159	-0.227	-0.279	-0.333	-0.343
- cumulated over 5 years	-0.169	-0.237	-0.289	-0.343	-0.353

Notes: Based on FE-model estimates of Table 4.3. For simplicity, we consider maternal leaves coinciding with calendar years. Entries are penalties expressed in log-differences (approximate proportions) of wage rates.

To clarify the relationship between the estimated penalties and the timing and duration of maternal leave we calculate cumulated wage losses in Table 4.4, based on the FE model estimates of Table 4.3. The first row presents wage losses in the first year after return to work, depending on the duration of the leave. The second to fifth rows present the cumulated wage losses incurred after a delay of 2 to 5 years. It becomes clear that wages take longer to recover from longer leaves. Remarkably, due to the strong effect of the $t - 4$ maternal time share, mothers are still confronted with a substantial wage penalty four years after re-entry. Our hypothesis is that the observed pattern is due to the late

returners. By taking the maximum length of protected parental leave, these mothers signal a relatively high degree of family attachment and a relatively low degree of labour market attachment to their (actual or potential) employers, and they are penalised for this in comparison to earlier returners. Since the estimated effect of 4 to 6% is calculated for all mothers including both early and late returners, the penalty to late returners may actually be several times higher.

Turning to the effects of other work interruptions in Table 4.3, we first observe that, as in previous studies distinguishing several types of interruptions, maternal leave does not affect wages in the same way as spells of unemployment or other types of leaves. Following Albrecht *et al.* (1999) and Beblo & Wolf (2002), this underscores the need for a theoretical explanation complementary to human capital depreciation. The occurrence of non-maternal interruptions leads to wage losses ranging between 7 and 9% in the following year. Much of this (5 or 6%) persists for two more years, but in the fourth year the penalty dwindles. Thus non-maternal breaks significantly affect wages for only three years. The finding that maternal leave sometimes has a more persistent effect on wages than other interruptions adds weight to our hypothesis that taking up the maximal length of protected parental leave entails additional signalling effects to employers.

The coefficients of the other explanatory variables are also interesting. The marginal returns to full-time work experience start at about 4% per year and are decreasing but remain positive for over 44 career years. An additional year of education is worth about 1.6% of the wage rate. Large firms (200 or more employees) pay 5 to 12% higher salaries. Unemployment experience, a temporary contract, and a mismatch between a woman's job and her education level have negative effects. The actual working hours coefficient is negative and quite significant, suggesting that jobs characterised by long hours tend to be less productive. Other things equal, women spending 8 more hours per week at work earn about 10% lower hourly wages. Also, women with a non-Western origin or living in former East Germany earn lower wages. The number of children has no direct or marginal effect on the wage rate, in addition to its indirect effect through the participation decision and the actual working hours. Finally, the effects of part-time work experience, training, and firm tenure are not clear-cut.

We further observe that the magnitude and in some cases the sign of the coefficients vary with the estimation method used. As would be expected, the largest differences are between the FE and the pooled estimates, with RE appearing as a compromise. Insufficient allowance for heterogeneity clearly leads to biases and, in particular, to overestimation of wage penalties. Compared to the pooled estimates of Beblo & Wolf (2002), the biases are attenuated here by the inclusion of the Mundlak terms as additional regressors. Although many of the estimated effects are smallest in the FE model, the negative effect of unemployment experience is largest (but its standard error increases proportionately). The selectivity correction terms (omitted from the tables) are statistically insignificant in the pooled estimation, but become very significant in the RE and FE models, even though they remain rather small in magnitude.¹⁸

Lastly, we also performed additional sensitivity analyses in which the selectivity or endogeneity corrections were omitted. Tables A1 and A2 in the appendix show the results

¹⁸The residual autocorrelation in the pooled regression deserves further comment. The symptoms of incomplete dynamics disappear once individual unobserved heterogeneity is accounted for (either in the RE or in the FE form). To a large extent dynamics and heterogeneity seem to capture each other's effects. Whereas the RE specification seems successful in terms of capturing heterogeneity, it is rejected in a Hausman test against the FE specification. We attribute the rejection to the distributional assumptions made; in our experience this is the usual outcome with large data sets and a price to be paid for the theoretical gain in efficiency.

of these two exercises. The tables show that in the absence of the corrections, the maternal wage penalty in the first year varies between 6% and 7% in the FE estimations. Hence, without a correction for endogeneity of working hours or selectivity, the wage penalty is likely to be underestimated. The control analyses also show that the finding of a jump or resurgence in the penalty (4-6%) for the time spent on maternal leave in year $t - 4$ is robust to the inclusion or exclusion of the corrections.

4.5 Conclusion

In this chapter we analysed wage losses due to maternal leave, using data on women in their fertile age from GSOEP 1994-2005 and explicitly taking into account their labour market participation decision. We extended the estimation framework proposed by Wooldridge (1995) and Semykina & Wooldridge (2006) for panel data with selectivity. In order to exploit the working hours information available in GSOEP we developed the case of a censored tobit participation model replacing the standard probit selection equation. Previous studies overlooked the fact that the choice faced by mothers is not just whether or not to work, but also how many hours to work.

Applying corrections for sample selectivity, unobserved heterogeneity and the endogeneity of working hours, we found a wage penalty of 10 to 14% for a one-year maternal leave in the year following the return to work. Other, non-maternal types of interruptions resulted in a somewhat lower penalty. Delayed effects revealed intriguing patterns. The wage penalty is small or inexistent for maternal leave taken two, three or five years ago, suggesting that mothers catch up quickly over time, whereas the effect of non-maternal interruptions decreases more gradually. However, mothers taking a fourth year of maternal leave face a jump in the wage penalty.¹⁹ No such pattern was observed for other types of interruptions.

On a methodological note, we conclude that the allowance for the endogeneity of labour supply in the wage equation, in addition to the correction for unobserved heterogeneity and sample selectivity, had a considerable impact on the estimates. If we did not treat working hours as endogenous, the maternal leave wage penalty was underestimated by almost half. Omitting the selectivity correction terms also seemed to lead to some underestimation of the penalty.

How do our results compare to previous estimates based on German data? Obviously, differences in the models, in the methodology and in the populations sampled impair all comparisons. In particular, our data set is not restricted to full-time workers. In the appendix (Table 4.A2) we report FE estimates of 6 to 7% which, in terms of the estimation method used, are best comparable to the FE estimates of Kunze (2002) and Beblo & Wolf (2002). The former study found a persistent wage penalty starting at about 18% and gradually declining to 13% six years after the maternal leave; the latter found only small and statistically insignificant penalties for maternal leave at all displacements. Our estimates occupy the middle ground, since we found a substantial wage penalty in the first year and a catch up process in later years.

It is not entirely straightforward to compare our estimates to those based on matching methods, partly because it is not easy to disentangle the maternal leave penalty from the overall motherhood penalty in those studies. Kunze & Ejrnaes (2004) reported a range of penalties for one-year leaves from 5 to 15% depending mainly on the new mother's qualification, which tallies with our figures. However, their model implied only a small

¹⁹ A similar though less clear pattern may be recognised in Kunze (2002), Table 5.

rebound effect due to increased returns to experience in the years following the return to work. The results of Beblo *et al.* (2009) suggest that new mothers suffer a motherhood penalty that depends on the duration of the leave taken, in the order of 20 to 30% (which includes the drag of foregone work experience and tenure). Our model focused on maternal leave, controlling for experience and tenure. We found that the maternal leave penalty is indeed sensitive to the duration of the leave, but only temporarily so, since five years after re-entry the maternal wage rate seems to have caught up.

We can therefore conclude on a relatively optimistic note. The catch-up process of maternal wages indicates that the deterioration of human capital during maternal leave is made up for in the period following the return to work (cf. Mincer & Polacheck, 1974; Mincer & Ofek, 1982). The timing and pattern of the restoration process depend on the length of the leave, but cannot be fully explained in the light of human capital theory. Differences in the size and time profile of the effects between maternal leave and other types of career interruptions, as well as the penalty supplement associated with a fourth year of maternal leave, indicate that other factors than human capital deterioration are likely to be at work. As an explanation for the resurgence we propose the hypothesis of a signalling effect accompanying prolonged maternal leave. This effect may be stronger in our data set than in others which are limited to full-time workers because there is more uncertainty about the labour-market commitment of part-time workers than about that of full-time workers. The observed time pattern of the wage penalty deserves further research and may help to verify the signalling hypothesis and other hypotheses concerning the determination of the cost of maternal leaves.

References

- Albrecht J.W., P-A. Edin, M. Sundström & S.B. Vroman (1999): "Career Interruptions and Subsequent Earnings: A Reexamination Using Swedish Data". *The Journal of Human Resources* 34, 294-311.
- Beblo M., S. Bender & E. Wolf (2009): "Establishment-level wage effects of entering motherhood". *Oxford Economic Papers* 61 (this issue).
- Beblo M. & E. Wolf (2002): "Wage Penalties for Career Interruptions: An Empirical Analysis for West Germany". ZEW Working Paper 02-45.
- Budig M.J. & P. England (2001): "The Wage Penalty for Motherhood". *American Sociological Review* 66, 204-25.
- Chamberlain G. (1982): "Multivariate Regression Models for Panel Data". *Journal of Econometrics* 18, 5-46.
- Dustmann C. & M.E. Rochina-Barrachina (2007): "Selection Correction in Panel Data Models: An Application to Labour Supply and Wages". *Econometrics Journal* 10, 263-93.
- Dustmann C. & A. Van Soest (1998): "Public and Private Sector Wages of Male Workers in Germany". *European Economic Review* 42, 1417-41.
- Heckman J.J. (1976): "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for such models". *Annals of Economic and Social Measurement* 5, 475-92.
- Heckman J.J. (1979): "Sample Selection Bias as a Specification Error". *Econometrica* 47, 153-61.
- Kreyenfeld M. (2001): *Employment and fertility in East Germany in the 1990s*. Dissertation, Max-Planck Institute for Demographic Research, Rostock.

- Kunze A. (2002): “The Timing of Careers and Human Capital Depreciation”. IZA Discussion Paper 509.
- Kunze A. & M. Ejrnaes (2004): “Wage Dips and Drops around First Birth”. IZA Discussion Paper 1011.
- Kyriazidou E. (1997): “Estimation of a Panel Data Sample Selection Model”. *Econometrica* 65, 1335-64.
- Lundberg S. & E. Rose (2000): “Parenthood and the Earnings of Married Men and Women”. *Labour Economics* 7, 689-710.
- Mincer J. & H. Ofek, (1982): “Interrupted Work Careers: Depreciation and Restoration of Human Capital”. *Journal of Human Resources* 17, 3-24.
- Mincer J. & S. Polachek (1974): “Family Investments in Human Capital: Earnings of Women”. *Journal of Political Economy* 82, Part 2, s76- s108.
- Mundlak Y. (1978): “On the Pooling of Time Series and Cross-Section Data”. *Econometrica* 46, 69-85.
- Olivetti C. & B. Petrongolo (2006): “Unequal Pay or Unequal Employment? A Cross-Country Analysis of Gender Gaps”. IZA Discussion Paper 1941.
- Rochina-Barrachina M.E. (1999): “A New Estimator for Panel Data Sample Selection Models”. *Annales d’Economie et de Statistique* 55/56, 153-81.
- Semykina A. & J. Wooldridge (2006): “Estimating Panel data models in presence of endogeneity and selection: Theory and application”. Discussion paper, Michigan State University.
- Simonsen M. & L. Skipper (2006): “The Costs of Motherhood: An Analysis using Matching Estimators”. *Journal of Applied Econometrics* 21, 919-34.
- Spence M. (1973): “Job market signaling”. *Quarterly Journal of Economics* 87, 355-79.
- Stock J.H. & M. Yogo (2005): “Testing for Weak Instruments in Linear IV Regression”. Chapter 5 in D.W.K. Andrews and J.H. Stock (eds.): *Identification and Inference in Econometric Models: Essays in Honour of Thomas J. Rothenberg*. Cambridge University Press, Cambridge, UK.
- Waldfogel J. (1998): “The Family Gap for Young Women in the United States and Britain: Can Maternity Leave Make a Difference?” *Journal of Labor Economics* 16, 505-45.
- Wooldridge J. (1995): “Selection corrections for panel data models under conditional mean independence assumptions”. *Journal of Econometrics* 68, 115-32.
- Wooldridge J. (2002): *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Cambridge, MA.

Appendix 4.A: Alternative estimates of the wage equation

Table 4.A1 presents results of pooled, RE and FE estimations without correction for the self-selection of women into non-participation. Omitting the selection correction resulted in smaller or very similar coefficients for most variables in all three specifications, though not for temporary contract, education, East, and country of origin. In the pooled and RE estimations, maternal leave penalties remain almost the same or slightly lower with the same time pattern as in Table 4.3. However, the FE estimates without selection correction, presented in the last column of Table 4.A1, differ substantially from their counterpart with selection correction. The actual working hours and full-time work experience variables turn insignificant, and the wage penalty due to maternal leave in the previous year decreases to 7% (from 10% in Table 4.3). Still, the time pattern of the delayed effects of maternal leave interruptions remains unchanged with a resurgence of almost 5% for the year $t - 4$.

Table 4.A2 shows the results of the pooled, RE and FE estimations without accounting for the possible endogeneity of the actual working hours. Also here, a majority of the effects are smaller than those in Table 4.3. Focusing on the FE estimates we find, without using instruments for the working hours, short-term wage penalties of 6.1% for maternal leave and of 6.2% for other (non-maternal) interruptions, which are smaller than the penalties found in Table 4.3. However, the table also shows that the time pattern of the delayed effects of maternal leave remains unchanged.

Table 4.A1 Wage equation without selectivity correction (2SLS)

Dependent var.: log hourly wage	Pooled		RE model		FE model	
Maternal leave in year $t - 1$	-0.193	(0.037)	-0.136	(0.027)	-0.070	(0.030)
Maternal leave in year $t - 2$	-0.029	(0.030)	-0.035	(0.020)	-0.018	(0.021)
Maternal leave in year $t - 3$	0.007	(0.027)	0.001	(0.018)	-0.004	(0.019)
Maternal leave in year $t - 4$	-0.065	(0.025)	-0.058	(0.017)	-0.049	(0.017)
Maternal leave in year $t - 5$	-0.014	(0.021)	-0.012	(0.015)	-0.010	(0.016)
Other interruptions, year $t - 1$	-0.081	(0.017)	-0.071	(0.012)	-0.058	(0.013)
Other interruptions, year $t - 2$	-0.052	(0.015)	-0.051	(0.011)	-0.038	(0.011)
Other interruptions, year $t - 3$	-0.062	(0.015)	-0.056	(0.010)	-0.043	(0.011)
Other interruptions, year $t - 4$	-0.023	(0.014)	-0.004	(0.009)	0.007	(0.010)
Other interruptions, year $t - 5$	-0.035	(0.013)	-0.023	(0.009)	-0.016	(0.010)
Number of children	0.015	(0.022)	-0.009	(0.017)	0.001	(0.019)
Actual working hours (hours/40)	-0.383	(0.086)	-0.337	(0.089)	0.072	(0.141)
Education (institutionalised years)	0.019	(0.005)	0.020	(0.005)	0.019	(0.005)
Full-time work exper. (years/100)	3.646	(0.278)	4.108	(0.344)	0.692	(1.316)
Idem squared	-1.707	(0.491)	-3.932	(0.674)	-4.053	(1.113)
Part-time work exper. (years/100)	-0.079	(0.366)	0.549	(0.319)	-1.614	(1.229)
Unemployment exper. (years/100)	-1.261	(0.555)	-2.727	(0.700)	-6.859	(1.929)
Temporary contract (1 if temp.)	-0.062	(0.015)	-0.085	(0.013)	-0.109	(0.014)
Training (1 if trained)	0.062	(0.012)	0.009	(0.009)	-0.010	(0.009)
Mismatch (1 if mismatch)	-0.148	(0.008)	-0.123	(0.009)	-0.094	(0.011)
Firm tenure (years/100)	0.403	(0.071)	0.170	(0.079)	-0.136	(0.097)
Firm size (1 if ≥ 200 workers)	0.179	(0.008)	0.113	(0.008)	0.046	(0.011)
East (1 if living in Eastern region)	-0.218	(0.011)	-0.207	(0.017)	-0.226	(0.052)
Country of origin (1 if non-Western country)	-0.054	(0.012)	-0.054	(0.019)		
Mundlak terms	Included		Included		Redundant	
Selectivity correction terms	Omitted		Omitted		Omitted	
Standard error individual effects			0.268		0.380	
Standard error residual	0.323		0.187		0.187	
Autocorrelation of residuals	0.730	(0.008)	-0.053	(0.013)	-0.084	(0.013)
Hausman test RE vs FE: χ^2_{29} (p -value)					119.1	(0.000)
Number of effective observations	9,137		9,137		9,137	
Number of individuals	2,901		2,901		2,901	

Notes: Standard errors (in parentheses) should be interpreted as asymptotic lower bounds. Maternal leave is expressed as a fraction of the year; other interruptions as a 0-1 dummy. Year dummies are included (though not reported). Mundlak terms are included in the pooled and RE models but redundant in the FE model. Actual working hours are treated as endogenous. Instruments used: age and age squared, other household income (log), house ownership, number of children, dummy variables indicating the age of the youngest child, bad health, and marital status.

Table 4.A2 Wage equation without endogeneity correction

Dependent var.: log hourly wage	Pooled		RE model		FE model	
Maternal leave in year $t - 1$	-0.103	(0.032)	-0.063	(0.023)	-0.061	(0.024)
Maternal leave in year $t - 2$	-0.006	(0.029)	-0.011	(0.019)	-0.010	(0.020)
Maternal leave in year $t - 3$	0.011	(0.026)	0.011	(0.018)	0.002	(0.018)
Maternal leave in year $t - 4$	-0.044	(0.024)	-0.048	(0.016)	-0.045	(0.017)
Maternal leave in year $t - 5$	-0.012	(0.021)	-0.006	(0.015)	-0.008	(0.016)
Other interruptions, year $t - 1$	-0.055	(0.016)	-0.059	(0.012)	-0.062	(0.013)
Other interruptions, year $t - 2$	-0.031	(0.014)	-0.041	(0.010)	-0.041	(0.011)
Other interruptions, year $t - 3$	-0.044	(0.014)	-0.048	(0.009)	-0.044	(0.010)
Other interruptions, year $t - 4$	-0.013	(0.013)	-0.002	(0.009)	0.007	(0.010)
Other interruptions, year $t - 5$	-0.021	(0.012)	-0.019	(0.009)	-0.015	(0.009)
Number of children	0.053	(0.021)	0.026	(0.016)	0.004	(0.017)
Actual working hours (hours/40)	0.051	(0.026)	0.005	(0.022)	-0.050	(0.026)
Education (institutionalised years)	0.012	(0.005)	0.016	(0.005)	0.017	(0.005)
Full-time work exper. (years/100)	3.026	(0.239)	3.614	(0.299)	1.590	(1.080)
Idem squared	-1.758	(0.475)	-3.213	(0.651)	-3.894	(1.054)
Part-time work exper. (years/100)	1.268	(0.242)	0.997	(0.287)	-0.750	(1.047)
Unemployment exper. (years/100)	-1.909	(0.522)	-2.693	(0.690)	-7.588	(1.843)
Temporary contract (1 if temp.)	-0.089	(0.013)	-0.101	(0.011)	-0.100	(0.013)
Training (1 if trained)	0.043	(0.011)	0.003	(0.008)	-0.008	(0.009)
Mismatch (1 if mismatch)	-0.134	(0.007)	-0.117	(0.009)	-0.096	(0.011)
Firm tenure (years/100)	0.432	(0.069)	0.165	(0.077)	-0.117	(0.096)
Firm size (1 if ≥ 200 workers)	0.167	(0.007)	0.107	(0.810)	0.048	(0.011)
East (1 if living in Eastern region)	-0.256	(0.008)	-0.246	(0.014)	-0.231	(0.052)
Country of origin (1 if non-Western country)	-0.062	(0.012)	-0.059	(0.019)		
Mundlak terms	Included		Included		Redundant	
Selectivity correction (p -value)	Yes	(0.00)	Yes	(0.000)	Yes	(0.000)
Standard error individual effects			0.267		0.359	
Standard error residual	0.312		0.185		0.185	
Autocorrelation of residuals	0.694	(0.009)	-0.043	(0.013)	-0.071	(0.013)
Hausman test RE vs FE: χ^2_{35} (p -value)					228.5	(0.000)
Number of effective observations	9,137		9,137		9,137	
Number of individuals	2,901		2,901		2,901	

Notes: Standard errors (in parentheses) should be interpreted as asymptotic lower bounds. Maternal leave is expressed as a fraction of the year; other interruptions as a 0-1 dummy. Year dummies and selection correction terms (tobit residuals) are included. Furthermore, Mundlak terms are included in the pooled and RE models but redundant in the FE model. The selection correction terms are statistically significant in the RE and FE models.

Chapter 5

Consumption smoothing in Russia¹

Abstract

Using a panel from the Russian Longitudinal Monitoring Survey (1994-2004), this chapter investigates to what extent Russian households have been able to maintain their living standards while suffering income shocks. Consumption smoothing is modelled by means of an equilibrium correction mechanism, which disentangles short-run dynamics and long-run equilibrium adjustments. GMM estimation is used to control for individual household effects in the presence of dynamics. Additionally, we differentiate between food and non-food consumption, positive and negative shocks, rural and urban areas, and several levels of poverty risk. We find that dynamics are important in the consumption equation, and that estimates are sensitive to imputation errors in home food production. No strong claims can be made with respect to heterogeneity in smoothing behaviour.

5.1 Introduction

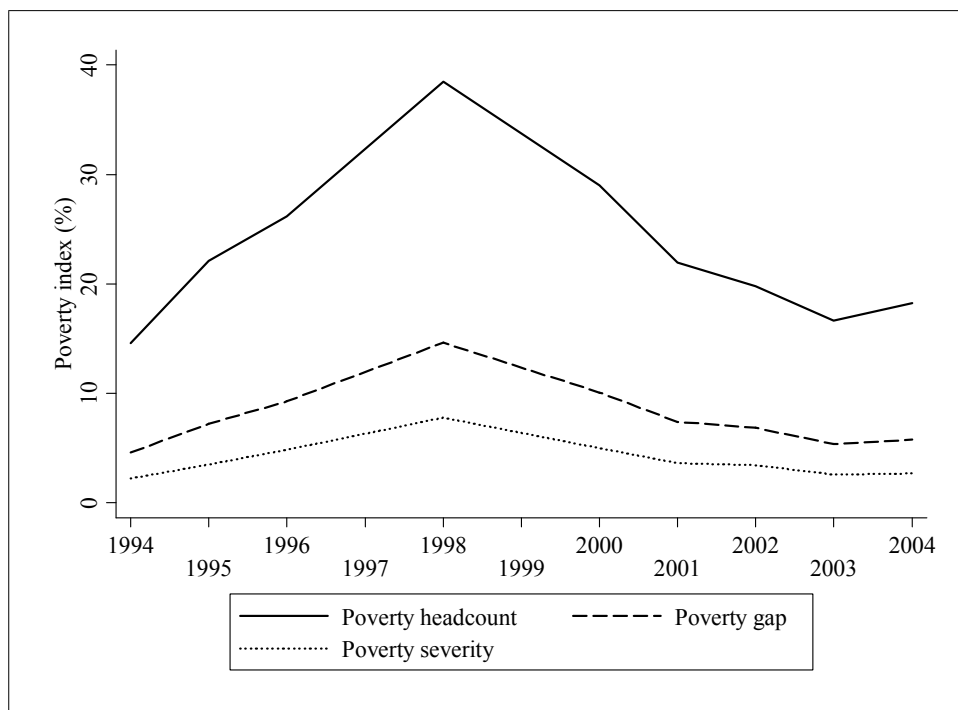
For Russian households after 1990, the transition from a centrally planned economy to a market based economy has been full of bumps, potholes and off the road experiences. Unemployment was a practically unknown phenomenon in pre-transition Russia. The closing down or privatization of the large Soviet-era industrial and agricultural companies resulted in mass unemployment and decreased job security. New skills and habits had to be acquired almost overnight and those still in a job faced wage arrears and forced leave arrangements. The cutting down of subsidies on food and energy caused surges in the cost of living. Soon after the fledgling market economy had been showing some hesitant signs of recovery, the 1998 financial and economic crisis led to defaults on domestic and foreign debts followed by a wave of bankruptcies, a devaluation of the ruble and a collapse of the stock market (Brown, 1999; Buchs, 1999; Sapir, 1999; Slay, 1999). Many households lost (large parts of) their lifetime savings. In addition to the systemic uncertainty caused

¹This chapter was co-authored with Geranda Notten (Graduate School of Public and International Affairs, University of Ottawa). Earlier versions appeared as working papers (Notten & de Crombrughe, 2006, 2007), and as Chapter 5 in Notten (2008). We are grateful to Chris Elbers, Jan Willem Gunning, Tony Shorrocks, two anonymous referees, and especially Sybrand Schim van der Loeff for their suggestions. Additionally, as requested, “we thank the Russia Longitudinal Monitoring Survey Phase 2, funded by the USAID and NIH (R01-HD38700), Higher School of Economics and Pension Fund of Russia, and provided by the Carolina Population Center and Russian Institute of Sociology for making these data available.” JEL classification: D12, D31, D91, I32.

by the structural sea-changes taking place during transition, Russians were exposed to the usual range of idiosyncratic risks including illness, disability or death of a household member, crime, job loss or crop failure. Not before 1999 did a new period of sustained economic growth begin, while the restructuring process went on.

Figure 5.1: Poverty trends in Russia (1994 – 2004)

Source: RLMS cross-sectional surveys



Perhaps not surprisingly for such a volatile period, poverty statistics show that many Russian households were unable to maintain their previous consumption levels. We illustrate this with the Russian Longitudinal Monitoring Survey (RLMS, 1994 to 2004, to be described more extensively in the sequel). Figure 1 graphs the evolution of three poverty index series, a poverty headcount, a poverty gap, and a poverty severity measure². The trends in Figure 1 are consistent with those reported by the World Bank (2004). From 1994 the poverty headcount in our panel of households rose to a peak of 38% in 1998, after which it steadily declined to about 18% in 2004. Clearly, privation was a common

²The three measures belong to the Foster-Greer-Thorbecke (1984) class of decomposable indexes, with power parameter $\alpha = 0, 1, 2$, respectively. Consumption expenditures and poverty lines will be defined more precisely in Section 3. The reported poverty measures are calculated in terms of individuals rather than households, and are estimated using the RLMS nationally representative cross-sections. The longitudinal data used for estimation in later sections are in terms of households and include only those observed at least three times in the period from 1994 to 2004.

experience. About 58% of households in our sample experienced at least one episode of poverty between 1994 and 2004, while 10% had expenditures below the poverty line *every* time they participated in the survey. Even among households remaining above the poverty line, many had living standards little higher than subsistence levels; a majority reported average expenditures less than twice the poverty threshold. For many of those, an episode of poverty would have been no more than one adverse shock away.

How do people cope with such a volatile economic environment? There is ample evidence that household consumption expenditures vary less over time than household income. Shocks affecting the earnings of household members do not (and certainly not instantaneously) affect consumption to the same extent. This observation is at the heart of Friedman's (1957) permanent income hypothesis. From a policy-making point of view, the ability of a household to deal with income uncertainty and other risks is an important aspect of its well-being. This is particularly relevant in places where large numbers of people live close to subsistence levels. Townsend (1994) and Ravallion and Chaudhuri (1997) were landmark contributions to the understanding of consumption smoothing processes in the developing world. Recently, thanks to the wide availability of the Russia Longitudinal Monitoring Survey (RLMS), a number of studies have focussed on the case of Russia; see Stillman (2001), Skoufias (2003), Mu (2006), and Gerry and Li (2008). These studies confirmed that only a fraction of all income shocks is transmitted to consumption expenditures.

What smoothing mechanisms do households use? A number of coping strategies is theoretically possible and all of them may be applicable to some extent. First, households may try and maintain their consumption levels by depleting savings, selling assets or durable goods, obtaining credit or borrowing money. Such strategies can be understood as forms of intertemporal 'self-insurance' (Skoufias, 2003). Secondly, people facing hardship may obtain help from public institutions and social networks like family, friends, communities and employers, through formal or informal transfer arrangements and insurance policies (Deaton, 1997; Fafchamps & Lund, 2003). The principle on which those arrangements and policies are based is that of 'risk sharing' among members of society. Thirdly, people will be stimulated to cultivate any available land, produce basic goods and services in autarchy, adjust their labour supply, or deliver productivity efforts. However, such reactions endeavour to maintain income rather than consumption *per se*.

Although the empirical literature on intertemporal consumption smoothing is extensive, there are more questions on which it provides no closure. One such question is how consumption smoothing in the short run relates to the necessity of balancing the accounts in the long run. There is a tendency for empirical analysis to focus on the instantaneous responses of consumption to shocks, apparently on the presumption that households adjust at once and that the adjustment does not depend on previous income and consumption levels or accumulated debts. This is not a realistic simplification. When faced with a decline in income, households may be able to maintain their consumption standards for a while in the hope that the decline is reversed, but if that hope proves vain the imbalance will become unsustainable. At some point in time savings and assets will be depleted, loans will have to be repaid, assistance from family and friends will dry up, and consumption expenditures will ultimately have to be adjusted in accordance with the permanent income stream.

A second incompletely answered question in the literature concerns heterogeneity. Exactly how much do households differ in their ability to smooth out income shocks, in particular as a consequence of unequal access to financial markets or social networks? Previous papers have shown how the individual resources and possessions of a household

and the surrounding social structures and institutions interact to determine its scope for risk management (Notten, 2008; Banerjee & Duflo, 2007). For instance, the lack of financial or physical collateral will constrain many households in their access to official loans. Some of these may still be able to obtain advances from relatives on the basis of 'social capital' (kinship). Regardless of the availability of collateral, credit rationing may occur if recent institutions are incomplete or do not yet function properly, while old institutions or services are being abolished or drastically reformed. Since individual resources and local environments differ, everyone will be able to overcome shocks in different ways and to a different extent. It seems clear that such differences may matter more than ever in a period of economic transition.

A third question which is not exhaustively treated is how to deal with the consequences of irregularities and measurement errors in income and consumption data. Infrequent purchases and the standard attenuation bias are not our main concerns here. In many household surveys, income and expenditure totals contain common imputed components representing the production and consumption of home-produced goods, known as 'auto-consumption'. (This mainly consists of food, and is the dominant component in the case of subsistence agriculture.) Imputations are by definition approximations and, since they occur on both the income and the expenditure sides of the equation, any errors of appreciation they include will tend to exaggerate the estimated income elasticities (hence offsetting the classical attenuation bias). Moreover, imputation errors are likely to be serially correlated.

This chapter proposes some new answers to those three questions. Using the economic transition in Russia as a case study, our aim is to contribute to the broader empirical literature on consumption smoothing behaviour. In a nutshell, the object of this literature is to find out what determines the ability of households to protect their ordinary consumption expenditures from systemic and idiosyncratic income shocks. Our main points are the following.

First, we investigate whether shock response behaviour is instantaneous, as implicit in previous contributions, or whether it has repercussions over a longer period. To introduce the possibility of delayed adjustments we propose an equilibrium-correction model (ECM) of consumption. Although it is linear in the variables, the ECM specification provides for flexible dynamics and introduces an explicit distinction between short-term reactions and long-term equilibrium adjustments.

Secondly, in regard to heterogeneity, we test for differences in consumption smoothing behaviour associated with the following factors: (i) the sign of income shocks (windfalls or setbacks); (ii) the type of environment (rural or urban); (iii) the household's rank in the distribution of living standards (as approximated by the level of consumption expenditures); and (iv) previous experience of poverty spells.

Thirdly, we implement a natural instrumental variable to deal with the issue of imputation errors causing spurious correlation between income and expenditures in the consumption equation and exaggerating the estimated income elasticities. This approach is general and easily applicable provided, as is frequently the case, the imputed amounts are known.

Beyond these three methodological points, the ultimate objective of the chapter remains to assess the factual ability of a particular population of households to steady their consumption expenditures. Our findings confirm that during the transition, Russian households found ways to smooth their *food* expenditures to a considerable extent, especially in the short run. For a 10% income shock the adjustment of food expenditures was as low as 1.5% in the short run and 2.3% in the long run. Previous studies of Russia

may have somewhat underestimated short-run smoothing abilities, and did not consider long-run implications. We can draw no comparable conclusions in the analysis of *non-food* consumption because specification tests give unfavourable outcomes.

Our investigation of possible heterogeneity produces only limited evidence that households reacted differently to positive and negative income shocks. The downward adjustment of food expenditures in case of a negative shock was only slightly stronger than the upward adjustment in case of a positive shock. Similarly, we obtain limited evidence that the response of food consumption to income shocks is stronger in urban environments than in rural areas, indicating that smoothing is somewhat easier in the latter. The differences between parameter estimates for the various subgroups in the sample are small, and the hypothesis of structural homogeneity of behaviour is not strongly rejected. Finally, we find that the correction for the endogeneity of income due to errors of approximation in the imputations for home production has a large impact on the estimates.

The remainder of this chapter is structured as follows. Section 2 introduces the ECM specification for consumption expenditures and discusses the empirical implications for our econometric model. Section 3 outlines how we constructed the household panel from the available 1994-2004 RLMS data, and how we defined the main variables. Section 4 reports our estimates, including specification tests and sensitivity analyses. Section 5 presents tests for the presence of several possible types of heterogeneity in consumption smoothing abilities, and Section 6 concludes.

5.2 The adjustment process of household consumption

5.2.1 Background

The notion of consumption smoothing is not really clear-cut. Perhaps the most straightforward interpretation refers to the finding that consumption is less volatile over time than income. However, the term is used more often to indicate a degree of insulation of consumption from income shocks, which depends as much on the covariance between consumption and income as on their variances. The term has also been used in the context of an efficient process of adjustment ensuring that all necessary adjustments in consumption are effectuated immediately and no predictable changes are delayed (Hall, 1978). In any case, to analyse smoothing we need to make a distinction between short-run and long-run adjustment. In the long run spending is constrained by the intertemporal (life-time) budget constraint. In the short run (because of convex preferences) consumers wish to level their living standards over time.

In this chapter, consumption smoothing refers to the empirical ability of a household to smooth its consumption expenditures, that is, the extent to which it actually succeeds in isolating its consumption expenditures from all kinds of income shocks.

The premise of many consumption smoothing models for households is that the threat of being unable to satisfy the needs of its members in the (near) future motivates precautionary behaviour, which is geared towards the maintenance of a minimal living standard under all circumstances. In a world with perfectly functioning risk-management institutions, households would have the possibility to fully insure against income risks. In the real world, and certainly in a country in transition, households can formally insure their living standards only very partially or not at all.

Disruptions may affect household (real) income either directly or indirectly. Job loss, wage payment arrears and involuntary (unpaid) leave have a direct impact on household income. Events such as illness, disability or death of a household member also have a

direct impact when the person involved was actively participating in income generating activities. Even when such events involve non-active family members there might be an indirect impact on the household income through adjustments in the internal household task division. Active members may have to reduce labour supply so as to free time for caring activities or other household tasks. Extremes in climate (e.g., drought, floods), epidemic diseases, crop failures, apart from the direct damage they may cause to households, often also reduce the amount of home produce that can be sold or consumed. Inflation or price adjustments affect the real value of consumption that can be attained, particularly if (as is usually the case) income does not or not fully compensate for price increases.

The empirical measurement of consumption smoothing abilities in developing countries has been inspired by two strands of intertemporal consumer choice models, often labelled risk-sharing and permanent-income or life-cycle models (Deaton, 1992, 1997). In the first strand of models opportunities for risk sharing between risk adverse consumers arise because their income distributions differ across 'states of the world'. Depending on the formalisation of the mechanism, the risk sharing may be achieved either through support networks within a community (e.g., kinship, social network, village economy) or by means of a market on which state contingent claims (so-called Arrow securities) are traded. In the case of full sharing, household consumption will grow at the same rate as that of the community. Shocks to a household's resources, be they positive or negative, will be absorbed in the common pool. This yields the testable 'perfect insurance' hypothesis that, when controlling for changes in community resources, household consumption should be unresponsive to income changes. While general prosperity will benefit all members, the insurance community will be powerless when faced with aggregate shocks (an economic crisis, a drought, the bankruptcy of a large firm). Such risk sharing models are the theoretical basis underlying many empirical studies of consumption smoothing (among others Altonji *et al.*, 1992; Cochrane, 1991; Mace, 1991; Townsend, 1994; Ravallion & Chaudhuri, 1997), including some concerning Russia in particular (see below).

The second strand of models derives from Milton Friedman's permanent income theory which postulates that consumption is determined by the present value of life-time resources. Forward-looking consumers use savings and credit markets to stabilise their consumption over future time periods (see, e.g., Deaton 1992). The testable implication is that anticipated income changes should not affect consumption at all. Only unforeseen shocks may have an effect on consumption, depending on how persistent they are perceived to be. If a shock is transitory, consumption will hardly be affected; if a shock is permanent, consumption will change commensurately as consumers review their life-time consumption plans.

Friedman's permanent income model has been extended in various ways in order to allow for the presence of market failures such as liquidity constraints, credit rationing, or information problems blocking access to credit for certain groups of consumers. Indeed, liquidity constrained or low wealth consumers may find themselves unable to buffer consumption expenditures, even with respect to anticipated income changes. Buffer stock models predict that some liquidity constrained consumers could even reduce their consumption by more than the size of a negative income shock because their aim is to maintain a sufficient precautionary stock of cash. The closely related life-cycle models have been used to explain not only consumption smoothing over the life-time, but also smoothing patterns over shorter intervals such as agricultural production cycles (within a year), business cycles (year to year), and working life (Browning and Crossley, 2001). In the case of Russia in transition, a careful empirical implementation of the permanent-income/life-cycle model is that of Stillman (2001).

5.2.2 Static specification

As emphasised by Stillman (2001), a direct test of the validity of the permanent-income/life-cycle model would require either unrealistic simplifications or information that is not available. Therefore, attempts to measure smoothing behaviour typically rely not on full structural specifications but on relatively simple (linear) reduced form models. Deaton (1992, 1997) reviews earlier contributions in that area. We review here only the main empirical specifications that have been used in the recent studies of Russian consumption smoothing behaviour by Stillman, (2001), Skoufias (2003), Mu (2006), and Gerry and Li (2008).³

These studies share a similar basic behavioural equation which can be summarised as follows:

$$\ln c_{i,t} = \beta_1 \ln y_{i,t} + \sum_{j=1}^J \gamma_j x_{j,i,t} + \sum_{\ell=1}^L \delta_\ell D_{\ell,i,t} + \nu_i + \varepsilon_{i,t} \quad (5.1)$$

where $\ln c_{i,t}$ and $\ln y_{i,t}$ denote the logarithms of consumption and income per capita, respectively, for household i in period t ; $x_{j,i,t}$, $j = 1, \dots, J$ denote household characteristics affecting the marginal utility of consumption, the so-called ‘taste shifters’; $D_{\ell,i,t}$, $\ell = 1, \dots, L$ are binary variables identifying each local community separately in each survey round t ;⁴ ν_i is an unobserved household-specific individual effect; and $\varepsilon_{i,t}$ is an independent ‘idiosyncratic’ white noise (or, at least, serially uncorrelated) disturbance term. The income elasticity of consumption, β_1 , is the key parameter measuring the extent of (or rather, lack of) consumption smoothing. In the case $\beta_1 = 0$, consumption is entirely insensitive to income changes and in this sense income shocks are ‘smoothed out’. At the other extreme, $\beta_1 = 1$, and income changes affect consumption expenditures proportionally; in this case no smoothing takes place over time or across households, at least at the frequency of observation. The interpretation of the empirical relationship will depend on whether we view smoothing behaviour as resulting from risk-sharing or from life-cycle behaviour (or a combination of both). Unless the structural behaviour underlying equation (5.1) can be identified, there is no way that unambiguous evidence in favour of a single smoothing mechanism can be obtained. What equation (5.1) does provide is a framework to estimate the behavioural (in)sensitivity of consumption to income changes, which can be interpreted as the extent to which households are able to isolate consumption from income shocks, whatever the means.

5.2.3 Risk-sharing interpretation

Under the *risk sharing* interpretation of smoothing behaviour, the theory would be that a household’s consumption grows or declines at the same rate as that of its risk-sharing community. The dummies $D_{\ell,i,t}$ capture effects specific to local community ℓ at time t , hence they control for community-level uninsurable shocks. The taste-shifters $x_{j,i,t}$ measure household characteristics affecting consumption needs, such as demographic composition. The estimate of β_1 can be used to test the full insurance hypothesis. If risk-sharing institutions function perfectly then idiosyncratic income shocks affecting a household should

³These studies all use data from the Russia Longitudinal Monitoring Survey (RLMS) albeit with differences in the included waves, households, variables, and the definitions used. For space reasons we limit our discussion to those aspects that are relevant to the purpose of this paper.

⁴A local community is defined as an RLMS Primary Sampling Unit and comprises a geographical area that is roughly equivalent to the administrative definition of a ‘raion’ in the Russian Federation.

leave its consumption entirely unaffected ($\beta_1 = 0$). A positive income elasticity ($\beta_1 > 0$) indicates less than full insurance.

While the precise model specifications and estimation strategies vary, the three papers by Skoufias (2003), Mu (2006), and Gerry and Li (2008) test and reject the full insurance hypothesis under the risk-sharing interpretation. We will discuss them briefly, but first mention for future reference this first-differenced transform of equation (5.1):

$$\Delta \ln c_{i,t} = \beta_1 \Delta \ln y_{i,t} + \sum_{j=1}^J \gamma_j \Delta x_{j,i,t} + \sum_{\ell=1}^L \delta_\ell \Delta D_{\ell,i,t} + \Delta \varepsilon_{i,t}. \quad (5.2)$$

First-differencing is one straightforward way to eliminate the unobserved household effects ν_i from equation (5.1) without making restrictive assumptions about their dependence on the observed household characteristics.

Skoufias (2003) differenced incomes and expenditures in equation (5.1), eliminating ν_i and replacing $\varepsilon_{i,t}$ by its first difference $\Delta \varepsilon_{i,t}$, but he did not difference the $x_{j,i,t}$ and $D_{\ell,i,t}$ terms and so estimated a variant of the following type:

$$\Delta \ln c_{i,t} = \beta_1 \Delta \ln y_{i,t} + \sum_{j=1}^J \gamma'_j x_{j,i,t} + \sum_{\ell=1}^L \delta'_\ell D_{\ell,i,t} + \Delta \varepsilon'_{i,t}. \quad (5.3)$$

Estimating this by Pooled OLS, Skoufias found an income elasticity of total consumption expenditures of 0.20. Instrumenting income changes using household-level shock indicators (wage payment arrears, forced leave and unemployment) he obtained a slightly higher estimate of 0.23 (with some indication that the higher estimate was due to urban households only). Replacing the (undifferenced) community dummies ($D_{\ell,i,t}$) by the *change* in average community income resulted in a very similar elasticity of 0.24 with respect to average community income. Splitting consumption into food and non-food expenditures suggested that the community income effect was due exclusively to food expenditures. Skoufias concluded that he found ‘strong evidence of partial insurance and community risk sharing in food consumption’ (p.79).

Mu (2006) used a demeaning (‘within’) transformation to eliminate the ν_i , and instrumented income with a refined set of household-level shock indicators. His fixed-effects instrumental-variables estimate of the income elasticity of total household expenditure came out practically identical to that of Skoufias at 0.23. He further expanded equation (5.1) with additional terms interacting income with indicators of initial wealth and human capital (education) levels. These terms revealed some differences in smoothing ability between households with different stocks of physical assets and levels of education, at least after the panel was split in rural and urban samples. Wealth seemed to increase the smoothing ability of rural households significantly, whereas education did the same in urban households. For instance, Mu estimated income elasticities of 0.16 for urban households with primary or lower schooling, and 0.26 for rural households with low wealth levels; whereas in favourable cases elasticities as low as 0 were found.

Gerry and Li (2008) used the same differenced form of equation (5.1) as Skoufias, but then replaced income change in the specification by a combination of dummy variables indicating types of idiosyncratic shocks (e.g., unemployment, wage arrears, forced leave) and coping mechanisms available to households (e.g., pensions and social benefits, sale of assets, help from relatives and home production of food). They focussed on the three RLMS rounds surrounding the Russian financial crisis of 1998, namely those of 1996, 1998 and 2000. They did not report an overall income elasticity of consumption but tested the

full insurance hypothesis through the various estimated shock and response effects, using pooled OLS as well as quantile regression. The shock variables they included had large and statistically significant effects; for instance the 1998 crisis decreased consumption by over 60%, unemployment did so by over 10%, and payment arrears did by over 7%. Like those of the previous studies, these results seem to imply both a clear rejection of the full insurance hypothesis and the active use by households of available smoothing mechanisms.

5.2.4 Life-cycle interpretation

In contrast to those three contributions, Stillman (2001) based his tests on a permanent income or life-cycle hypothesis, and accordingly did not include the community dummies ($D_{\ell,i,t}$ terms) in the model. Distinguishing between anticipated and unanticipated as well as between permanent and transitory income shocks, he focussed on unanticipated, transient, aggregate shocks by instrumenting income changes with a set of exogenous shock variables (oil prices, exchange rates, and community-level variations in wage and pension arrears). The model was estimated by 2SLS after a demeaning (within) transformation to eliminate the ν_i . The estimates of the income elasticity of food and total non-durable expenditures were very high, ranging from 0.70 to 1.10, implying little smoothing and a strong rejection of the permanent income hypothesis. Remarkably, poorer households faced with an income shock not only reduced their consumption but simultaneously increased their savings. Since such precautionary behaviour is not predicted by the standard life-cycle model, Stillman interpreted it as the possible outcome of a buffer stock model in which credit-constrained consumers deal with the uncertainty of future incomes by achieving a target level of cash savings.

The common conclusion emerging from the studies reviewed here is that Russian households in the late 1990s found some ways to smooth their consumption but they still remained vulnerable to many shocks. A common limitation is that they focus on the instantaneous responses of consumption to income, implying that households understand the exact nature of shocks as they occur and adjust their behaviour independently of previous income and consumption levels. This is a convenient but bold simplification. Households will need time to find out the transient or definitive character of a shock, and while they may be able to maintain their consumption standards temporarily in the hope that the decline is reversed, if that hope proves vain then at some point in time assets will be depleted, loans will have to be repaid, assistance from family and friends will dry up. Ultimately, consumption expenditures will have to be adjusted and made sustainable for the longer term. We therefore extend the empirical model discussed above with an equilibrium-correction mechanism allowing for delayed adjustments and long-term dynamics.

5.2.5 Equilibrium Correction

Equilibrium-correction models, or error-correction models (ECM) as they were originally called, were introduced in the context of aggregate consumption functions by Davidson *et al.* (1978) and related formally to the concept of cointegration by Engle and Granger (1987). They are representations of autoregressive distributed lag models designed to be easily interpretable in terms of adjustments to long-run economic equilibria and short-run deviations from equilibrium. The specification is general enough to encompass the more restrictive distributed lag and partial adjustment mechanisms. In terms of consumption $c_{i,t}$ and income $y_{i,t}$ per capita as defined before, we start from a first-order autoregressive-

distributed-lag process including individual effects and observed household characteristics:

$$\begin{aligned} \ln c_{i,t} &= \alpha \ln c_{i,t-1} + \beta_1 \ln y_{i,t} + \beta_2 \ln y_{i,t-1} \\ &+ \sum_{j=1}^J \gamma_j x_{j,i,t} + \sum_{\ell=1}^L \delta_\ell D_{\ell,i,t} + \nu_i + \varepsilon_{i,t}. \end{aligned} \quad (5.4)$$

An interesting reparameterisation emphasising the ECM interpretation of this process reads

$$\begin{aligned} \Delta \ln c_{i,t} &= \beta_1 \Delta \ln y_{i,t} - (1 - \alpha) (\ln c_{i,t-1} - \theta \ln y_{i,t-1}) \\ &+ \sum_{j=1}^J \gamma_j x_{j,i,t} + \sum_{\ell=1}^L \delta_\ell D_{\ell,i,t} + \nu_i + \varepsilon_{i,t}, \end{aligned} \quad (5.5)$$

where we have imposed the stability condition $\alpha < 1$, and defined

$$\theta \equiv \frac{\beta_1 + \beta_2}{1 - \alpha} \quad \text{for} \quad \alpha < 1. \quad (5.6)$$

The term $(\ln c_{i,t-1} - \theta \ln y_{i,t-1})$ in equation (5.5) can be interpreted as the (relative) deviation of consumption from its equilibrium level in the previous period; as long as this term is nonzero, even in the absence of a current change in income, there is pressure for adjustment (since we imposed $\alpha < 1$). The coefficient θ parameterises the income elasticity at equilibrium and is easily seen to be the cumulative ‘long-run’ response of consumption to income changes; whereas β_1 measures only the immediate or ‘short-run’ response. The coefficient $[-(1 - \alpha)]$ of the deviation from equilibrium is a ‘speed of adjustment’ parameter. If α is close to 1 the speed is practically zero and the adjustment is infinitely slow, whereas if $\alpha = 0$ the speed is maximal and the lagged adjustment is completed in a single period.

Equations (5.2) and (5.3), unlike the ECM in (5.5), ignore errors of adjustment inherited from the past. Either the necessary readjustments are relegated to the idiosyncratic error term $\varepsilon_{i,t}$, disregarding the likelihood of induced serial correlation; or it must be the case that $\alpha = 1$ (as proscribed here), in which case the data contain no information on any relationship binding consumption levels to income in the long run (even if such a relationship still makes sense).

In equation (5.5), we can give consumption smoothing a precise quantitative meaning. Since β_1 measures the short-run response of consumption to income changes, the extent of ‘short-run smoothing’ can be defined as $(1 - \beta_1)$; that is the proportion of income shocks that is not transmitted to consumption expenditures within the same year. By analogy, since θ distinctly measures a long-run response, one could interpret $(1 - \theta)$ as the extent of ‘long-run smoothing’; that is, more precisely, the proportion of income shocks that never gets transmitted to consumption expenditures. However, smoothing is essentially a delaying tactic. A conceptually attractive measure of smoothing behaviour could therefore be based on the delay of adjustment, that is, the time it takes for income shocks to be transmitted to consumption. The *mean delay of adjustment* is easily calculated from an ECM with regular coefficient values, as follows:

$$\begin{aligned} \bar{\ell}_y &\equiv \frac{\alpha \beta_1 + \beta_2}{(\beta_1 + \beta_2)(1 - \alpha)} = \frac{1}{1 - \alpha} \left(1 - \frac{\beta_1}{\theta} \right) & \text{for} & \quad \beta_1 \geq 0, \beta_2 \geq 0, \beta_1 + \beta_2 > 0, \\ & & \text{and} & \quad 0 \leq \alpha < 1. \end{aligned} \quad (5.7)$$

This is a decreasing function of β_1 , and an increasing function of α and β_2 (or θ); in other words, the mean delay parameter is higher the larger the extent of short-term smoothing ($1 - \beta_1$), and the lower the correction speed ($1 - \alpha$). Instantaneous adjustment is only possible if $\alpha = \beta_2 = 0$. The closer α comes to 1, the longer households can delay the adjustment of their expenses to their ‘permanent’ income stream. Hence $\bar{\ell}_y$ measures how long consumers faced with a crisis can put off tightening their belts, on average.

In ECM models the distinction between short-run and long-run effects is natural, quantifiable and testable. In general, long-run income elasticities are thought to be (substantially) larger than short-run ones, and we may therefore expect $(1 - \beta_1)$ to be (substantially) larger than $(1 - \theta)$. Smoothing behaviour (delaying) is what causes the difference between the short and the long run. In the aggregate and in the cross-sectional dimension, household income and household consumption seem to move more or less proportionally, suggesting unit elasticities in the long run. If it is the case that $\theta = 1$, then the mean delay of adjustment $\bar{\ell}_y$ is clearly the more relevant measure of the extent of household smoothing behaviour.

5.2.6 The GMM solution to the estimation problem

The estimation of the ECM in equations (5.4) or (5.5) is not without problems: $\ln c_{i,t-1}$ is necessarily correlated with the unobserved household effect, ν_i , and so is any further lag of $\ln c_{i,t}$. As shown by Nickell (1981), this correlation is not obliterated by the usual within transformation (demeaning). Arellano & Bond (1991), Arellano & Bover (1995), Blundell & Bond (1998), and others have developed a large family of consistent and more or less efficient Generalized Method of Moments (GMM) estimators for this type of dynamic panel equation. The attractiveness of the GMM approach depends (among other things) on the efficient use of available instruments, the optimal weighting of the orthogonality conditions between these instruments and the idiosyncratic disturbances, and above all a large sample size in the cross-sectional dimension of the panel.

The household unobserved effects, ν_i , are first removed by an appropriate transformation; since the within (demeaning) transformation fails to eliminate dynamic panel bias, the obvious appropriate transformation is first differencing. However, as an alternative, Arellano & Bover (1995) proposed an adaptation of the within transformation which has some advantages over first differencing in panels with data gaps, namely the ‘forward orthogonal deviations’ obtained (essentially) by subtracting the average of all available *future* observations from the current observation (and applying a simple scale factor). We indicate either transformation, first differencing or forward orthogonal deviation, by three floating dots in superscript ($\overset{\dots}{\cdot}$), as in the following transformed equation:

$$\begin{aligned} \ln \overset{\dots}{c}_{i,t} &= \alpha \ln \overset{\dots}{c}_{i,t-1} + \beta_1 \ln \overset{\dots}{y}_{i,t} + \beta_2 \ln \overset{\dots}{y}_{i,t-1} \\ &\quad + \sum_{j=1}^J \gamma_j \overset{\dots}{x}_{j,i,t} + \sum_{\ell=1}^L \delta_\ell \overset{\dots}{D}_{\ell,i,t} + \overset{\dots}{\varepsilon}_{i,t}. \end{aligned} \quad (5.8)$$

If $\varepsilon_{i,t}$ is white noise and first differencing is used, then $\overset{\dots}{\varepsilon}_{i,t} \equiv \Delta \varepsilon_{i,t}$ is a first-order moving average process or MA(1), implying first-order serial correlation (though none of a higher order) in the transformed disturbances. In the case where forward orthogonal deviations are used, the transformed disturbances retain the serial uncorrelatedness (and even the white noise property) possessed by $\varepsilon_{i,t}$. The lagged dependent variable $\ln \overset{\dots}{c}_{i,t-1}$ is still endogenous in equation (5.8), but the panel itself provides instruments in the form of $\ln c_{i,t-2}$ and further lags (untransformed). As the longitudinal dimension of the panel

increases, more and more lagged forms can be added as instruments, and actually care must be exercised not to let their number (which grows quadratically with T) become disproportionate.

Instead of removing the unobserved effects, ν_i , by differencing or demeaning the data, there is also the possibility of estimating equation (5.1) in levels using as instruments lagged differences that are uncorrelated with the ν_i . Blundell & Bond (1998) explored the conditions under which such instruments would be valid, and designed a ‘system GMM’ estimator combining moment conditions from both the differenced and the level formulations. It is plausible that this approach would allow a more efficient use of our RLMS data, where the cross-section is large, the time dimension is very short, consumption levels vary up and down without excess persistence, and we expect these levels to be informative. We will present the different estimates in a later section.

In principle, the GMM approach outlined above can resolve the ‘dynamic panel bias’ that complicates the estimation of a consumption equation including lagged levels. However, before we proceed, it is opportune to mention another potential source of bias. The income variable $y_{i,t}$ is certain to be measured imperfectly and likely to contain endogenous adjustments. On both counts, it should not be treated as exogenous and must itself be instrumented. The classical attenuation bias is not our only (or main) concern here. In the RLMS data, income and consumption both contain a substantial imputed component representing the production and consumption of home-produced goods (mostly food), known as ‘autoconsummation’. Even if the reported quantities of home produce are precise, the imputed prices are unlikely to be accurate. Since the imputed component is present on both the left and the right-hand side of the equation, it may cause an upward bias in the current income coefficient (hence offsetting the ordinary attenuation bias). Furthermore it is thinkable that such imputation errors are serially correlated. To resolve this problem we propose to instrument the income variable with its non-imputed part (‘income net of imputations’). Clearly, omitting imputed components from the instrument will not resolve any measurement error or endogeneity bias due to non-imputed income and more action will be needed. We will return to this issue in the following sections.

5.3 Data and measurement issues

5.3.1 Data

As mentioned earlier, the data base for this chapter is Phase II of the Russia Longitudinal Monitoring Survey (RLMS) covering the period 1994-2004.⁵ The RLMS is considered the first nationally representative random sample of the Russian population (Heeringa, 1997; Zohoori *et al.*, 1998; Mroz *et al.*, 2005). It was constructed as a multi-stage probability sample of about 4000 ‘dwelling units’ (addresses) covering a huge territory. There were survey rounds in most years though unfortunately not in 1997 and 1999. Data collection took place almost systematically in the last three months of the year. The year by year poverty estimates in this chapter are based on the repeated (cross-sectional) surveys, but the longitudinal dimension of the data makes it possible to investigate poverty dynamics and consumption smoothing abilities. Observations with missing information on the required variables were dropped. The longitudinal sample consists of an unbalanced panel of households that have been observed for at least three consecutive years.

⁵Detailed information is provided on the RLMS website, at the following url:
<http://www.cpc.unc.edu/projects/rlms/home.html>.

A possible point of concern is that of attrition in the panel. Analyses of the response rates show that these are satisfactory (typically above 80%), although less so in Moscow and St-Petersburg, and amongst small households with young members (cf. Heeringa, 1997). Some households are inevitably lost from the panel as a consequence of moving house, splitting up, or other more or less natural causes of attrition.⁶

The RLMS attempts in a number of ways to prevent nonresponse and other drop-outs from affecting the representativity of the survey sample. From 1996 on, households who moved away from the originally sampled address were tracked if possible to their new address and included in subsequent surveys. ‘Lost’ households were in principle replaced by ‘fresh’ households who moved into the originally sampled address after this was vacated. So-called ‘offspring’ households, newly created when a participating household split up, were also invited to participate. Fresh contingents of households from Moscow and St. Petersburg were added to the RLMS sample precisely to compensate for the relatively high metropolitan drop-out rates.⁷ It is therefore reasonably likely that the representativeness of the survey sample was maintained if not improved over the years.

In order to satisfy specific demands on data availability, Mu (2006) constructed a balanced, much more selective sample than ours; when he accounted for selectivity in his sample, he found that an attrition correction based on inverse probability weighting had very little effect on his estimates. Our panel is more inclusive, also incorporating newly observed households from later surveys, on the sole condition that they were observed in three consecutive rounds. Therefore we argue that our sample is if anything much less likely than Mu’s to be affected by attrition bias.

As controls for household demographics and other characteristics, our consumption equations will include the number of household members in each of six age categories: children aged 0-6 (denoted $x_{1,i,t}$), children aged 7-18 ($x_{2,i,t}$), males aged 19-60 ($x_{3,i,t}$), females aged 19-55 ($x_{4,i,t}$), males aged 60 and above ($x_{5,i,t}$), and females aged 55 and above ($x_{6,i,t}$). In addition, we use binary dummy variables (denoted generically by $D_{\ell,i,t}$) to indicate the survey round and the regional location of the household’s dwelling.⁸ Finally, poverty status was determined by comparing the household’s total expenditures (as defined in the next subsection) to a household-specific poverty line constructed by the RLMS group itself. The RLMS poverty lines are based on regional age-and-gender-specific food baskets valued at regional prices, and adjusted for the demographic composition of the household; that is, they take into account the household’s composition and (regional) location.⁹

5.3.2 Composition of income and expenditures

Income data are retrieved from the RLMS constructed income files; expenditures data from the RLMS constructed expenditures files. We will now provide details on the construction

⁶A referee has pointed out that adverse income shocks may contribute to the decision of a household to move house and, hence, to drop out of the survey. Furthermore, Friebel and Guriev (2005) argue that “the widespread use of in-kind wages and wage arrears in Russia may be explained as an attachment strategy” restricting the mobility of workers; they also present some evidence corroborating this. Though beyond the scope of our paper, such hypotheses deserve attention.

⁷Indeed, in the early years of the RLMS, we find that panel households have somewhat lower income and expenditures than average and that rural areas are somewhat overrepresented, whereas from 1998 on such differences become much less clear cut (Notten, 2008, pp. 135-137; Mu, 2006).

⁸The effect of location was also tested using more detailed dummies for ‘communities’, defined as the survey’s primary sampling units, as well as interactions between time (or wave) dummies and location. Such extensions did not affect the coefficients of interest in a substantial way.

⁹The RLMS poverty lines are on average about 66% of the Minimum Subsistence Level used by the Russian authorities to determine eligibility for means-tested benefits (Notten and Gassmann, 2008, p. 89).

of our household income and expenditure totals. We start with discussing some data issues affecting both totals before discussing them separately.

First, it should be noted that different income and expenditure categories are based on different recall periods varying from a week for food purchases and a month for wage income, to three months for purchases of durables and one year for the amount of home produced crops. Subsequently all amounts are adjusted to a 30-day basis that broadly corresponds with the monthly income and consumption levels at the time of the interview.

Secondly, the value of food produced and consumed at home ('autoconsumption') is imputed by valuing the reported quantities of foodstuffs at local market prices. In Russia, home-produced food represents an important source of household income and consumption, especially but not only in the Russian countryside, in terms of both the number of households involved and the share of consumption or income (cash and in kind). Home-grown food is therefore included in the income and consumption definitions. This poses a specific measurement error issue since imputation errors will affect both sides of the equation equally.

Thirdly, there have been changes in the RLMS questionnaires over the years. Inconsistencies in questionnaire items may obviously induce spurious changes in consumption and income which will distort the evidence on consumption smoothing. After a thorough comparison of the questionnaires and examination of summary statistics we decided that it was safer to exclude expenditures on 'health' and 'other services' from our aggregates.¹⁰

On the income side, from a theoretical perspective, one would like to separate genuine shocks from induced income smoothing responses which attenuate the magnitude of the shock. In practice, it is hard to make a clear distinction between genuine shocks and induced household responses. To illustrate, the wage loss due to a household member losing her job could be compensated (partly) by another member working overtime. We did not attempt to identify such compensating behaviour and included all wage earnings in income. Similarly, although income from certain social protection arrangements like sick pay and unemployment benefits may be considered as post-shock income, they are unlikely to be endogenous to consumption and are included in our income definition.¹¹ Conversely, income from property and jewellery sales, transfers received from friends and relatives, and cash borrowings, were excluded because the use of these resources is more likely to be induced by the need to finance current consumption expenditures. Summing up, we calculate household income as the total household income from salary, rent, interest receipts, investments, pension benefits, unemployment benefits, child allowances, maternity benefits, apartment allowances, stipends, the value of home produced food (cash and in kind) and other income; excluding depletion of assets, property and jewellery sales, transfers received from friends and relatives, and cash borrowings.

On the expenditures side, the problem is to measure current household consumption. The standard ambiguities arise with real estate, durable goods and physical assets since

¹⁰In 2000, the expenditure section of the household questionnaire was adjusted, including more detailed questions about expenditures on 'health' and 'other services'. Since that year, the 'health' category represents roughly 5% of the total budget, and 'other services' roughly 12%. The share of the original 'services' category remained unaffected, at about 14% of the budget before and after 2000. Hence it does not seem that introducing the new categories has displaced expenditures from the previously existing 'services' category.

¹¹Less than 1% of the participants report receiving any unemployment benefits; we checked and found that the exclusion of this source of income does not have a major influence on the estimates. We also note that in Russia during the 1990s, few benefits were means-tested. Since about 2000 new income-tested benefits have been introduced (e.g., apartment benefits), and some previously existing benefits have become income-tested (e.g., child benefits). Empirical studies suggests that targeting errors are large (Gassmann, 2003; Notten and Gassmann, 2008).

they serve for future as well as current consumption, and may also be used as a means to store value (especially in times of high inflation and exchange rate instability; cf. Lokshin & Ravallion, 2004). A specific problem concerning housing expenditures is that no informed imputations are provided by the RLMS for the value of housing to home owners. Finally, it is a quite common phenomenon for a Russian household not to pay its rent or utility bills; tenants enjoy strong legal protection and the utility infrastructure makes it difficult to cut off single dwelling units (World Bank, 2003). Thus, unpaid bills seem to function as a sustainable way to make ends meet when cash resources fall short (i.e., to relax the liquidity constraint).¹² For these reasons we chose to exclude from our consumption total all expenditures on durables, luxury goods and housing. We calculated two subtotals: food expenditures, including both cash expenditures on foodstuffs and the estimated value of 'autoconsumption' (food consumed and produced at home); and non-food expenditures, which includes expenditures on tobacco, clothing, fuel, services, and meals taken outdoors (considering that these contain a large share of service rather than just food).

Table 5.1: Income and expenditures in the RLMS (1994-2004)

Year	Number of households	Average Expenditures				Average Income			Ratio
		Total	Food	Non-food		Total	Incl.	Excl.	
				Incl.	Excl.				
1994	3586	3964	2344	943	677	3066	2740	326	1.29
1995	3441	3409	1978	970	461	2764	2161	603	1.23
1996	3320	3040	1630	963	447	2588	2066	522	1.17
1997	-	-	-	-	-	-	-	-	-
1998	3426	2240	1231	663	346	1859	1588	271	1.20
1999	-	-	-	-	-	-	-	-	-
2000	3634	2989	1271	838	880	2284	1865	419	1.31
2001	4039	3569	1380	1024	1165	2684	2302	382	1.33
2002	4213	3733	1352	1113	1268	3071	2559	512	1.22
2003	4295	4076	1337	1193	1546	3502	2766	736	1.16
2004	4246	4079	1327	1174	1578	3519	2986	533	1.16

Source: RLMS yearly survey rounds (cross-sections).

Notes: Entries are monthly per capita averages expressed in June 1992 rubles.

Excluded expenditures consists of the following RLMS categories:

durables, luxuries, rent, health, other services, miscellaneous,

bonds, savings, and other payments (insurance, loans).

Excluded income consists of the following RLMS categories:

property sales, insurance benefits, other benefits, help received

from family and friends, and sale of jewellery.

¹²See Freinkman (1998). Notten (2008) provides a more detailed discussion in her Chapter 6, "Managing Risks: What do Russian households do to smooth consumption?"

Table 5.1 presents summary statistics of the income and expenditures totals and subtotals (food and non-food) in per capita values; all amounts are expressed in constant June 1992 ruble prices, and the total of excluded income and expenditure categories are also reported. Clearly, the 1998 crisis had a profound negative impact on household income and expenditures, in all categories. We will discuss Table 5.1 in more detail below in relation to measurement error.

5.3.3 Measurement errors

We have already mentioned the possibility (or rather the practical certainty) of measurement errors in income and expenditures. A first likely source of error is the inclusion of imputations in both income and consumption totals for the value of ‘autoconsommation’. This consists mostly (and in our case entirely) of food that is home-produced and home-consumed. Households report quantities of domestically produced items and these are subsequently valued at local market prices observed at the time of the survey. Since the same imputed values are included in both income and consumption, any imputation errors will cause a bias away from zero in the current income coefficient (hence reversing the ordinary attenuation bias). As suggested by Deaton (1997), we avoid the upward bias by instrumenting the income and lagged expenditures variables with their non-imputed parts. We point out that this intervention will not resolve any attenuation or other endogeneity bias due to non-imputed income.

Previous studies used instrumental variables in order to prevent independent measurement errors (of the traditional variety) biasing tests towards accepting the full insurance hypothesis. Several authors cited earlier used information on wage arrears, pension arrears, forced leave, and unemployment to create instrumental variables for current income, on the premise that these events are associated with unexpected income shocks (Skoufias, 2003; Mu, 2006; Gerry and Li, 2008). We followed the standard GMM logic of exploiting information internal to the panel, and used lagged levels of the imputation-free part of income as instrumental variables for total income changes.¹³

A second measurement error issue concerns inconsistencies due to changes in the survey questionnaires. Some idea of their importance can be gleaned from Table 5.1. Recall that we excluded expenditures on ‘health’ and ‘other services’ from our consumption totals mainly because of such inconsistencies; we also excluded asset sales and borrowings from our income total because of their potential endogeneity. While the value of excluded income sources is rather modest and stable over time, the value of excluded expenses rises from 17% in 1994 to 39% in 2004. Analysis of detailed expenditure categories suggests that this rise can be attributed in large part to the inclusion of new items in the 2001 questionnaire, leading to a higher recall of expenses on health and various services.¹⁴ It is most probable that the 2001 changes led to a substantial improvement in the accuracy of the data, yet such inconsistencies over time are problematic for the estimation of a dynamic model since all increases in expenses are liable to be interpreted as responses to income shocks. Given the magnitude of the increases, the two affected spending categories were excluded from the non-food expenditure total.

A third type of measurement error is one that has plagued household budget surveys ever since that of Ducpétiaux, which Engel analysed in his famous 1857 article: apparent ‘overconsumption’, or underreporting of income. Although not always discussed, let alone explained, this phenomenon seems to be quite common. The total income and expendi-

¹³Guariglia & Kim (2003) argued that the internal instruments should not be considered exogenous.

¹⁴Results not shown here are available from the authors on request.

ture categories in Table 5.1 give an indication of the size of the problem in the RLMS. Average total expenditures are systematically higher than average incomes in every survey round, and the implied negative savings represent up to one third of reported income. Such systematic discrepancy suggests a violation of the household budget constraint. In their studies, Atkinson *et al.* (1995), Deaton (1997), and Ravallion (1994) attribute the problem merely to the tendency of respondents to underreport income from informal or moonlighting activities.¹⁵

There is no detectable coherency between the implied (negative) savings rates and those reported in national income statistics for Russia. Official Goskomstat sources put the saving rate above 20% in the years 1994 to 1997 and fluctuating between 15 and 20% in the years 1998 to 2004. *Russian Economic Trends*, an independent monthly periodical discontinued in 2003, published figures corrected for double counting of hard currency sales that were between 6 and 10 percentage points lower than the Goskomstat rate.

Apart from the underreporting of incomes there are two other straightforward explanations for negative balances between income and consumption. One is that of negative savings by depletion of financial or non-financial reserves, including stocks of durable goods. In view of inflationary pressures, instability in the banking system, and uncertainty about monetary reform, it is possible that durable goods end up being used to store value and play the role of a wealth buffer. The second explanation is that of (increases in) wage arrears, with earnings due but not paid being seen as unspent income, hence a form of forced saving. Both phenomena are likely to be common in transition Russia and can actually be traced in the RLMS data. However, in the RLMS data, both phenomena seem to occur on far too small a scale to explain the extent of ‘overconsumption’.¹⁶

To illustrate the regularity of the problem in our data we graph the logarithmic income and expenditure series for the sequence of survey years in *Figure 2*. Panel 2A presents the evolution of overall averages, Panel 2B the evolution of four representative percentiles (the 20th, 40th, 60th, and 80th) of the frequency distributions of household incomes and expenditures. We interpret the striking parallelism between the two totals as signifying that the extent of income underreporting is relatively constant as a proportion of the actual total. This encourages us to assume that we can treat the implied measurement error in household income as proportionally constant. Its effect will accordingly be absorbed by a constant term (see equation (5.5)).

A different check on the likelihood of measurement errors in incomes and expenditures is done by looking at changes in both quantities. There would be no issue of consumption smoothing if these amounts did not vary, but changes by a factor of 10, 100, or even more, raise suspicions. As a conservative measure, for the estimations reported below, we excluded observations implying income or total expenditure shocks beyond the range (e^{-5}, e^5).¹⁷

The preceding discussion acknowledges that the RLMS income and expenditures data are far from perfect. Nevertheless, they are very seriously managed and constitute by far the best available source for the Russian confederation. One can interpret the fact that they have been used in numerous earlier studies as a confirmation of this.

¹⁵To obtain positive saving rates, Denizer *et al.* (2002) and Foley & Pyle (2005) purge the RLMS data of ‘acute dissavers’, i.e., households that report expenditures exceeding income by more than 50 or 100%, respectively.

¹⁶Gregory *et al.* (1999) revise the official saving rate of 28.7% for 1994 downward and propose a smaller estimate of 12%; this is apparently calculated by imputing wage arrears and counting them as unspent current income in the RLMS.

¹⁷A shock beyond the range (e^{-5}, e^5) means the change factor is smaller than 1/148 or exceeds 148. Depending on the sample considered, this concerns between 1% and 2% of the data.

Figure 5.2: Evolution of income and expenditures (log per capita)

Figure 5.2A: Averages

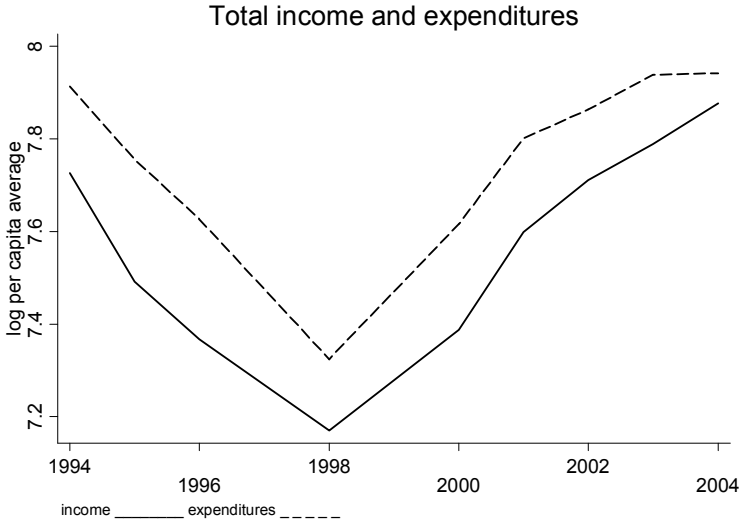
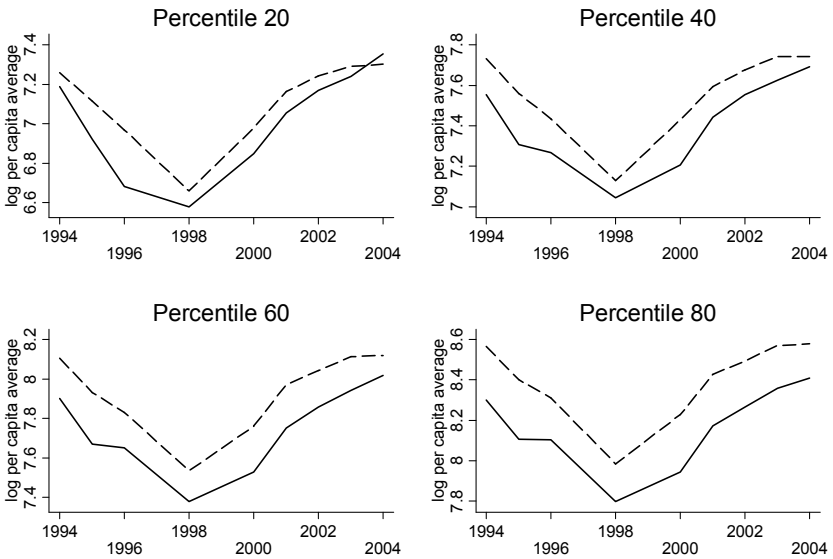


Figure 5.2B: Percentiles



There are reasons to think that the quality of the data is better for the food than the non-food expenditures. In comparison with non-food products, foodstuffs are by nature more perishable. In comparison with clothing and other durables and some services, food is usually purchased at regular intervals, much of it in non-lumpy quantities (except for staples), and it can be used (consumed) only once. An important point is that, in contrast to health and other services, the food items in the questionnaire have remained unchanged over the entire period under consideration. Finally, food is a large part of people's daily needs and evidently the major component of household expenditures with respect to the issue of consumption smoothing. In the sequel, we will therefore focus the discussion primarily, although not exclusively, on food consumption.

5.4 Estimation results and sensitivity analysis

5.4.1 The incomplete panel 1994-2004

In the preceding sections we prepared the ground for the estimation of the consumption equation. In this section we report the key estimates, some tests of the validity of the choices made, and some sensitivity analyses. There is a total of nine RLMS rounds available between 1994 and 2004; there are gaps in 1997 and 1999. We test our model on two different data sets. First, in this subsection, we use the most complete possible panel, comprising nine irregularly spaced waves (1994-1996, 1998, and 2000-2004, with gaps in 1997 and 1999). Exploiting the irregularly spaced data fully is only feasible under stringent random-effects assumptions that are not justifiable here and that we therefore prefer to avoid. Hence the 1998 data play no direct role in the estimates, although they do serve as instrumental variables in the 2SLS and GMM procedures when relevant. In the next subsection, we will see how the results are affected when we use the shorter but regular annual panel for 2000-2004 ($T = 5$). Throughout, we will estimate separate models for food and non-food consumption.¹⁸

Beginning with the extensive data set we present OLS and 2SLS estimates of equations (5.2) and (5.8) for food in Table 5.2; and GMM estimates of equation (5.8) in Table 5.3. For non-food, GMM estimates of equation (5.8) are given in Table 5.4.

Columns [i] to [iv] of Table 5.2 contain estimates of the static model of equation (5.2) applied to food consumption. This is similar to, although distinct from, the model of Skoufias (2003) and Gerry and Li (2008), who estimated equation (5.3). Column [i] reports pooled OLS estimates. The short-run income elasticity of food consumption is positive and statistically significant at any conventional level. A 10% income shock would affect the household's food consumption by a predicted 2.3%. This compares closely with the pooled OLS estimate of Skoufias (2003), which was 2.0%.

We have already pointed out the reasons for treating the income variable $y_{i,t}$ as endogenous. Columns [ii], [iii], and [iv] of Table 5.2 report several versions of instrumental variable (IV or 2SLS) estimates with different instrumentation choices, based on both imputation-free and lagged levels of income. These are essentially variants of the Anderson & Hsiao (1981) estimators. In column [ii] income changes are instrumented with current imputation-free income levels; the effect of a 10% income shock drops to 1.3%. On the face of it this is consistent with OLS containing an upward bias due to imputation errors in both income and consumption.

¹⁸Let us point out that estimating the model twice, for food and non-food expenditures separately, or once, for the aggregate of the two, requires the same strong separability assumptions about the underlying utility functions.

Table 5.2: OLS and IV (2SLS) estimates for food expenditures 1994-2004

Method	OLS	IV	IV	IV	IV	IV
Column	[i]	[ii]	[iii]	[iv]	[v]	[vi]
$\Delta c_{i,t-1}$					0.064** (0.028)	0.063** (0.027)
$\Delta y_{i,t}$	0.230*** (0.009)	0.130*** (0.010)	0.240*** (0.019)	0.074*** (0.027)	0.062 (0.041)	0.077 (0.063)
$\Delta y_{i,t-1}$						0.011 (0.028)
$\Delta x_{1,i,t}$	-0.153*** (0.022)	-0.173*** (0.022)	-0.151*** (0.022)	-0.183*** (0.023)	-0.182*** (0.031)	-0.180*** (0.032)
$\Delta x_{2,i,t}$	-0.162*** (0.019)	-0.177*** (0.019)	-0.161*** (0.019)	-0.186*** (0.020)	-0.164*** (0.027)	-0.162*** (0.027)
$\Delta x_{3,i,t}$	-0.152*** (0.017)	-0.149*** (0.018)	-0.152*** (0.017)	-0.147*** (0.018)	-0.141*** (0.024)	-0.142*** (0.024)
$\Delta x_{4,i,t}$	-0.208*** (0.019)	-0.225*** (0.020)	-0.206*** (0.020)	-0.234*** (0.020)	-0.221*** (0.027)	-0.219*** (0.028)
$\Delta x_{5,i,t}$	-0.165*** (0.040)	-0.159*** (0.040)	-0.165*** (0.040)	-0.156*** (0.040)	-0.156*** (0.054)	-0.157*** (0.054)
$\Delta x_{6,i,t}$	-0.195*** (0.029)	-0.204*** (0.030)	-0.194*** (0.029)	-0.209*** (0.030)	-0.205*** (0.038)	-0.205*** (0.038)
Instrumented:						
$(\Delta c_{i,t-1})$					$\check{c}_{i,t-2}$	$\check{c}_{i,t-2}$
$(\Delta y_{i,t})$		$\check{y}_{i,t}$	$y_{i,t-1}$	$\check{y}_{i,t-1}$	$\check{y}_{i,t-1}$	$\check{y}_{i,t-1}$
$(\Delta y_{i,t-1})$						$\check{y}_{i,t-2}$
IV count		12	12	12	13	14
Obs. count	18668	18668	18668	18668	10939	10939
Hh. count						
FD-AR(1)					0.000	0.000
FD-AR(2)					0.759	0.769
(<i>p</i> -values)						

Notes:

Standard errors in parentheses.

Significance tests: *** means p -value<0.01, ** p -value<0.05, * p -value<0.1.

Year (survey round) dummies included but not reported.

The IVs $\check{y}_{i,t}, \check{y}_{i,t-1}, \check{y}_{i,t-2}$ denote income and its lags net of imputations.The IV $\check{c}_{i,t-2}$ denotes consumption net of imputations, lagged twice.

The IV count also includes the demographic variables and time dummies.

In [v] and [vi], data levels are transformed using forward orthogonal deviations.

FD-AR(1),(2) refer to Arellano-Bond type tests for serial correlation of the first and second order in residual first differences.

Source: RLMS unbalanced panel (1994 - 2004).

In column [iii] of Table 5.2 income changes are instrumented with the lagged total income level (including the imputed component). The lagged income level is an appropriate instrument if income is an accurately measured ‘predetermined’ variable. The instrument allows for endogenous income, in the sense that income changes may contain instantaneous responses to current shocks, but it does not resolve the common imputation error problem. It is not surprising then that, apart from increased standard errors, the instrumental variable estimates in column [iii] hardly differ from the OLS estimates in column [i].

In column [iv] income changes are instrumented with the lagged level of income net of imputed components. This instrument allows for endogeneity in contemporaneous income changes (fast responses to current shocks) in addition to the effect of common imputation errors.¹⁹ The effect of a 10% income shock now drops further to 0.7%. Henceforth, we opt for the last type of instrumental variable.

The last two columns of Table 5.2 augment the consumption equation (5.2) with dynamic terms. Column [v] introduces the lagged first-differenced consumption and Column [vi] introduces in addition the lagged first-differenced income. The resulting estimating equation is a case of equation (5.8) where the transform used is the first-difference transformation; first differences are instrumented with appropriate lagged levels. The short-term income elasticity loses statistical significance although it does not depart in magnitude from the value estimated in Column [iv]. The lagged consumption variable (which is instrumented with a further lag of the imputation-free consumption level) is significant at standard levels in both models; the lagged income variable fails to have a significant effect. Two serial correlation tests (AR1, AR2), due to Arellano & Bond (1991), are calculated from first-differenced data rather than the forward orthogonal deviations used for estimation; they remain in line with expectations, displaying negative first-order autocorrelation but no second-order autocorrelation, and give no indication of dynamic misspecification. However, the inclusion of the dynamics in the model comes at a cost. As more than 40% of the observations are lost, the 2SLS estimator loses precision.

The ECM equation (5.8) is re-estimated in the first two columns of Table 5.3 using the optimal (relatively efficient) GMM method. Column [i] presents the result without and Column [ii] with the lagged income term. A first efficiency gain should come from using forward orthogonal deviations rather than first differences in equation (5.8); this reduces the loss of observations from over 40% to 29%. A second source of efficiency gain is the use of additional instruments; the number of instrumental variables increases from 14 to 35. A third improvement is expected from using an appropriate (theoretically optimal) weighting matrix for the different moment conditions involving the various instruments.²⁰ Both current income and lagged consumption are now statistically significant; even lagged income is marginally significant (at the 10% level) in Column [ii]. Again the autocorrelation tests, due to Arellano & Bond (1991), are based on first-differenced data and remain in line with expectations. Since we use more instruments than the minimum required for identification, the model is overidentified and we can test the validity of the overidentifying restrictions (i.e., moment conditions); the Hansen test indicates no violation and hence does not suggest that suspect instruments were included.

Across estimates, the coefficients of the ‘taste shifters’ $(x_{1,i,t}, \dots, x_{6,i,t})$ vary relatively little. Since consumption expenditures and income are expressed in logarithmic per capita terms, the coefficients of $x_{1,i,t}, \dots, x_{6,i,t}$ can be interpreted roughly as relative changes in per capita expenditures following a modification in household composition. These coefficients are all negative, reflecting economies of scale in the household’s food needs as

¹⁹It would require another lag to resolve classical measurement error.

²⁰The weighting matrix is obtained from a preliminary (first-step) GMM estimation round.

the number of members increases. Most costly to feed (and soak) are adult men, including elderly ones. Somewhat less demanding are adolescent and especially young children. As any stereotype would have it, adult women seem to be the least voracious. Not all these differences are statistically significant but several are.

Table 5.3: GMM (IV) estimates for food consumption expenditures

Sample period	1994-2004		2000-2004	
Column	[i]	[ii]	[iii]	[iv]
$c_{i,t-1}$	0.093*** (0.027)	0.125*** (0.033)	0.074** (0.031)	0.110*** (0.038)
$y_{i,t}$	0.097*** (0.026)	0.151*** (0.041)	0.095** (0.041)	0.221** (0.088)
$y_{i,t-1}$	- -	0.048* (0.029)	- -	0.055 (0.034)
$x_{1,i,t}$	-0.198*** (0.022)	-0.176*** (0.025)	-0.161*** (0.030)	-0.129*** (0.036)
$x_{2,i,t}$	-0.145*** (0.015)	-0.130*** (0.017)	-0.148*** (0.025)	-0.128*** (0.028)
$x_{3,i,t}$	-0.130*** (0.015)	-0.134*** (0.015)	-0.135*** (0.023)	-0.135*** (0.022)
$x_{4,i,t}$	-0.175*** (0.017)	-0.166*** (0.018)	-0.202*** (0.026)	-0.182*** (0.029)
$x_{5,i,t}$	-0.099*** (0.028)	-0.105*** (0.027)	-0.158*** (0.044)	-0.163*** (0.044)
$x_{6,i,t}$	-0.174*** (0.025)	-0.170*** (0.024)	-0.219*** (0.037)	-0.221*** (0.036)
IV count	35	35	19	19
Observation count	13286	13286	8994	8994
Household count	4464	4464	3822	3822
FD-AR(1) p -value	0.000	0.000	0.000	0.000
FD-AR(2) p -value	0.469	0.268	0.661	0.402
Hansen p -value	0.201	0.279	0.345	0.875

Notes: Standard errors in parentheses.

Significance tests: *** means p -value<0.01, ** p -value<0.05, * p -value<0.1.

Year (survey round) dummies included but not reported.

The IV count includes demographic variables and time dummies.

The data levels are transformed using forward orthogonal deviations.

FD-AR(1),(2) refer to Arellano-Bond type tests for serial correlation of the first and second order in residual first differences.

Source: RLMS unbalanced panel (1994 - 2004).

5.4.2 The annual panel 2000-2004

Next we check how sensitive the estimates are to the use of the two-year time intervals. To this end we re-estimate equation (5.8) with and without the lagged income term using only the annual 2000-2004 observations. The results are reported in columns [iii] and [iv] of Table 5.3, following the comparable estimates for the incomplete 1994-2004 panel. Although the results provide some indication that the lower number of observations reduces precision, they remain very similar, and no modification of our substantive conclusions is required. We can conclude that, as far as food consumption is concerned, we detect no obvious misspecification of the dynamic ECM model or violation of the conditions for consistency of the GMM estimator. The estimates seem robust in the sense that they are not very sensitive to the selection of the time interval and to moderately different definitions of the set of instruments used.²¹

In contrast to the findings on food consumption, our estimates using non-food consumption did not pass the specification tests. Table 5.4 reports the GMM estimates for the non-food total. The complete specification produces *negative* income elasticities with relatively large standard errors. Marginal rejections by the Arellano-Bond second-order autocorrelation test (in first-differences) and strong rejections by the Hansen test point towards violations of the model assumptions. Excluding some of the instruments improved the outcome of these tests at the cost of losses of precision in the estimates.

Although the proposed specification was rejected for non-food consumption, we also estimated the model using total consumption expenditures as the dependent variable (results not shown). Depending on the choice of instruments, the null hypotheses of the specification tests were marginally rejected at 5 or 10% significance levels. The point estimates, however, seemed relatively stable across specifications and remained close to those of the food consumption equation. In fact, they were very comparable to the results reported in the previous studies of Russian consumption smoothing cited above.

It emerges that food and non-food consumption behaviour differ in transition Russia, and that the food component dominates in the analysis of (total) consumption smoothing. There is no shortage of potential explanations for differences in observed behaviour. Since many non-food items include an element of durability, one would actually not have expected the dynamics of adjustment to be identical for the two categories of goods. They may enter the household utility functions in asymmetric ways. Add to that differences in the recall period and in the method of recording expenditures and it will be clear that also the measurement error processes are likely to differ.

²¹For obvious reasons, not all results are shown here; more are available from the authors on request. See Roodman (2008) for an interesting discussion on ways of keeping the number of instruments limited without giving up too much information.

Table 5.4: GMM (IV) estimates for non-food consumption expenditures

Sample period	1994-2004		2000-2004	
Column	[v]	[vi]	[vii]	[viii]
$C_{i,t-1}$	0.149*** (0.019)	0.132*** (0.020)	0.123*** (0.026)	0.093*** (0.027)
$Y_{i,t}$	0.113* (0.062)	-0.088 (0.098)	0.103 (0.069)	-0.598*** (0.197)
$Y_{i,t-1}$	- -	-0.211*** (0.078)	- -	-0.346*** (0.089)
$x_{1,i,t}$	-0.050 (0.042)	-0.120** (0.051)	0.019 (0.060)	-0.150* (0.080)
$x_{2,i,t}$	0.133*** (0.032)	0.082** (0.038)	0.186*** (0.045)	0.078 (0.058)
$x_{3,i,t}$	0.128*** (0.035)	0.145*** (0.038)	0.069* (0.040)	0.085* (0.045)
$x_{4,i,t}$	0.064 (0.041)	0.037 (0.044)	-0.032 (0.052)	-0.122* (0.063)
$x_{5,i,t}$	0.044 (0.076)	0.063 (0.080)	-0.096 (0.102)	-0.076 (0.106)
$x_{6,i,t}$	-0.157** (0.063)	-0.169*** (0.065)	-0.106 (0.083)	-0.108 (0.088)
Total IV count	35	35	19	19
Observations count	12759	12759	8673	8673
Households count	4370	4370	3768	3768
FD-AR(1) p -value	0.000	0.000	0.000	0.000
FD-AR(2) p -value	0.041	0.085	0.082	0.464
Hansen p -value	0.000	0.000	0.000	0.000

Notes: Standard errors in parentheses.

Significance tests: *** means p -value < 0.01, ** p -value < 0.05, * p -value < 0.1.

Year (survey round) dummies included but not reported.

The IV count includes demographic variables and time dummies.

The data levels are transformed using forward orthogonal deviations.

FD-AR(1),(2) refer to Arellano-Bond type tests for serial correlation of the first and second order in residual first differences.

Source: RLMS unbalanced panel (1994 - 2004).

5.4.3 Further interpretations

As far as food consumption is concerned, the short-run effects in Table 5.2 are broadly consistent with the findings of Skoufias (2003), Mu (2006), and Gerry and Li (2008). The interpretation of the short-run income elasticity (β_1) is the same in the static and in the dynamic models; it measures the instantaneous adjustment of consumption to income shocks. The full insurance hypothesis ($\beta_1 = 0$) is rejected in each model, whatever the estimation method used. The estimates also confirm that current income shocks are not fully transmitted to current consumption ($\beta_1 < 1$); households find ways to smooth their food consumption, or at least a large proportion of it.

At the same time, our findings suggest that the point estimates are sensitive both to the correction for imputation errors, and to the introduction of long-term dynamics. Controlling for imputation errors by the instrumental variable technique yields a substantially lower short-run income elasticity in the static model (viz., a reduction from 2.3% to 0.7% for a 10% income change). Introducing long-term dynamics lets the short-run elasticity come back half-way (up to 1.5% in the complete specification).

The other parameters are also of interest in this relationship. The coefficient of the equilibrium correction term in equation (5.5), $[-(1 - \alpha)]$, should be negative as households need to compensate for under- or overconsumption in the previous period, but not exceed 1 in absolute value if systematic overshooting is to be avoided. Hence, the meaningful values of α are in the range $0 \leq \alpha < 1$. The estimated values of α are significantly larger than zero but in a range close to 0 and far from 1, namely 0.0625 to 0.125, confirming that delayed adjustments are important.

For regular values of the coefficients (as found here), the mean delay of adjustment ($\bar{\ell}_y$) was defined by formula (5.7). Our estimated values (in years) of $\bar{\ell}_y$ are 0.386 (a little over $4\frac{1}{2}$ months) in the incomplete 1994-2004 panel, and 0.322 (a little under 4 months) in the annual 2000-2004 panel. These are not long respite periods!

The long-term income elasticity of food consumption was defined in equation (5.6). This is estimated at 0.228 in Column [ii] and 0.310 in Column [iv] of Table 5.3. These estimates are significantly larger than 0 and, more surprisingly, they are far smaller than unity (both substantively and statistically). They are, nonetheless, somewhat higher than their short-term counterparts, supporting perhaps the proposition that it becomes more difficult to ‘smooth’ consumption over the longer term.

The fact that we find long-term income elasticities far below unity means that we fail to reconcile the panel evidence with the cross-sectional and aggregate time-series evidence. If we construct a cross-section from the RLMS using ten-year differences, then the relationship between consumption expenditures and income is roughly proportional. Cross-sections based on short periods of one or two years exhibit elasticities somewhat lower than 1 but still strong. Similarly in the time-series dimension, the cross-sectional averages (or aggregates) exhibit a unitary long-term elasticity and lower short-term elasticities. But even short-term elasticities are a multiple of those found in the panel estimates.²²

To the extent that the hypothesis $H_0 : \beta_2 = 0$ holds, there would be no difference between the ECM model of equation (5.4) or (5.5) and a classical partial adjustment model. In any case, we have to conclude that the static specification of equation (5.2) is dynamically incomplete for food consumption, and this comment is bound to apply to the Skoufias specification in equation (5.3) as well.

²²There is a similar puzzle concerning the time-series properties of income and consumption. In the aggregate, the series seem to have an autoregressive coefficient indistinguishable from unity. In our panel estimates, this is far from being the case.

In the case of non-food consumption, the situation is worse. Even the generality of the ECM is insufficient to capture the dynamics in the data. The problem may go unnoticed in a static equation for total consumption, especially when food expenditures dominate the budget. This result casts doubt on efforts to model the aggregate of food and non-food or consumable and durable expenditures, and the conclusions drawn from them.

It is important to keep in mind that the estimates reflect the outcome of a mix of consumption smoothing strategies, including all applicable combinations of self-insurance plans and risk sharing arrangements. Furthermore, as we pointed out in the introduction, the quality and the workings of the existing financial institutions constrain the choice of plans and arrangements that households can make. Given differences in households's resources and their environment, it is likely that they will not be able to manage risks and absorb shocks exactly to the same extent. Hence in the next section we explore differences in smoothing responses across households.

5.5 Heterogeneity in smoothing behaviour

To investigate differences in smoothing responses across households we follow an exploratory approach. We estimate the food consumption ECM separately for various groups of households and compare the results across groups; in this way we also test whether the ECM specification is able to accommodate smoothing responses from households that find themselves in rather different circumstances.²³ The splits in the sample of households are guided by a set of indicators each of which is arguably associated with enhanced or reduced smoothing abilities: (i) the type of income shock (positive or negative), (ii) the environment (rural or urban area), (iii) average standard of living (low, middle, high), and (iv) previous experience of poverty spells (chronically poor, occasionally poor, never poor); the last two groupings will be crossed.

In a standard life cycle model with perfectly functioning financial markets, there is no reason for households to respond differently to positive or negative (unanticipated) income shocks. However, if access to credit or insurance is restricted, the urgency of adjusting to positive or negative shocks may no longer be the same. Households with no or limited access to risk management institutions may be forced to cut food expenditures faster in the face of an adverse shock than they will increase expenses after a windfall gain. Given the dramatic structural changes that characterize the economic transition in Russia, involving deep reforms of the financial markets, the government, the social protection system, and labour relations, one would expect that many Russian households did face restricted access to credit and insurance facilities. We test this proposition by separately estimating the ECM for households experiencing positive shocks and those experiencing negative shocks; the results are summarized in Table 5.5. We find the short-run income elasticity is slightly higher for negative income shocks (1.4% for a 10% income shock) than it is for positive income shocks (1.3% for a 10% income shock) but the difference is not statistically significant. The reverse holds for the long-run income elasticity, which is a little lower for negative shocks (2.1%) than for positive shocks (2.3%). The speed of correction coefficient ($1 - \alpha$) on the error correction term is also somewhat higher for negative shocks. Thus the estimates suggest that households make somewhat faster and larger adjustments in their food expenses after a negative income shock but the result is far from conclusive as the confidence intervals of the estimated parameters largely overlap.²⁴

²³ An alternative approach is to introduce dummy variables and interaction terms; see, for an example, Mu (2006).

²⁴ A joint test will assess the differences between the three main coefficients of interest in Table 5.5.

Table 5.5: Overall and split-sample estimation by sign of shock

	All shocks	Positive shocks	Negative shocks
$c_{i,t-1}$	0.125*** (0.0329)	0.136*** (0.0427)	0.129*** (0.0376)
$y_{i,t}$	0.151*** (0.0414)	0.128** (0.0533)	0.141*** (0.0333)
$y_{i,t-1}$	0.048* (0.0289)	0.070* (0.0403)	0.040 (0.0318)
IV count	35	35	35
Observations	13286	9082	6380
Households	4464	3984	3444
FD-AR(1) p -value	0.000	0.000	0.000
FD-AR(2) p -value	0.268	0.233	0.416
Hansen p -value	0.279	0.299	0.388

Notes: Standard errors in parentheses.

Significance tests: *** means p -value < 0.01, ** p -value < 0.05, * p -value < 0.1.

Demographic variables and year dummies included in all estimations.

The IV count includes demographic variables and year dummies.

The data levels are transformed using forward orthogonal deviations.

FD-AR(1),(2) refer to Arellano-Bond type tests for serial correlation of the first and second order in residual first differences.

Source: RLMS unbalanced panel (1994 - 2004)

The second proposition of heterogeneity we check is whether consumption smoothing behaviour differs between rural and urban households. Given the large differences of development between rural and urban areas in Russia, households's access to and use of risk management institutions may be very different. Skoufias (2003) found that rural households have lower smoothing abilities than urban households, but he was concerned that these findings may have been driven by measurement error. Mu (2006) presented evidence suggesting that different factors determine smoothing abilities in rural and urban areas. While in urban areas higher consumption smoothing abilities are associated with higher levels of education (but not with wealth), in rural areas higher consumption smoothing abilities are associated with higher wealth (but not with education). However, neither Skoufias nor Mu controlled for measurement error in home food production, and although home food production is important for both rural and urban households in Russia (in 2003 half of the urban households reported using land for food production), the prevalence and importance of home food production is much larger for rural households (Notten, 2008 pp. 173-175). Table 5.6 reports our split-sample estimates for rural and urban households. For the urban sample, lagged income is not significant but the error correction coefficient is close to one (0.926). In the rural model the long-term income elasticity is significant and rather high with 3.7%, while the error correction coefficient is 0.813. The short-term income elasticity is considerably higher for rural households (2.0%) than for urban households (1.1%), suggesting that rural households reduce food expenses

more strongly in response to current income shocks. Standard 95% confidence intervals of the parameter estimates for the two subsamples overlap slightly.²⁵ The rural-urban split is the most common one investigated in previous studies, even though the splitting line is not incontrovertible and remains a topic of research.²⁶

Table 5.6: Overall and split-sample estimation by type of area

	All areas	Urban areas	Rural areas
$c_{i,t-1}$	0.125*** (0.033)	0.074** (0.038)	0.187*** (0.050)
$y_{i,t}$	0.151*** (0.041)	0.106** (0.053)	0.204*** (0.056)
$y_{i,t-1}$	0.048* (0.029)	0.001 (0.041)	0.094** (0.041)
IV count	35	35	35
Observations count	13286	8699	4587
Households count	4464	3046	1418
FD-AR(1) p -value	0.000	0.000	0.000
FD-AR(2) p -value	0.268	0.324	0.571
Hansen p -value	0.279	0.820	0.429

Notes: Standard errors in parentheses.

Significance tests: *** means p -value<0.01, ** p -value<0.05, * p -value<0.1.

Demographic variables and year dummies included in all estimations.

The IV count includes demographic variables and year dummies.

The data levels are transformed using forward orthogonal deviations.

FD-AR(1),(2) refer to Arellano-Bond type tests for serial correlation of the first and second order in residual first differences.

Source: RLMS unbalanced panel (1994 - 2004)

Finally we will try and assess whether households with a lower living standard or having previously experienced one or more episodes of poverty have different ways of adjusting expenditures. Even though we focus on food, it is not entirely clear *a priori* what relationship to expect. On the one hand, households with a lower living standard would tend to have less resources to manage risks and more restricted access to risk management institutions; one would therefore expect them to respond more strongly to income shocks. On the other hand, it is sometimes argued that well-to-do households have more means to economise in planning, buying, conserving and preparing food, while unequipped households are at the mercy of a local and irregular supply.

We attempt a split of the sample of households according to their living standard (low, middle or high) and their previous experience of poverty spells (chronically poor, occasionally poor, never poor). Both criteria are derived from the history of actual household

²⁵ A joint test will assess the differences between the three main coefficients of interest in Table 5.6.

²⁶ We use the RLMS definition of *urban* and aggregate all other areas (including the ‘poselki gorodskogo tipa’ or ‘peri-urban’ areas) in a *rural* (or, strictly, ‘non-urban’) category.

expenditures.²⁷ To construct a proxy for living standard, a series of ‘household-equivalent expenditures’ is generated as the ratio of actual household expenditures to the household-specific poverty line. The resulting ratio is then averaged over time. Thus, a time average below 1 indicates that a household lived in poverty during the transition period; and conversely for an average above 1. Then we group households into low (ratio below 1.5), middle (ratio between 1.5 and 2.5) or high (ratio of 2.5 or higher) living standard categories. The proportion of households in each category is roughly one third; from low to high, 30%, 37% and 33%, respectively.

Next we divide households into three categories: frequently poor, occasionally poor, and never poor. Frequently poor households are households observed to fall below the poverty line in at least three survey rounds (22%); occasionally poor households experienced one or two spells of poverty (36%); never poor households remained above the poverty line in all the survey rounds in which they participated (43%).²⁸

The criteria of living standard and poverty frequency are then combined or ‘crossed’ into six subgroups: frequently poor (1); occasionally poor by low (2), medium (3), and high (4) living standard; and never poor by medium (5) and high (6) living standard.

The results are summarized in Table 5.7. Only in the three largest of the six subgroups, one or more of the estimated adjustment parameters are statistically significant: frequently poor (1); occasionally poor, medium standard (3); never poor, high standard (6). We attribute this result to the combination of a computationally demanding estimator (GMM) with relatively low numbers of observations in the subsamples. Focusing only on the estimates for the larger subgroups, the results suggest that there is not much difference in food consumption smoothing behaviour between frequently poor and occasionally poor, middle standard households.²⁹ High standard, never poor households tend to make larger short term and long term adjustments in response to income shocks; this result stood when we split the sample in only three living standard categories (results not shown here). A higher income elasticity for high standard households is consistent with the explanation that well-to-do households have more means to economise on food expenditures than unequipped households. Again, the wide and overlapping confidence intervals of the estimated income parameters indicate that these results should not be taken at face value.

²⁷We disregard the implied endogenous selectivity on the grounds that its effects are likely to be minor. Only expenditures on non-durable goods are taken into account in these calculations.

²⁸This classification is chosen to reflect a spells-based concept of poverty (Hulme & Shepherd, 2003). The ‘frequently poor’ category is comparable to the indicator for ‘at risk of persistent poverty’ proposed by Atkinson *et al.* (2002). Using the RLMS, Luttmer (2001) finds that one fifth of the poor in Russia are erroneously classified as a consequence of transitory shocks and noisy data.

²⁹A joint test will assess the differences between the three main coefficients of interest in Table 5.7.

Table 5.7: Overall and split-sample estimation by poverty experience and living standard

	All		Frequently poor		Occasionally poor & Usual living standard			Never poor & Usual living standard		
			low	medium	high	medium	high	medium	high	
$c_{i,t-1}$	0.125*** (0.033)	0.078 (0.049)	-0.010 (0.083)	0.006 (0.061)	0.004 (0.086)	0.045 (0.109)	0.303*** (0.082)			
$y_{i,t}$	0.151*** (0.041)	0.146** (0.074)	-0.001 (0.138)	0.152** (0.077)	0.141 (0.135)	0.075 (0.091)	0.263*** (0.083)			
$y_{i,t-1}$	0.048* (0.029)	0.060 (0.055)	-0.048 (0.088)	0.009 (0.055)	0.060 (0.103)	-0.013 (0.070)	0.132*** (0.050)			
Instrument count	35	35	35	35	35	35	35			
Observations	13286	2703	1086	2592	1043	1818	3124			
Households	4464	784	470	767	306	663	1142			
FD-AR(1) p -value	0.000	0.000	0.000	0.000	0.000	0.001	0.000			
FD-AR(2) p -value	0.268	0.414	0.442	0.840	0.471	0.321	0.169			
Hansen p -value	0.279	0.502	0.154	0.681	0.183	0.280	0.251			

Notes: Standard errors in parentheses. Significance tests: *** means p -value < 0.01, ** p -value < 0.05, * p -value < 0.1.

Demographic variables and year dummies included in all estimations.

The IV count includes demographic variables and year dummies.

The data levels are transformed using forward orthogonal deviations.

FD-AR(1),(2) refer to Arellano-Bond-type tests for serial correlation of 1st and 2nd order in residual first differences.

Source: RLMS unbalanced panel (1994 - 2004)

5.6 Conclusion

In comparison to the models typically used in the empirical literature on consumption smoothing, this chapter developed and tested a more flexible (ECM) model of dynamic consumption smoothing behaviour. The ECM model distinguishes between immediate and delayed adjustments of consumption expenditures to income shocks and thereby takes into account that income and consumption can temporarily deviate even though they should balance in the long run. We tested and rejected the full insurance hypothesis, i.e., that current consumption is entirely detached from contemporary income shocks. The inevitability of a lifetime budget constraint suggests the necessity of a long-term connection between consumption and income. We therefore also investigated whether observed behaviour is consistent with a tendency of consumption and income to return to a long term equilibrium after income shocks. We failed however to detect the implied proportionality relationship between consumption and income at the household level in the RLMS data, and thus to reconcile the panel evidence with stylised facts from cross-sectional or aggregate time-series analyses.

Undoubtedly more research will be required to find out exactly how complex the dynamics of adjustment need to be; and why the implied long-run income elasticity of (food and non-food) consumption expenditures is so much smaller than unity.

In our empirical estimation strategy we compared a range of short and long term models, using several estimation techniques and making various adjustments to reduce the impact of certain types of measurement error on our estimates. To take these factors simultaneously into account we used a state of the art instrumental variable estimation technique, the Generalized Methods of Moments. The ECM model was subjected to a number of specification and robustness tests. Our results broadly confirmed the findings of previous studies on consumption smoothing behaviour in Russia, mainly the rejection of the full insurance hypothesis. However, they also show that dynamics are necessary in the consumption equation, and that smoothing estimates are sensitive to the measurement error caused by imputations in home food production. The heterogeneity tests in the last section of the chapter provide some limited indication that consumption smoothing abilities may vary according the type of shock (positive or negative), the level of urbanisation, standards of living and previous experience of poverty spells. However, small differences in the point estimates and overlapping confidence intervals prevent strong claims with respect to heterogeneity in smoothing behaviour.

Concluding, the results presented in this chapter very well illustrate that the task of gaining a better understanding of the empirical realities of consumption behaviour is highly complex and that, despite the many recent advances in the availability of data and estimation techniques, our power to answer deep questions is still limited.

References

- Altonji J.G., F. Hayashi & L.J. Kotlikoff (1992): "Is the extended family altruistically linked? Direct tests using micro data". *The American Economic Review* 82 (5) 1177-1198.
- Anderson T.W. & C. Hsiao (1981): "Estimation of Dynamic Models with Error Components". *Journal of the American Statistical Association* 76 (1981) 598-606.

- Arellano M. & S. Bond (1991): "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations". *Review of Economic Studies* 58 (2) 277-297.
- Arellano M. & O. Bover (1995): "Another Look at the Instrumental Variable Estimation of Error-Component Models". *Journal of Econometrics* 68, 29-51.
- Atkinson A. B., Cantillon, B., Marlier, E. & Nolan, B. (2002). Social indicators: the EU and social inclusion. Oxford; New York: Oxford University Press.
- Atkinson A. B., L. Rainwater & T.M. Smeeding (1995). *Income distribution in OECD countries: the evidence from the Luxembourg Income Study (LIS)*. Paris: Organisation for Economic Co-operation and Development.
- Banerjee A.V. & E. Duflo (2007): "The economic lives of the poor". *Journal of Economic Perspectives* 21 (1) 141-167
- Blundell R. & S. Bond (1998): "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models". *Journal of Econometrics* 87 (1998) 115-143.
- Brown A. (1999): "The Russian crisis: Beginning of the end or end of the beginning?" *Post-Soviet Affairs* 15 (1) 56.
- Browning M. & T.F. Crossley (2001): "The life-cycle model of consumption and saving". *The Journal of Economic Perspectives* 15 (3) 3-22.
- Buchs T. D. (1999): "Financial crisis in the Russian Federation". *Economics of Transition* 7 (3) 687.
- Cochrane J.H. (1991): "A simple test of consumption insurance". *The Journal of Political Economy* 99 (5) 957-976.
- Davidson J.E.H., D.F. Hendry, F. Srba & S. Yeo (1978): "Econometric modelling of the aggregate time-series relationship between consumers' expenditure and income in the United Kingdom". *Economic Journal* 88, 661-692.
- Deaton A. (1992): *Understanding Consumption*, Clarendon Lectures in Economics. Oxford University Press, Clarendon Press.
- Deaton A. (1997): *The analysis of household surveys: A microeconomic approach to development policy*. Johns Hopkins University Press for the World Bank.
- Denizer C., H.C. Wolf & Y. Ying (2002): "Household saving in the Transition". *Journal of Comparative Economics* 30 (2002) 463-475.
- Engel E. (1857): "Die Productions- und Consumptionsverhältnisse des Königreichs Sachsen". *Zeitschrift des Statistischen Büreaus des Königlich Sächsischen Ministerium des Innern*, No. 8 u 9, 1-54. Reprinted as an appendix to "Die Lebenskosten Belgischer Arbeiter-Familien früher und jetzt" in *Bulletin de l'Institut International de la Statistique IX* (1895), 1e livraison Anlage 1, 1-54.
- Engle R.F. & C.W.J. Granger (1987): "Cointegration and Error Correction: Representation, Estimation and Testing". *Econometrica* 55, 251-276.
- Fafchamps M. & S. Lund (2003): "Risk-Sharing Networks in Rural Philippines". *Journal of Development Economics* 71 (2) 261-287.
- Foster J., J. Greer & E. Thorbecke (1984): "A Class of Decomposable Poverty Measures". *Econometrica* 52 (3) 761-766.
- Freinkman L. (1998): "Russian Federation Housing and Utility Services: Policy Priorities for the Next Stage of Reforms". Washington, DC: World Bank, Report No. 17483-RU.
- Friebel G. & S.M. Guriev (2005): "Attaching Workers Through In-kind Payments: Theory and Evidence from Russia". *World Bank Economic Review* 19 (2) 175-202.

- Friedman M. (1957): *A theory of the consumption function*. Princeton University Press.
- Gassmann F. (2003): *Review of Targeted Social Assistance in Russia; Description, Analysis and Recommendations*. Gassmann Consulting, Eindhoven.
- Gerry C.J. & C.A. Li (2008): Consumption smoothing and vulnerability in Russia. *Applied Economics*, iFirst 05 Dec 2008, doi: 10.1080/00036840701765403.
- Gregory P., M. Mokhtari & W. Schrettl (1999): "Do Russians Really Save That Much? Alternate Estimates from the Russian Longitudinal Monitoring Survey", *The Review of Economics and Statistics* 81 (1999) 694-703.
- Guariglia A. & B.Y. Kim (2003): "Wage arrears uncertainty and precautionary saving in Russia", *The Economics of Transition* 11 (3) 493-512.
- Hall R. (1978): "Stochastic Implications of the Life Cycle-Permanent Income Hypothesis: Theory and Evidence", *Journal of Political Economy* 86 (1978) 971-987.
- Heeringa S.G. (1997): "*Russia Longitudinal Monitoring Survey (RLMS), Sample Attrition, Replenishment, and Weighting in Rounds V-VII*". Survey Design and Analysis Unit, Institute for Social Research, University of Michigan, Ann Arbor, MI 48106-1248, March 1997.
- Hulme D. & A. Shepherd (2003): Conceptualizing chronic poverty. *World Development*, 31 (1) 403-423.
- Lokshin M. & M. Ravallion (2004): "Household Income Dynamics in Two Transition Economies". *Studies in Nonlinear Dynamics and Econometrics* 8 (3), article 4, 1-31.
- Luttmer E. (2001): "Measuring Transition Dynamics and Inequality in Transition Economies". World Bank Policy Research Working Paper No. 2549.
- Mace B.J. (1991): "Full insurance in the presence of aggregate uncertainty". *The Journal of Political Economy* 99 (5) 928-957.
- Mroz T.A., L. Henderson & B.M. Popkin (2005): *Monitoring Economic Conditions in the Russian Federation: The Russia Longitudinal Monitoring Survey 1992-2004*. Report submitted to the U.S. Agency for International Development, Carolina Population Center, University of North Carolina at Chapel Hill, North Carolina. May 2005.
- Mu R. (2006): "Income Shocks, Consumption, Wealth, and Human Capital: Evidence from Russia". *Economic Development and Cultural Change* 55 (2006) 857-892.
- Notten G. (2008): *Measuring and Managing Poverty Risks*. Dissertation, Boekenplan. (Chapter 5: "Poverty and consumption smoothing in Russia"; Chapter 6: "Managing Risks: What do Russian households do to smooth consumption?") (Available on URL: <http://arno.unimaas.nl/show.cgi?fid=11319>)
- Notten G. & D. de Crombrugghe (2006): "Poverty and consumption smoothing in Russia". Working paper, MGSOG/2006/WP004, Maastricht Graduate School of Governance, Maastricht University (URL: <http://www.governance.unimaas.nl/home/publications/2006WP004.pdf>)
- Notten G. & D. de Crombrugghe (2007): "Poverty risk and consumption smoothing abilities in Russia". MPRA Paper No. 5314, 15 Oct. 2007.
- Notten G. & F. Gassmann (2008): "Size matters: poverty reduction effects of means-tested and universal child benefits in Russia". *Journal of European Social Policy*, 18 (3) 260-274.
- Ravallion M. (1994). *Poverty Comparisons Vol. 56*. Harwood Academic Press.

- Ravallion M. & S. Chaudhuri (1997): "Risk and Insurance in Village India: Comment". *Econometrica* 65 (1) 171-184.
- Roodman D. (2008): "A Note on the Theme of Too Many Instruments". *Oxford Bulletin of Economics and Statistics* 71 (1) 135-158.
- Russian Economic Trends, 1998-2003. (Various issues; discontinued in 2003)
- Sapir J. (1999): "Russia's Crash of August 1998: Diagnosis and Prescription". *Post-Soviet Affairs* 15 (1) 1-36.
- Skoufias E. (2003): "Consumption smoothing in Russia: Evidence from the RLMS". *Economics of Transition* 11 (1) 67-91.
- Slay B. (1999): An Interpretation of the Russian Financial Crisis. *Post-Soviet Geography and Economics* 40 (3) 206-214.
- Stillman S. (2001): "The response of consumption in Russian Households to economic shocks." IZA Discussion Paper No. 411, Institute for the Study of Labor, Bonn.
- Townsend R.M. (1994): "Risk and Insurance in Village India". *Econometrica* 62 (3) 539-591.
- World Bank (2003): *Housing and Communal Services in Russia: Completing the Transition to a Market Economy, Final Report*. The World Bank Infrastructure and Energy Department, Europe and Central Asia Region.
(http://siteresources.worldbank.org/INTRUSSIANFEDERATION/Resources/305499-1094736798511/518266-1095320717197/housing_eng.pdf)
- World Bank (2004): *Russian Federation: Poverty Assessment*. 28923-RU, Poverty Reduction and Economic Management Unit, Europe and Central Asia Region, Washington D.C.: The World Bank.
- Zohoori N., T.A. Mroz, B. Popkin, E. Glinskaya, M. Lokshin, D. Mancini, P. Kozyreva, M. Kosolapov & M. Swafford (1998): "Monitoring the Economic Transition in the Russian Federation and its Implications for the Demographic Crisis - the Russian Longitudinal Monitoring Survey". *World Development* 26 (11) 1977-1993.

Chapter 6

Rule of law and the lure of growth: A panel data re-examination¹

Abstract

Recent shifts in the way international investors and aid donors determine their priorities in the developing world has stimulated interest in the governance-growth nexus. Indicators measuring various aspects of the quality of state governance have been elaborated, a fundamental one being the maintenance of the rule of law. In this chapter we compare several well-established sources of rule of law indicators, and analyse their contribution to explaining economic growth over the last decades in a worldwide panel of countries. We assess the strength of the statistical relationship and its robustness with respect to different choices regarding the source of the indicator, the composition of the sample, the details of the specification, or the treatment of heterogeneity. It turns out that all these choices matter. We can find no solid empirical basis for linking the economic growth in a country (and hence the effectiveness of investments or aid funds) to its performance in terms of the most popular rule of law indicators.

6.1 Introduction

Since the late 1990s, the focus of development policies has moved away from purely macro-economic policy concerns and towards improving the quality of state governance, not least because of a growing perception that the costly stabilisation policies of the two preceding decades had failed to deliver growth. The World Bank, for instance, has placed an increasing emphasis on policy performance and made the improvement of public sector governance one of the main themes of its lending portfolio as well as its nonlending services.² The theoretical foundation for the new development practice was provided by the insights of North (1990, 2005) and the ‘New Institutional Economics’ he spearheaded, which made evident the importance of a country’s system of governance for its success in terms of long-term economic growth, enhancement of human welfare, and societal development.

A parallel development is the leap made by both foreign direct investment and foreign portfolio investment in emerging market economies, with about a twentyfold increase in

¹This chapter was co-authored with Christiane Arndt (OECD). An earlier version appeared as Chapter 6 in Arndt (2009). We thank Orsetta Causa for her constructive comments.

²Successive World Bank Annual Reports document the resulting concentration of lending in ‘countries with sound policies where IDA resources can be used most effectively’. See also Arndt (2009).

annual levels over the last two decades. According to the OECD, differences among developing countries' perceived credibility in policy implementation, and above all in the quality of their systems of governance, have emerged as the single most important determinant of their investment location decisions.³

The newfound interest in the quality of state governance in developing countries has spurred the demand for 'objective' performance indicators that could help to inform donors, lenders and investors. Some were pre-existing, like the Fraser Institute's 'Economic Freedom Index' and the Policy Risk Services (PRS) Group's 'International Country Risk Guide' (ICRG). New ones were created in the 1990s, like the Heritage Foundation's Index of Economic Freedom and the World Bank Institute's 'Worldwide Governance Indicators' (WGI). Nowadays, a significant proportion of the fund allocation decisions by donors, lenders and investors seems to be influenced by the performance and rankings of the destination countries on those indicators.⁴

The question whether there is a verifiable link between good governance as measured by these indicators and outcomes as measured by economic growth or prosperity is therefore of great practical and political importance. Not only are the indicators influential in determining the current access to development funds, credit and foreign investment, they are also likely to be used as objective arguments in any negotiation about the future great policy orientations on the international scene. This motivates our interest in the relation between publicly available governance indicators and actual economic outcomes.

The aim of this chapter is to investigate the existence and solidity of the link between published governance indicators and the resulting economic growth. However, good governance is a gem with multiple facets, and each source presents its own carving of it. Previous studies have often emphasised one aspect of good governance that is thought to be fundamental in creating the right conditions for development and the adoption of new technologies, namely, the observance and maintenance of the rule of law.⁵ There is sufficient consensus about the importance of the concept for it to be found in multiple sources in comparable forms.

Yet 'rule of law' is itself a multi-sided idea, and to exploit the different sources we will need to specialise it further. In current market-based thinking, the main (though certainly not the only) economic implication of the rule of law concerns legal security and limited transaction costs for private investors. Hence we searched the available sources for (sub)indicators designed to proxy the observance and maintenance of the rule of law relevant for the private sector, including mainly the protection of property rights and contract enforcement. This is not to deny the primacy of human rights, or the part that the rule of law plays in creating a just and fair society; however these are objectives in their own right, and may be more remotely connected to growth.⁶

³See Oman, 2000; OECD, 2002. Foreign investment data are from World Development Indicators Online 2008 (World Bank 2008).

⁴See, e.g., Arndt & Oman (2006). Several more indicators were created by or under the auspices of the World Bank, but are not available to us with sufficient historical depth to be useful here, like the Country Policy and Institutional Assessment (CPIA, see IDA, 2007); and the 'Doing Business Indicators' based on the Business Environment and Enterprise Performance Surveys Initiative ([©]2002-2004).

⁵See, for instance, Knack & Keefer (1995), Kaufman & Kraay (2002), and Rigobon & Rodrik (2004). According to The Economist (March 13th, 2008), rule of law has become a staple of economic thinking and "the motherhood and apple pie of development economics". North (1990) initially emphasised the importance of solid formal and informal institutions to reduce transaction and enforcement costs. Dam (2006) examined the underlying mechanisms through which the law, the judiciary and the legal profession influence the economy.

⁶Rodrik (2007), on rule of law: "Am I the only economist guilty of using the term without having a good fix on what it really means?" Our use of the term 'rule of law' corresponds to what Trebilcock &

Having determined a short-list of rule of law indicators, we seek to determine to what extent those indicators can contribute (if at all) to the explanation of economic growth over the last decades, in the largest possible set of countries. We use panel data methods and propose solutions to a number of specification problems. In order to prevent superficial conclusions to be drawn, our growth equations will be submitted to tests of robustness with respect to different choices regarding the source of the indicator, the composition of the sample, the details of the specification, and the treatment of heterogeneity.

The literature on the economic, technological and institutional determinants of growth is large and growing fast. North's seminal work stimulated a number of statistical studies, which tended to confirm the existence of the positive effect of good governance on economic prosperity and the effectiveness of aid; perhaps the most influential of those was Burnside & Dollar (2000). For a time after the publication of their results, one could believe that a consensus was emerging on the issue.⁷ Yet, there were also criticisms, and perusing the recent empirical literature, it is hard to avoid the impression that the governance-growth link still is a subject of ongoing debate rather than an established fact.⁸ We tried to zero in on some of the problems standing in the way of a clear conclusion to the debate.

One major difficulty in identifying the effect of improvements in governance on economic development is their likely endogeneity, which might be the consequence of simultaneity, measurement error, omitted variables, and even selection bias. Mauro (1995) attempted to correct for reverse causality by constructing exogenous instrumental variables for his governance index, namely an index of ethnic fractionalization and a set of colonial heritage dummies. The exogeneity of these instruments was soon questioned (see Easterly and Levine, 1997; Zak and Knack, 2001). Acemoglu *et al.* (2001) used settler mortality as an instrument for the quality of institutions, and inferred the existence of a strong positive impact of institutions on long-term economic growth. Some other studies adopted this instrument (e.g. Kaufmann and Kraay, 2002, Rodrik *et al.*, 2004), still others criticised it. Sachs argued that institutions are almost entirely endogenous to growth (see Sachs and McArthur, 2001; Sachs and Warner, 1995). Glaeser *et al.* (2004) demonstrated that settler mortality may simply capture the human capital of the European colonists, rather than early institutional strengths. Albouy (2006) found that Acemoglu *et alii*'s results are highly sensitive to the choice of the source of information on settler mortality, with multiple mortality estimates being available for a number of countries. Rigobon and Rodrik (2005) avoided the delicate search for instrumental variables altogether and instead ingeniously achieved identification through heteroskedasticity. They, too, concluded that the rule of law has a strong positive effect on incomes.

Those attempts at solving the identification problem are predominantly data-driven. This suggests that the fundamental problem is the 'open-endedness of growth theories', as Brock and Durlauf (2001) put it. Various competing models of growth, development and the effectiveness of aid are circulating, and there exist so many mutually compatible theoretical arguments that it is very hard to justify exclusion restrictions. Hence, practically any candidate instrument is likely to be correlated with potentially relevant omitted variables, thereby violating the exogeneity requirement.

Daniels (2008) call its 'thin' conception. They describe it as focussing on property rights and the efficient administration of justice in order to provide for stability. A more general, 'thicker' concept of rule of law would also encompass physical, legal and political security, linking it to civil liberties and democracy – which is not the focus of the present study. Thomas Carothers (2003) from the Carnegie Endowment for International Peace laments the lack of knowledge at many levels of conception, operation and evaluation surrounding the expanding "rule-of-law-promotion" in developing countries.

⁷See, e.g., Kaufman & Kraay (2002) and Rodrik *et al.* (2004).

⁸Compare, for instance, Rodrik *et al.* (2004) with Glaeser *et al.* (2004) in the JEG.

Another problem is the conflation of economic development or prosperity with economic growth. Many of the early studies investigating the effects of governance on economic performance focussed on per capita income *levels* rather than income *growth*.⁹ The existence of some long-run causal link from good governance to high income *levels* seems to be almost universally accepted, even though its exact nature is under debate.¹⁰ By contrast, the existence of a link between good governance and economic *growth* within the policy-relevant time horizon of one or two decades is far from clear. Rodrik (2008, p. 19) states that he is “not aware of any strong econometric evidence that relates standard governance criteria to growth (all the evidence is about income levels)”. The argument of Glaeser *et al.* (2004) is that empirically, human capital is a much more important determinant of growth than are institutional factors.

Predating the governance-growth debate, a number of studies have been published that questioned the robustness of cross-country growth regressions on different grounds. Levine and Renelt (1992) demonstrated how sensitive growth regressions can be to small changes in the set of conditioning variables. Sala-i-Martin (1997) followed suit with ‘two million regressions’, and his somewhat more lenient criteria permitted more positive conclusions. Temple (1998) demonstrated the sensitivity of technology and convergence parameter estimates to measurement errors. Kalaitzidakis, Mamuneas and Stengos (2000) examined the sensitivity of growth regressions to the introduction of some non-linear effects. Those studies did not address specifically the governance-growth debate, but some more recent studies did. Easterly *et al.* (2004) expanded and extended the sample used by Burnside & Dollar (2000) and, using exactly the same specification, found that the earlier result entirely broke down. Glaeser *et al.* (2004), as mentioned, argued that Acemoglu *et al.*’s (2001) settler mortality is standing in as an instrument for human capital rather than institutions.

The panel data approach to growth empirics was launched by Islam (1995), but governance indicators were not part of the picture then. In view of the very limited availability of historical series, early studies of the economic importance of governance were mostly based on cross-sectional data.¹¹ As argued above, it is practically hopeless to establish the simultaneous causal relationships between growth and governance indicators in a purely static, cross-sectional setting, as some of those studies purported to do, for want of generally convincing instrumental variables. Recently, stimulated by the increasing availability of data, the trend towards the use of panel data and adapted methods has been gaining strength in the governance-growth debate. Bhattacharya (2009) focuses on the effect of four distinct institutional variables on economic growth, and finds a positive impact of the ICRG law and order indicator. Jacob & Osang (2007) confirm the positive effect of institutions on growth with panel data methods, drawing on the WGI rule of law among several other institutional variables.

The longitudinal (time) dimension in panel data changes the perspective. While the panel raises the issue of heterogeneity and other complications, at the same time it improves the prospect for formulating plausible identifying exclusion restrictions. The relevance of the governance-growth linkage is not with respect to short-term fluctuations but rather with respect to medium to long-term performance in economic growth. It seems

⁹ Among many: Kaufman & Kraay (2002), Rodrik *et al.* (2002), Alcalá & Ciccone (2003).

¹⁰ Indeed, the relationship between good governance and prosperity is sometimes seen as tautological.

¹¹ Notable exceptions are Chong and Calderon (2000); Durham (1999); Olson *et al.* (2000). Chong and Calderon (2000), in particular, exploit the variation over time in growth and governance data to investigate the issue of causal feedback. Applying a version of the ‘Granger causality’ test, they find strong evidence of causality in both directions.

reasonable to expect a time lag of a few years (why not of the order of 5) for changes in governance, institutions and procedures to seep through economic activity and exert their impact on growth. Similarly, any real feedback from economic prosperity towards institutions and procedures of governance will take time to be implemented. Admittedly, the perceptions of experts and survey respondents about the quality of governance are likely to be influenced much more rapidly by economic conditions in their country. By the same token, any instantaneous effect of growth on governance indicators is more likely to be the result of these perceptions than evidence of a real feedback or substantive instantaneous changes in the institutions and procedures of governance. Hence, exploiting the slow dynamics of institutions may help to resolve the endogeneity issues; unfortunately, the open-endedness of growth theories is also depriving us of theoretical guidelines for empirical dynamics.

The bottom line of our brief recapitulation of the literature is that the robustness of growth regressions is endangered from many sides. In this chapter we investigate whether recently accrued data sets, panel data methods and new insights from the quality of governance debate help resolve this fragility. We concentrate attention less on the choice of the other conditioning variables (Levine & Renelt, 1992; Sala-i-Martin, 1997), and more on the choice among several available proxy indicators for the rule of law as well as the way they are scaled.

The contribution we hope to make to the already vast literature on growth is the following. First, we constitute a comprehensive collection of existing rule of law indicators for the largest possible set of countries. This is the raw material for an important fragility test: Does the evidence of a link between rule of law and economic growth depend on the particular choice of indicator used? In other words, we let the four most popular rule of law indicators run a ‘horse race’ and compare the intensity and the robustness of their effects.

Second, we exploit the time dimension of the data and rely on panel data methods to identify the effect of rule of law on economic growth over the policy-relevant horizon of one or two decades. Such work was relatively rare in the past because of the scarcity of governance data with any historical depth. Our data-gathering exercise aims at the widest possible country coverage and the longest possible historical series, encompassing many of the data sets used in previous studies.

Third, we propose solutions to a number of econometric problems that are likely to be relevant. (a) We deal with heterogeneity in the form of unobserved random effects but we allow them to be related to some explanatory variables in the spirit of Mundlak (1978). We also consider certain forms of nonlinearity or heterogeneity in the slope coefficients, e.g., by allowing for interactions involving the governance indicators. (b) We tackle measurement errors in initial income (which are likely to induce spurious correlation with growth and an artificial endogeneity problem) by standard instrumental variables methods. The same method may also capture some omitted variable (heterogeneity) bias along the way. (c) Even though governance indicators are often presented on pseudo-metric scales, they really reflect qualitative rankings only. We avoid any cardinal interpretations of the indicators and the standard linearity assumption about their relationship to growth. We transform the published indicators into categorical variables indicating whether a country is perceived to have a markedly high or a markedly low quality of rule of law as compared to the median pack of countries.

In the end, we arrive at a somewhat unexpected conclusion. We fail to track down a robust relationship between the most popular rule of law indicators and economic growth. Assembling more data seems to raise more doubts about the ability of existing indicators

to capture aspects of the rule of law that foster growth. We can therefore provide no empirical justification for the current policy of targeting development aid and credit at countries scoring well on such indicators.

The structure of the chapter is as follows. Section 2 presents the standard growth model on which our specifications are based. Section 3 describes the database and the set of rule of law indicators used. Section 4 discusses specification issues and other methodological choices to be made. Section 5 presents empirical results, in particular as they exhibit the fragility of the effects found. Section 6 discusses what policy conclusions can be drawn from these findings.

6.2 Model formulation

This section presents the standard growth model on which our specifications in the sequel will be based, and discusses how to deal with the heterogeneity in a panel of countries from all continents. Related issues of interest include a clarification of the income level/growth dilemma and how it relates to the neoclassical versus endogenous growth debate; the relevance or otherwise of the so-called convergence debate; and time series aspects.

Neoclassical growth models following Solow (1959) predict parallel steady state growth paths for the living standards (per capita income or output) of all the countries in the world. The universal rate of growth in the steady state is exogenously determined by worldwide technological progress (usually represented by means of a time trend). The common trend does not imply a common standard of living, though. Output and income per capita in the steady state depend on the rate of saving or capital accumulation, possibly for distinct categories of capital (physical and human, private and public), as well as on population growth (which dilutes the capital intensity). If countries maintain different saving habits, investment policies and fertility rates, then disparities in standards of living will also subsist, even if all growth rates converge to that of the worldwide technological trend.

It is possible to assume that the technology of social organisation, like the technology of production, progresses along some exogenously driven trend and, accordingly, to subsume the quality of institutions or governance into ‘total factor productivity’.¹² However, the plausibility of such an exogenously driven trend in the technology of social organisation is disputable. If countries can maintain different preferences toward savings, investment and fertility, why would they not keep different preferences with respect to the principles of social organisation? It is not hard to find examples of states that refuse to conform to hypothetical worldwide trends in governance practices.

To facilitate the discussion we introduce some formal notation. Let $y_{i,t}$ represent the natural log of GDP (income or output) *per capita* for country i in time period t ; and $x_{i,t}$ a vector of regressor variables thought to determine the steady-state output per person: saving or investment rates in physical and human capital, labour force or population growth rate, and a technological time trend.¹³ The subscript i refers to a worldwide

¹²In neoclassical growth models, ‘total factor’ or ‘multifactor’ productivity growth is defined and measured as the component of growth that is not explained by growth in inputs. It is usually attributed to progress in the technological and organisational efficiency of the production process (insofar as this is not embodied in physical or human capital).

¹³Following Mankiw, Romer & Weil (1992), the labour force or population growth rate has often been increased with a fixed fraction of the order of 5 to 7%, representing the sum of rates of (labour-augmenting) technological progress and (both physical and human) capital depreciation. The introduction of human capital (as distinct from physical capital) as a production factor is also a relatively recent addition to the Solow growth model; see Barro (1991); Barro & Sala-i-Martin (1992); Mankiw, Romer & Weil (1992);

cross-section of countries ($i = 1, \dots, N$) and the subscript t to a fixed set of calendar periods (say, *quinquennia* indicated by their closing date: $t = 1946, 1951, \dots, 2001, 2006$). Empirical models based directly on a Solowian steady state relate $y_{i,t}$ to $x'_{i,t}$ without the complication of dynamics. Such a relationship can be interpreted as a description of the steady state path of per capita output (living standards) as driven by capital accumulation, population growth and the technological trend. Deviations of a country from its steady state path are likely to contain some persistence over time, and therefore we would expect the model disturbances to be serially correlated. This has not been a concern in studies estimating such equations using a purely cross-sectional data set. On the other hand, studies investigating the time series properties of per capita output cast some doubt on the trend-reverting behaviour it should exhibit in a neoclassical world with a fixed technological trend.

It is conceptually straightforward to extend the set of regressors with other variables possibly affecting either steady-state output or deviations from the steady state. There are several reasonable candidates: (i) political stability (Alesina & Rodrik, 1994); (ii) financial sector development (King & Levine, 1992); (iii) trade openness (Frankel and Romer, 1999); (iv) the quality of institutions and governance (Knack & Keefer, 1995; Rodrik *et al.*, 2002); (v) trust or social capital (Knack and Keefer, 1997); (vi) income inequality (Alesina & Rodrik, 1994; Persson & Tabellini, 1994; Barro, 2000; Banerjee & Duflo, 2003); and (vii) geography (Sachs 2003). To fix the ideas, suppose the relationship between $y_{i,t}$ and $x_{i,t}$ is extended with a vector $z_{i,t}$ of governance indicators and other control variables as additional explanatory variables. The serial correlation in the disturbances is likely to be mitigated somewhat but not entirely resolved by this extension. In fact, a major concern is that if the neoclassical model with its exogenous technological trend is incorrect, then deviations from the steady-state outcome might not even be stationary, rendering estimation of the relationship in the time dimension meaningless. (Cross-sectional regressions are not affected by this problem.)

Growth models à la Mankiw, Romer & Weil build in deviations from the steady-state path and are derived as a process of linear approximation around it. Let p denote the length of a standard period of time; since the publication of Islam (1995) a standard period length of 5 years is usually chosen, which serves to smooth out short-run business cycles as well as gaps in governance and other indicators. We find it convenient to denote annualised growth over a future period $[t, t + p]$ by

$$g_{i,t+p} = \frac{1}{p} (y_{i,t+p} - y_{i,t}) . \quad (6.1)$$

A typical growth equation, relating growth to the initial output level, the labour force, physical and human capital accumulation, a technological trend, and one or more governance indicators or other controls, might read as follows:

$$g_{i,t+p} = c_i + y_{i,t}\alpha + x'_{i,t}\beta + z'_{i,t}\zeta + u_{i,t+p} . \quad (6.2)$$

The c_i represent country-specific time trends and (unless they coincide) symbolise intercept heterogeneity; $(y_{i,t}, x'_{i,t}, z'_{i,t})$ is the (row) vector of regressors; α, β, ζ are the corresponding (vectors of) parameters; and $u_{i,t+p}$ is a random error term, known as an ‘idiosyncratic’ disturbance. The vector of regressors will in general include fixed time effects (period dummy variables) as well.¹⁴ A limited form of heterogeneity in slope coef-

Benhabib & Spiegel (1994).

¹⁴In a one-shot cross-section of countries, the individual country effects are absorbed in the disturbance term and the time effect cannot be identified separately.

ficients can be introduced simply by augmenting the equation with interactions between (by way of example) one indicator, say $z_{1,i,t}$, and the $x'_{i,t}$ vector:

$$g_{i,t+p} = c_i + y_{i,t}\alpha + x'_{i,t}\beta + z'_{i,t}\zeta + z_{1,i,t}x'_{i,t}\xi_1 + u_{i,t+p}. \quad (6.3)$$

Much more general treatments of parameter heterogeneity are possible but will not be attempted in this chapter.¹⁵

The growth equations above include unobserved heterogeneity in the form of time-constant individual effects denoted c_i . These individual effects represent time trends specific to the panel units and may be related to the levels of the included regressors; in fact, unless the c_i coincide they are necessarily correlated somewhat with $y_{i,t}$. The standard way to deal with this problem in linear models is to apply a data transformation (usually ‘within demeaning’ or ‘first differencing’) designed to eliminate the unobserved effects from the model. The cross-sectional variation is then differenced away (apart from individual-specific trends in growth). The advantage of this elimination is that ‘dynamic panel bias’ is avoided. The cost of elimination is that cross-sectional effects that are invariant over time are not estimated, and other effects of interest must be inferred from the within-country time variation only. Hence, the coefficients of variables that vary slowly over time are estimated imprecisely. Furthermore, the consequences of measurement error may be exacerbated. Since institutional features are known to be both persistent and hard to measure, these problems are likely to be particularly relevant for the governance indicators in $z'_{i,t}$.

In contrast to within demeaning or first-differencing methods, random effects methods assume away or ignore any correlation between c_i and the regressors. An assumption of mean-independence between c_i and the regressors makes it possible to derive the covariance structure of the composite random term ($c_i + u_{i,t+p}$) and exploit both the cross-sectional and the time variation in the data efficiently by an application of Generalised Least Squares (GLS). Although the independence assumption cannot be strictly true, there is a trade-off between the loss of consistency and the gain of efficiency due to the use of the cross-sectional variation. This explains the attractiveness of random effects methods in country panels.

If the governance indicators in $z'_{i,t}$ are indeed slow-moving and affected by measurement error, there is a distinct possibility that their time variation contributes little information. In the time dimension it may be dominated by measurement error and overall it is dominated by cross-sectional variation. The simplest way to try and alleviate the effects of measurement errors (assuming the serial correlation in the measurement errors is not too high) is to ‘average them out’. Specifically, we define the time averages

$$\bar{z}_{i,\cdot} \equiv \frac{1}{T} \sum_{t=1}^T z_{i,t},$$

and interpret them as a set of persistent individual features.¹⁶ Equation (6.3) can then be modified into

$$g_{i,t+p} = c_i + y_{i,t}\alpha + x'_{i,t}\beta + \bar{z}'_{i,\cdot}\zeta + \bar{z}_{1,i,\cdot}x'_{i,t}\xi_1 + u_{i,t+p}, \quad (6.4)$$

¹⁵Lee et al. (1997) demonstrated the potential importance of allowing for more general slope heterogeneity, as in

$$g_{i,t+p} = c_i + y_{i,t}\alpha_i + x'_{i,t}\beta_i + z'_{i,t}\zeta_i + z_{1,i,t}x'_{i,t}\xi_{1,i} + u_{i,t+p}.$$

¹⁶Mundlak (1978) argued that, in general, time averages of all regressors should appear in the regression, in addition to their time-varying counterparts. Ultimately then, there is no difference left between the augmented random effects model and the fixed effects model.

with the idiosyncratic disturbance $u_{i,t+p}$ redefined appropriately. Apart from the interaction term, equation (6.4) amounts to a standard growth model including the governance effects only by way of individual country effects.

The aim of formulations like the last one is to improve the chances of detecting the effects of governance indicators in our growth equations. We try to alleviate the ‘attenuation bias’ we may expect as a result of measurement error, and avoid the loss of information required to eliminate the heterogeneous intercepts entirely. By doing so, we incur the risk of a ‘dynamic panel bias’ affecting primarily the estimation of α (because of the correlation implied between c_i and $y_{i,t}$) and also the estimation of β , but arguably less the estimation of ζ .

6.3 Data

The first step of our study was to assemble a comprehensive data set, including as many as possible of the publicly available governance indicators as well as macroeconomic variables. The main (secondary) sources of data we used were (i) the Quality of Government database of Teorell *et al.* (2008); (ii) the International Country Risk Guide from the Political Risk Services Group (commercially available online); (iii) the Penn World Tables 6.2 of Heston *et al.* (2006); (iv) the World Bank’s (2007) World Development Indicators; (v) the geography data set of the Center for International Development and Geography of the Harvard Kennedy School; and (vi) the historical statistics of Maddison (2003). Choices had to be made concerning the periodicities at which to construct the data, the scalings to be applied, the choice of rule of law indicators, and the control variables to be included. This section describes the choices we made.

6.3.1 Data periodicity

As an intermediary step we construct a large (unbalanced) panel data set with yearly data from 1946 to 2006 for a total of more than 200 countries. Since Islam (1995), it has been standard practice in studies of medium-term growth and development to reduce yearly data into five-year aggregates or *quinquennial* data. Such time aggregation is useful to smooth out short-term dynamics (business cycles) and random measurement error. Moreover, aggregating yearly data absorbs gaps in certain series constructed from low-periodicity surveys; this conveniently reduces the proportion of missing observations. Thus we constructed our database covering the twelve five-year periods from 1947-1951 to 2002-2006.¹⁷ Within each ‘quinquennium’ all stock and flow data are calculated as averages of the available yearly data. Since most variables are missing for at least some countries in the early decennia of the period, the panel is unbalanced, and we keep it so rather than to exacerbate the selectivity associated with orderly data production. Accordingly, reported regressions draw on different numbers of periods and observations depending on the set of variables used, which will be indicated in the tables. Since these different sample coverages are not neutral for the regression outcomes, results will also be reported under a ‘balanced samples’ restriction, which does not refer to balancing the panel dimensions but to balancing the countries and years for which all four governance indicators are available.

¹⁷ Actually we repeated all calculations for decennia (ten-year periods) and longer periods. This did not lead to systematically different results.

Table 6.1 Rule of Law indicators

	(1)	(2)	(3)	(4)
Indicator name and producer	FI Property Rights Fraser Institute	HF Property Rights Heritage Foundation	ICRG Law and Order PRS Group	WGI Rule of Law World Bank Institute
Primary sources	Expert assessment	Expert assessment	Expert assessment	Composite index
Indicator scale (higher=better)	0 to 10	0 to 100	0 to 6	-2.5 to 2.5
Country coverage	129	163	145	202
Time coverage	1970-2004	1994-2006	1984-2003	1996-2006
Frequency	Yearly	Yearly	Yearly (and more)	Yearly since 2002
Availability	Public	Public	Data not public	Aggregate indicators
Description	Legal Structure and Security of Property Rights	Security of private property against expropriation risk etc.	Country risk ratings by private rating company (PRS Group)	Composite of numerous expert assessments and perception surveys
Part of	F.I. Economic Freedom Index	H.F. Index of Economic Freedom	International Country Risk Rating (ICRG)	Worldwide Governance Indicators (WGI)
Advantages	Earliest, longest coverage, some detailed data available on site	Well-defined criteria, wide country coverage, comparability, changes in ratings documented	Single stable source, long coverage, basic methodology documented	Two-year frequency, wide country coverage, basic methodology documented
Shortcomings	Ambiguous concepts, varying sources, lack of comparability over time and countries	Changes in ratings are often due to revisions, has been criticised for Anglo-Saxon bias	Data not public, unexplained changes in ratings and revisions over time	Ambiguous concepts, varying composition, lack of comparability over time and countries
Abbreviation	(1) Fraser PR	(2) Heritage PR	(3) ICRG L&O	(4) WGI RoL

6.3.2 Choice of indicators

Despite or perhaps because of the existing doubts about what the right conception of ‘rule of law’ is and, more to the point, about which form of rule of law is most effective at different levels of development to stimulate economic growth, many different indicators are in circulation. They are produced by various international organisations in need of assessing which countries have well-functioning rule of law systems.

As mentioned in the introduction, we selected four rule of law indicators with sufficient country and time coverage to make a panel data approach feasible. These are the four indicators mentioned in the introduction: (i) the ‘Legal Structure and Security of Property Rights’ subcomponent of the Fraser Institute’s Economic Freedom rating; (ii) the ‘Property Rights’ subcomponent of the Heritage Foundation’s Index of Economic Freedom; (iii) the ‘Law and Order’ subcomponent of the International Country Risk Guide (ICRG), produced by the Political Risk Services (PRS) Group; and (iv) the composite ‘Rule of Law’ indicator which is part of the World Bank Institute’s Worldwide Governance Indicators (WGIs).¹⁸

Table 6.1 lists the four indicators along with some of their details, the type of primary sources they use, their coverage, and some advantages and shortcomings of each. All four indicators are focussed either mainly or exclusively on the functioning of the rule of law for the private sector; they are examples of what Trebilcock & Daniels (2008) call a ‘thin’ conception of rule of law. All four have been widely reported in the media and used in quantitative studies.

The four sources, obviously, are themselves secondary sources using compilations from different primary sources (surveys) and do not therefore have exactly the same coverage, neither in terms of geography (countries) nor in terms of history (years). For a part, the primary sources and precise contents of the four indicators differ. For example, while the Heritage Foundation’s ‘Property Rights’ index is more narrowly defined with criteria related to the protection of private property and the enforcement of contracts, the Worldwide Governance Indicator for ‘Rule of Law’ contains in addition many variables related to the incidence of crime. The composition and the methodology of construction also vary across indicators. The Worldwide Governance Indicators are most comprehensive in terms of information sources and are the only producer not to rely exclusively on expert assessments (although those are still predominant).

Nevertheless, the four indicators are very much interdependent. The WGI ‘Rule of Law’ itself uses, among other sources, both the Heritage Foundation’s ‘Property Rights’ indicator and the ICRG ‘Law and Order’ indicator. Both the Heritage Foundation and the World Bank Institute use data from The Economist Intelligence Unit in constructing their indicators. Both the WGI ‘Rule of Law’ and the Fraser Institute’s ‘Legal Structure and Security of Property Rights’ are partly based on data from the World Economic Forum’s Global Competitiveness Report. The Fraser Institute’s sources have been evolving and in recent years, they have included both the WGI ‘Rule of Law’ and the ICRG ‘Law and Order’ indicators.

Table 6.2 shows the correlations between the four indicators, in five-year levels (Panel A) and in five-year differences (Panel B). Although the correlations in the levels are substantial, they also leave room for independent variation. The highest correlation is between

¹⁸Respective references: (i) Fraser Institute’s Economic Freedom rating, Gwartney and Lawson (2006); (ii) Heritage Foundation (1995-2008); (iii) International Country Risk Guide (ICRG), Political Risk Services Group (2007); (iv) Worldwide Governance Indicators (WGIs), Kaufmann et al. (2007).

Table 6.2 Correlations between different Rule of Law indicators

Panel A: Five-year-average levels				
(Common sample of 325 observations)				
	Fraser PR	Heritage PR	ICRG L&O	WGI RoL
Fraser PR (Property Rights)	1			
Heritage PR (Property Rights)	0.806	1		
ICRG L&O (Law and Order)	0.860	0.695	1	
WGI RoL (Rule of Law)	0.929	0.878	0.831	1
$\ln GDP_{pc}$	0.735	0.717	0.838	0.681
$\Delta \ln GDP_{pc}$	0.316	0.260	0.333	0.330
Cheibub-Gandhi ‘regime type’	-0.267	-0.290	-0.150	-0.292

Panel B: Five-year changes				
(Common sample of 212 observations)				
	Fraser PR	Heritage PR	ICRG L&O	WGI RoL
Fraser PR (Property Rights)	1			
Heritage PR (Property Rights)	0.083	1		
ICRG L&O (Law and Order)	0.377	0.090	1	
WGI RoL (Rule of Law)	0.230	0.335	0.366	1

the WGI ‘Rule of Law’ and the Fraser Institute’s ‘Legal Structure and Security of Property Rights’ indicator, the lowest between the Heritage Foundation’s ‘Property Rights’ indicator and the ICRG ‘Law and Order’. As one would expect, the correlations between changes over time are much lower. The highest correlation is between changes in the ICRG’s ‘Law and Order’ indicator and in the Fraser Institute’s ‘Legal Structure and Security of Property Rights’ indicator; the lowest correlation is between changes in the latter and in the Heritage Foundation’s ‘Property Rights’ indicator. Finally, we remark that the four Rule of Law indicators are highly correlated with the level of GDP per capita and weakly correlated with growth in GDP per capita.

We mentioned in Table 6.1 what we think are the main strengths and weaknesses of the four indicators. Imperfect measurement is more than likely to be an unavoidable feature of all such constructed indicators. One specific problem occurring repeatedly with Heritage Foundation and ICRG is that a change in the rating from one year to the next does not always reflect an actual change in the quality of governance over time but may be due to a non-retrospective correction or revision by the data producing agency. The occurrence of such corrections suggests that unlike classical measurement errors which are assumed independent white noise, observation errors in the rule of law indicators may be persistent and serially correlated. Since the WGI indicator is a composite based on numerous other indicators, it is likely to inherit the persistence of measurement errors. The WGI indicator has an additional specific feature following from its design: it is normalised and scaled year by year. Therefore, the scale is not strictly comparable over time and the figures contain little or no time series information; they should be interpreted as a repeated cross-section rather than a longitudinal data set. More generally, governance scores provide ordinal

rather than cardinal information, and therefore have a pseudo-metric character. As will be discussed presently, we will try to alleviate the consequences of measurement errors, non-metricity and implied asymmetries or nonlinearities in the published indicators by simple categorical transformations.¹⁹

6.3.3 Scaling of indicators

The next choice to be made concerns scaling. Most indicators are published in cardinal form, i.e., as numbers on a specially designed pseudo-metric scale. Specifically, the scale goes from 0 to 10 for the Fraser Institute; from 0 to 100 for the Heritage foundation; from 1 to 6 for ICRG; and from -2.5 to +2.5 for WGI. This is done for practical reasons and probably for convenience of communication. Fundamentally, however, the information the indicators convey is at best of ordinal nature, and the scales are not always comparable over time, let alone across sources. Including such indicators with constant coefficients in linear regressions implies disputable assumptions of cardinality and linearity. To relax these to some extent, we extract the gist of the information in the data by a simple categorical transformation. We define a *low* value of a governance indicator as a value less than or equal to the lower quartile limit, and a *high* value as one equal to or above the upper quartile limit. The values in between we call the *median* values (approximately the middle half of the sample). We then generate two binary indicators (dummy variables), one for the *low* category and one for *high* category. This transformation does not retain the pseudo-metric information included in the rule of law indicators, and greatly facilitates the interpretation of the results; the sign and magnitude of coefficients are easy to interpret and to compare across indicator sources. These advantages extend to interaction terms involving the dummy variables.

6.3.4 Other variables

The appendix includes a list of the variables used, their definitions and sources. We only discuss the main ones very briefly here. Annual data for GDP per capita is obtained from the Penn World Tables (PWT 6.2), probably the most widely used source for worldwide comparative macro-economic data; see Heston, Summers and Aten (2006). Following tradition in empirical models of growth, the annual data are transformed into average levels and growth rates over five-year periods (quinquennia). Explanatory variables refer to the quinquennium prior to the period of growth explained; in effect, they are lagged one period. Apart from the initial level of per capita income, these explanatory variables comprise mainly an investment rate, a population growth rate, trade openness, type of regime (dictatorship or democracy), education, health, climate (temperate zone), and rule of law.

In the estimations we will report in the sequel, we used data for the investment rate, population growth, and trade openness, obtained from the World Development Indicators (World Bank, 2007). The investment rate is gross fixed capital formation as a percentage of

¹⁹For critical reviews of these and other problems with governance indicators, see for instance Arndt and Oman (2006); Charles *et al.* (2008); Davis (2004); Glaeser *et al.* (2004), Knack (2007); Kurtz and Schrank (2007); Thomas (2007); and van de Walle (2005). A reply to some of the comments on the WGIs is given by Kaufmann, Kraay and Mastruzzi (2007b). Glaeser *et al.* (2004) argued that most indicators used in studies to proxy for institutions measure “outcomes, not some permanent characteristics that North refers to”. Accordingly, we interpret the indicators as time specific outcomes, proxying one aspect of the quality of institutions and subject to evolution and change. We adopt the view that common definitions of the quality of governance are perfectly compatible with changes over a time-horizon of five or ten years.

GDP. Population growth is the growth rate in total (*de facto*) population. Trade openness is the sum of exports and imports of goods and services measured as a share of GDP.

To control for different types of political regime, we used the classification originally developed by Przeworski *et al.* (2000) and extended by Cheibub & Gandhi (2004). The advantage of the classification is threefold: i) it is transparent and replicable, ii) data is available for up to 198 countries since 1946, and iii) there is no conceptual overlap with our rule of law measures. The variable is coded 0 if a country's regime is classified as democratic, 1 if it is classified as a dictatorship. A regime is considered a dictatorship if the chief executive is not elected, the legislature is not elected, there is no more than one party, or there has been no alternation in power. Transition years are attributed to the emerging regime.

The regime type data provided by Cheibub & Gandhi are highly correlated with democracy variables proposed by other sources. At the same time, the regime classification can be clearly distinguished from the rule of law indicators, as illustrated by the low correlations shown in the last row of Panel A in Table 6.2. Evidently, excellent property rights protection and low crime rates are possible in dictatorships and one-party states. A prominent example is Singapore, which ranks at or near the top on all four rule of law indicators (next to countries such as Switzerland and the US), even though it is classified as a dictatorship in the Cheibub & Gandhi data.

The quality of human capital is proxied with both education and health. The average number of schooling years in the total population aged 15 and over is adopted from Barro & Lee (2000). Unfortunately, even the updated Barro & Lee schooling years data has a limited sample coverage. Alternative education variables from the World Development Indicators (labour force with primary, secondary and tertiary education) present a wider country coverage but do not so far have sufficient historical depth.²⁰

Health is proxied using a child immunisation rate against DPT (diphtheria, pertussis or whooping cough, and tetanus). A child is considered adequately immunised against DPT after receiving three doses of vaccine. The immunisation rate is defined as the percentage of children aged 12-23 months who received vaccinations before 12 months or at any time before the survey. The sources of the information are the World Development Indicators (World Bank, 2007), the World Health Organization and the United Nations Children's Fund.

Finally, we try to account for geographical (or at least climate) diversity by means of a 'temperate zones' variable obtained from Gallup *et al.* (1999), and defined as the percentage of a country's population living in so-called 'Koeppen-Geiger temperate zones' in 1995.

6.4 Specification issues and Estimation Methods

In this section we discuss remaining specification issues and the choice of estimation methods. We have introduced above four rule of law indicators from different sources and with different coverage, both in terms of geography (countries) and in terms of history (years). The most straightforward way to find out if they perform comparably as explanatory variables in growth regressions is to try them out one by one. Although nothing in principle prevents us from including them collectively to compare their effects and evaluate their complementarities, it will soon become obvious that this is too much to be asked from the

²⁰ An improved educational data set has recently been made available by Cohen & Soto (2007) and is currently under investigation.

data.²¹

Next a selection had to be made among a very large set of potentially relevant control variables. In the previous section we presented a shortlist of regressors comprising initial income per capita, the investment rate, the population growth rate, trade openness, type of regime (dictatorship), education, health, temperate zone, and rule of law. We consider pooled OLS (POLS), random effects (RE) and fixed effects (FE) specifications including all. To limit the proliferation of parameters and the loss of degrees of freedom, we also consider pooled estimates and random effects estimates in which several control variables are replaced by time-averages or initial values. Since few coefficients turn out significant, we will present shortened lists of regressors.

As is well-known, unless the individual effects are (mean-)independent of the regressors, POLS and RE methods will not deliver consistent estimates of the coefficients and their standard errors. From the point of view of robustness against unobserved heterogeneity, fixed effects estimates are the safest to use. Unfortunately, they are also the most demanding in terms of degrees of freedom, probably the most sensitive to measurement error, and they provide no direct estimates of the effects of country-specific, time-constant variables.

We have not yet addressed the possible endogeneity of initial income.²² At least two plausible reasons have been mentioned to expect the initial income level to be endogenous in growth equations like (6.2), (6.3), or (6.4). One reason is the correlation implied between c_i and $y_{i,t}$ (the cause of what is known as ‘dynamic panel bias’). The other is measurement error in income per capita levels; this is likely to cause on the one hand attenuation bias in its coefficient and on the other hand spurious correlation with growth (since the latter is measured as the forward difference in income per capita). The standard way to deal with these issues is by means of instrumental variable (IV) methods. To deal with the first issue (dynamic panel bias) we would need instruments uncorrelated with the individual effects, like lagged growth rates of income. Such instruments turn out to be extremely weak; furthermore, our interest focusses on the rule of law coefficients which are likely to be relatively little affected by this source of bias. We chose to target the second source of bias by instrumenting the initial income level with its own lag. We ran pooled, RE and FE versions of the IV estimations (to be abbreviated as P-IV, RE-IV and FE-IV, respectively).

In the next section we try to present a representative selection of estimation results. We will report two sets of estimates, one treating $y_{i,t}$ as exogenous and the second instrumenting $y_{i,t}$ using its lagged value $y_{i,t-p}$. The other regressors function as their own instruments. All specifications also include fixed time effects in the form of dummy variables.

The first set comprises the following four estimates, in increasing order of generality: (1) Pooled OLS for equations of the type (6.3), denoted ‘POLS’; (2) Random Effects GLS for equations of the type (6.4), denoted ‘RE-TA’ (for Time Averages; regressors appear either in time-varying form or in time-averaged form, but not both); (3) Random Effects GLS for equations of the type (6.3), denoted ‘RE’; (4) Fixed Effects (LSDV) for equations of the type (6.3), denoted ‘FE’.

²¹ Another avenue would be to construct a super-composite combining information from all four sources, or extract a common factor from them. The cost would be further conceptual dilution and remoteness from ‘actionability’. For the present, we focus on the use of the distinct published indicators

²² Unlike what might be the case in a cross-section analysis, we do not need to be concerned with the possibility of simultaneity between income and governance because of the time lags which are likely to be involved.

The second set consists of two-stage least squares (2SLS) analogues, treating the lagged income level as endogenous, and using the lag of the initial income level as an instrumental variable: (5) Pooled 2SLS for equations of the type (6.3), denoted ‘P-IV’; (6) Random Effects GLS with IV for equations of the type (6.4), denoted ‘RE-TA-IV’; (7) Random Effects GLS with IV for equations of the type (6.3), denoted ‘RE-IV’; and (8) Fixed effects IV for equations of the type (6.3), denoted ‘FE-IV’.

The pooled estimation ignores any form of heterogeneity and is most restrictive. The fixed effects estimation including the time-varying regressors only is the least restrictive, and can be used as a basis for comparison and for specification testing. The random effects estimates can be seen as different compromises between pooled and fixed effects estimation. For easy reference, we recapitulate the presented methods in Table 6.3.

Table 6.3: Summary of estimation methods

Abbreviation:	Method:	Equation:
Panel A:		
1. POLS	Pooled OLS	(6.2)
2. RE-TA	Random Effects GLS with Time Averages	(6.4)
3. RE	Random Effects GLS	(6.2)
4. FE	Fixed Effects or Within OLS (= LSDV)	(6.2)
Panel B:		
5. P-IV	Pooled 2SLS, initial income instrumented	(6.2)
6. RE-TA-IV	RE-TA with initial income instrumented	(6.4)
7. RE-IV	RE with initial income instrumented	(6.2)
8. FE-IV	FE with initial income instrumented	(6.2)

6.5 Empirical Results

This section presents empirical results, in particular as they exhibit the fragility of (the significance of) the effects found as we vary the choices made. In Subsection 5.1 we estimate our baseline growth equation, at first excluding rule of law indicators, yet varying the choice of method and of sample period. In Subsection 5.2 we extend the regression equations with the four main competing rule of law indicators, one by one. We present summary tables on the direction and significance of the effects found. In Subsection 5.3 we discuss the fragility of the estimated coefficients of control variables to the choice of rule of law indicators. In Subsection 5.4 we investigate whether investments in physical capital might be themselves determined by and thereby masking the indirect effect of the rule of law.

6.5.1 Baseline growth equations

In this subsection, we present estimation results for our baseline growth equation, excluding rule of law indicators for the time being, yet varying the choice of method and of sample period. We present Pooled OLS, random effects and fixed effects regressions (Table 6.4), as well as the analogous 2SLS regressions treating the initial income level as endogenous (Tables 6.5).

Table 6.4: Baseline growth equation

	(1) POLS	(2) RE-TA ^c	(3) RE	(4) FE
Panel 6.4(A): Unbalanced samples				
ln GDP per cap.	-0.474* (0.255)	-1.067*** (0.308)	-1.259*** (0.338)	-7.895*** (0.685)
Investment rate	0.0839*** (0.0225)	0.0672*** (0.0250)	0.0405 (0.0270)	-0.0196 (0.0322)
Population growth	-0.216* (0.119)	0.147 (0.127)	0.0108 (0.133)	0.195 (0.142)
Trade openness	0.00210 (0.00401)	-0.00148 (0.00631)	0.00819 (0.00529)	0.0325*** (0.00843)
Dictatorship	-0.351 (0.348)	-1.040 (0.672)	-0.477 (0.396)	-0.154 (0.453)
Education	0.123 (0.0961)	0.209 (0.146)	0.338*** (0.130)	0.196 (0.290)
Health	0.0124 (0.00895)	0.0319** (0.0152)	0.0131 (0.0100)	0.00622 (0.0108)
<i>Temperate zone^c</i>	<i>0.815*</i> (0.485)	<i>1.351**</i> (0.679)	<i>1.668**</i> (0.688)	
Observations	358	373	358	379
Countries		95	94	100
RMS Residual	2.410	2.081	2.017	1.691
Panel 6.4(B): Balanced samples				
ln GDP per cap.	-0.0935 (0.284)	-0.350 (0.411)	-0.465 (0.396)	-13.13*** (1.525)
Investment rate	0.106*** (0.0294)	0.110*** (0.0340)	0.104*** (0.0335)	0.0926** (0.0425)
Population growth	-0.399*** (0.129)	-0.287* (0.152)	-0.443*** (0.149)	-0.0203 (0.210)
Trade openness	0.00514 (0.00436)	0.00961 (0.00849)	0.00590 (0.00559)	0.0149 (0.0114)
Dictatorship	0.181 (0.393)	-0.101 (0.818)	0.386 (0.493)	0.254 (0.778)
Education	-0.0233 (0.103)	0.0791 (0.178)	0.0835 (0.147)	0.428 (0.798)
Health	0.00874 (0.0118)	0.00620 (0.0205)	0.00862 (0.0151)	0.00713 (0.0256)
<i>Temperate zone^c</i>	<i>0.683</i> (0.543)	<i>0.768</i> (0.867)	<i>0.887</i> (0.773)	
Observations	160	163	160	165
Countries		83	82	85
RMS Residual	1.799	1.189	1.173	0.970

Notes: Growth lags one 5-year period behind the regressors; time effects included; ordinary standard errors in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

^c Coefficients in italics refer to time-averaged or time-constant regressors.

Table 6.5: Baseline growth equation, instrumenting income

	(1) P-IV	(2) RE-TA-IV ^c	(3) RE-IV	(4) FE-IV
Panel 6.5(A): Unbalanced samples				
ln GDP per cap.	-0.885*** (0.265)	-1.406*** (0.322)	-1.596*** (0.360)	-9.226*** (0.887)
Investment rate	0.0847*** (0.0225)	0.0714*** (0.0250)	0.0416 (0.0271)	-0.00802 (0.0328)
Population growth	-0.215* (0.120)	0.158 (0.127)	0.00948 (0.133)	0.164 (0.143)
Trade openness	0.00209 (0.00403)	-0.00125 (0.00629)	0.00831 (0.00532)	0.0334*** (0.00849)
Dictatorship	-0.484 (0.350)	-1.214* (0.671)	-0.546 (0.398)	-0.129 (0.456)
Education	0.217** (0.0976)	0.278* (0.146)	0.418*** (0.134)	0.218 (0.293)
Health	0.0159* (0.00900)	0.0367** (0.0152)	0.0149 (0.0101)	0.00684 (0.0108)
<i>Temperate zone^c</i>	<i>0.962** (0.488)</i>	<i>1.439** (0.677)</i>	<i>1.815*** (0.695)</i>	
Observations	358	373	358	379
Countries		95	94	100
RMS Residual	2.419			
Panel 6.5(B): Balanced samples				
ln GDP per cap.	-0.428 (0.291)	-0.464 (0.420)	-0.628 (0.412)	-11.95*** (2.683)
Investment rate	0.107*** (0.0296)	0.113*** (0.0337)	0.104*** (0.0334)	0.0908** (0.0428)
Population growth	-0.406*** (0.129)	-0.277* (0.151)	-0.444*** (0.149)	-0.0494 (0.218)
Trade openness	0.00524 (0.00438)	0.00959 (0.00828)	0.00593 (0.00557)	0.0132 (0.0119)
Dictatorship	0.122 (0.395)	-0.171 (0.801)	0.355 (0.492)	0.352 (0.802)
Education	0.0423 (0.105)	0.0966 (0.174)	0.116 (0.148)	0.343 (0.817)
Health	0.0132 (0.0119)	0.00809 (0.0201)	0.0106 (0.0151)	0.00383 (0.0264)
<i>Temperate zone^c</i>	<i>0.869 (0.547)</i>	<i>0.830 (0.847)</i>	<i>0.984 (0.772)</i>	
Observations	160	163	160	165
Countries		83	82	85
RMS Residual	1.808			

Notes: Growth lags one 5-year period behind the regressors; time effects included; ordinary standard errors in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

^c Coefficients in italics refer to time-averaged or time-constant regressors.

As mentioned in the previous section, we regress the five-yearly growth in GDP per capita on five-year-lagged values of GDP per capita (in logarithms), the investment rate, population growth, trade openness, regime type ('dictatorship'), education, health and the proportion of the population living in temperate geographic zones, all (except the last one) averaged over five-year periods. We vary the sample composition and the extent of heterogeneity allowed in order to assess how the outcomes are affected. We report on the four basic specifications listed in Panel A of Table 6.3, in increasing order of generality: Pooled OLS; Random Effects GLS with time averages replacing slowly-varying control variables (RE-TA); Random Effects GLS proper, including all control variables in their time-varying form (RE); and Fixed Effects (Within OLS or LSDV). All specifications also include fixed time effects in the form of dummy variables, although to save space the corresponding coefficients will not be reported in the tables.

Table 6.4(A) presents the estimates using the widest possible sample composition, Table 6.4(B) presents them for a smaller sample restricted to those countries and years for which all four rule of law indicators are available (the 'balanced samples' case). The first striking outcome is that the estimates are very sensitive to the sample composition. The coefficients of the two Solowian determinants of growth, the investment rate and population growth, have varying signs and magnitudes and weak significance in the full samples; whereas in the restricted samples, the coefficient of the investment rate is close to 0.1 and significant for all methods used, and the coefficient of population growth is negative throughout (though insignificant under fixed effects estimation). Among the additional control variables, trade openness, education and temperate zone exhibit some significant coefficients in the full samples, but none in the restricted one. The coefficient of the lagged income level is systematically negative, but of very variable magnitude and significance.

As mentioned in the previous section, one possible problem with these estimates is measurement error in the lagged income level. Apart from attenuation bias, such measurement error would induce spurious correlation and endogeneity with respect to growth (the lagged income level being the denominator of the growth variable). We therefore repeat the estimations treating the lagged income level as endogenous and instrumenting it with the income level lagged a further five years. These estimates are presented in Table 6.5, with Panel (A) again referring to the full samples and Panel (B) to the restricted 'balanced' samples. In spite of some changes the results are qualitatively very similar to those in Table 6.4, suggesting that measurement error in the income level is not causing major biases.

6.5.2 Introducing the rule of law indicators

In this subsection we extend the growth equations with the rule of law indicators described in Section 6.3, and vary the details of specification and estimation in order to investigate the robustness of the results. Once again, we consider different sample compositions (full unbalanced samples versus balanced common samples) and different estimation methods (depending on the treatment of heterogeneity and the treatment of the endogeneity of lagged income). In addition, we have four different sources of indicators to consider for the rule of law, and two different ways of scaling the indicators (pseudo-metric versus categorical). We test the fragility of results using the four different rule of law indicators interchangeably. As previously, all specifications include time effects (which remain unreported for obvious space reasons).

Estimation results are reported in detail in Tables A.01 to A.08 of Appendix A to this chapter. In this section, for the sake of clarity, we only present a summary view of the

significance and sign of rule of law effects using different estimation methods. Tables 6.6 to 6.8 consist of two panels each, one panel (A) using the largest possible (unbalanced) sample, and the other (B) using only the common (balanced) sample, i.e., the observations available for all the indicators. In these summary tables, we denote estimates not significantly different from 0 at the 10% level by 0; those significant at the 10% but not at the 5% level by their sign (-, +); those significant at 5% but not the 1% level by twice their sign (--, ++); those significant at the 1% level by three times their sign (---, +++). Standard errors are calculated with a heteroskedasticity and autocorrelation-robust formula in Tables 6.6 and 6.7, and also in Table 6.8 whenever possible.

In none of our regressions does the Heritage Foundation indicator have a significant effect. The ICRG indicator only shows significant effects with the “Low” governance quality dummy. The two remaining indicators, WGI and Fraser, more frequently (though not systematically) show significant effects.

Since the four rule of law indicators are not available for the same countries and years, some differences in the results might be due to their different sample coverage. The estimated coefficients and their statistical significance are indeed sensitive to the sample period used. In general, we find more significant effects under the ‘balanced samples’ restriction than in the full samples, in spite of the diminished sample sizes. Nevertheless, the difference in significance between the different indicators cannot be explained by the choice of sample coverage. The contrast in significance between the Heritage and ICRG indicators on the one hand, and the WGI and Fraser on the other, remains present in the common sample, at least under the original pseudo-metric scaling of the indicators (and in a modified form after the categorical rescaling).

Pseudo-metrically scaled rule of law indicators from the Heritage Foundation and from ICRG fail to exhibit significant effects in any of the regressions. Similarly, in none of the regressions using categorically rescaled indicators does the ‘High’ quality of governance dummy variable have a significant positive effect on growth. The ‘Low’ quality of governance dummy performs better, except with the Heritage Foundation indicator, where none of the rule of law indicators seem to matter. Taking these results at face value, one possible interpretation is that only ‘bad’ governance matters, or at least what is rated bad governance by sources other than the Heritage Foundation.

Next we observe that the estimated effects are sensitive to the treatment of heterogeneity. Using Pooled OLS and random effects the pseudo-metric indicators get significant effects, but with fixed effects these weaken, and they vanish altogether in the complete samples. In general, since the quality of governance is unlikely to change quickly over time and has been ‘observed’ only relatively recently, we would expect its effects to be relatively hard to identify under a fixed effects specification (where all cross-sectional variation has been differenced away). Since the random effects with time averages is more economical with the variation in the data than the traditional random effects estimator, it also tends to produce statistically more significant effects. Our estimates conform with these expectations.

Turning to the categorically scaled indicators, the ‘High’ dummy never gets a significant coefficient and thus seems to be immaterial, whatever the method. The ‘Low’ dummy has a negative coefficient in a number of cases, in pooled regressions, but also in random effects regressions in the case of the Fraser Institute. The only indicator that shows some significant effect when full allowance is made for heterogeneity (fixed effects models) is ICRG.

The findings in term of significance of the indicators are unaffected when we treat the lagged income level as endogenous and instrument it with its own lag; the ICRG

coefficients even turn out somewhat more significant with the IV methods.²³

We conclude from these outcomes that none of the rule of law effects is very robust to the inclusion of fixed effects, and few are robust to the inclusion of random effects with individual-specific time averages. Pooled estimates are more significant but are in general inconsistent, and omitted variable bias may explain their apparent significance. Rescaling the indicators into categories so as to allow for a degree of nonlinearity leads to asymmetric estimates of the effects of a ‘high’ and a ‘low’ quality of rule of law: a low score reduces growth significantly, but a high score does not significantly enhance it. Allowing for a certain degree of endogeneity of the initial income level does not affect these conclusions.

Attempts to uncover stronger or more subtle nonlinear relationships were not successful. In particular, we allowed for different effects of rule of law in groups of countries defined by their income levels or their regional membership, using World Bank classifications. By these and similar methods, we failed to find any differences of significance between those various groups of countries.

Table 6.6: Pseudo-metric rule of law indicators

Panel 6.6(A): Unbalanced samples				
	Fraser PR	Heritage PR	ICRG L&O	WGI RoL
POLS	++	0	0	++
RE-TA	+	0	0	+++
RE	0	0	0	++
FE	0	0	0	0
Panel 6.6(B): Balanced samples				
	Fraser PR	Heritage PR	ICRG L&O	WGI RoL
POLS	+++	0	0	+++
RE-TA	+++	0	0	+++
RE	++	0	0	++
FE	+	0	0	0

Notes: Robust standard errors used
 0 = not significantly different from 0 at the 10% level
 -, --, --- = significantly negative at the 10%, 5%, 1% level, respectively
 +, ++, +++ = significantly positive at the 10%, 5%, 1% level, respectively

²³We should note, however, that not all standard errors are strictly comparable, since a variance formula robust to heteroskedasticity and serial correlation is not readily available in some of the cases covered in Table 6.8.

Table 6.7: Categorical rule of law dummies

	Fraser PR		Heritage PR		ICRG L&O		WGI RoL	
	Low	High	Low	High	Low	High	Low	High
Panel 6.7(A): Unbalanced samples								
POLS	---	0	0	0	--	0	---	0
RE-TA	--	0	0	0	0	0	0	0
RE	0	0	0	0	0	0	0	0
FE	0	0	0	0	0	0	0	-
Panel 6.7(B): Balanced samples								
POLS	---	0	0	0	-	0	---	0
RE-TA	--	0	0	0	-	0	-	0
RE	--	0	0	0	0	0	0	0
FE	0	0	0	0	-	0	0	0

Notes: Robust standard errors used
 0 = not significantly different from 0 at the 10% level
 -, --, --- = significantly negative at the 10%, 5%, 1% level, respectively
 +, ++, +++ = significantly positive at the 10%, 5%, 1% level, respectively

Table 6.8: Categorical rule of law dummies, instrumenting income

	Fraser PR		Heritage PR		ICRG L&O		WGI RoL	
	Low	High	Low	High	Low	High	Low	High
Panel 6.8(A): Unbalanced samples								
Pooled IV	---	0	0	0	--	0	---	0
RE-TA-IV	--	0	0	0	-	0	0	0
RE-IV	0	0	0	0	0	0	0	0
FE-IV	0	0	0	0	0	0	0	0
Panel 6.8(B): Balanced samples								
Pooled IV	---	0	0	0	-	0	---	0
RE-TA-IV	--	0	0	0	--	0	-	0
RE-IV	--	0	0	0	-	0	0	0
FE-IV	0	0	0	0	--	0	0	0

Notes: Robust standard errors used when available
 0 = not significantly different from 0 at the 10% level
 -, --, --- = significantly negative at the 10%, 5%, 1% level, respectively
 +, ++, +++ = significantly positive at the 10%, 5%, 1% level, respectively

6.5.3 Fragility of effects of other regressors

The significance of several of the other explanatory variables is highly sensitive to the choice of rule of law indicators. We will make reference here to the detailed tables of estimates presented in Appendix A (Tables 6.A.01 to 6.A.08).

The initial income per capita is significant in pooled estimation when using the ICRG or the Fraser Institute indicators (Table 6.A.01(a)); however, the significance disappears in the shorter, ‘balanced’ data set (Table 6.A.01(b)), indicating that the better significance is due to the longer historical depth achieved when using these sources. In fixed-effects analysis, initial income is highly significant irrespective of the rule of law indicator used (Table 6.A.03 and Table 6.A.04), and for both the full and the balanced data sets. Its coefficient is of a reasonable magnitude (-0.1, correcting for the fact that the growth rate is in percentages).

The investment rate tends to be highly significant and its coefficient is of the order of 0.1, at least under the common samples restriction (Table 6.A.01(b)). This is clearly the case in all pooled cross-section regressions and in the random effects models with time averages. However, the relationship breaks down in the ICRG and Fraser Institute cases when the samples are unrestricted and full heterogeneity is allowed for. Again, this may be related to the much longer samples in these cases, possibly involving a higher risk of structural instability.

The coefficient of population growth is much more erratic. It is significantly negative and of the order of -0.2 to -0.4 in the pooled OLS estimates, and similarly in the random effects estimates under the common samples restriction; however, its statistical significance breaks down with fixed effects, and similarly with random effects in full samples.

The remaining five controls have largely insignificant coefficients, with only a few exceptions.

Trade openness is not significant in any of the pooled or random effects regressions, although its coefficient is almost always positive and of the order of 0.005 to 0.015. It is significant only in fixed effects analysis (Table 6.A.04) when using the ICRG or the Fraser Institute indicators, although much less so when the data set is reduced.

The coefficient of the dictatorship variable has varying signs and magnitudes and is hardly ever significant. This is not too surprising, given that there are few changes in regime type over time.

Education does not fare much better. Only in some random effects models with ICRG and Fraser Institute data does a significant positive impact on growth show up, of the order of 0.2 to 0.3. The other human capital variable, health, exhibits a similar pattern.

Finally, the geographical ‘temperate zone’ variable, which is redundant in fixed effects models (since it is time-constant), always has a positive effect, though it is not usually significant. Its magnitude varies widely, from 0.2 to 1.8.

6.5.4 Rule of law and the investment rate

In this subsection we investigate whether investments in physical capital might be themselves determined by the rule of law and therefore masking its effect to the extent these are indirect. Tables 6.B.01 to 6.B.04 in Appendix B report estimates of an equation explaining the investment rate in every five-year period by income growth and the other main variables in the preceding five-year period. The predictor variables that matter most are the (lagged) investment ratio and population growth and, to a lesser extent, trade openness; however, few effects remain significant in the fixed effects model. In no case does economic growth turn out to be a significant predictor of the investment rate. As

to the rule of law, there is no persistent pattern and therefore (in our view) no robust effect. In the pooled and random effects (RE-TA) estimates, a low Heritage Foundation score seems oddly enough to enhance investment, but this effect disappears in full random effects (RE) and fixed effects (FE) models. In the fixed effects model and under the restriction of balanced samples, a low Fraser Institute score seems to discourage investment. For the remainder, there seems to be no more significance than would be expected by chance. We conclude that we can observe no substantial indirect channel of causality from rule of law to growth through the investment ratio.

6.6 Conclusions

In this chapter we compiled a comprehensive data base of available rule of law indicators and tested the robustness of their estimated effects on economic growth. We found little evidence of robust effects once heterogeneity across countries is allowed for. Among the four indicator sources considered, the rule of law indicator of the Heritage foundation had the least, and that of the Fraser Institute the most significant effects on growth. The pseudo-metric scalings commonly used for the indicators do not seem to add much value to a simple categorical score of the type ‘good/average/bad’. In fact, the effects of ‘good’ and ‘bad’ governance are probably asymmetric (if they exist). One striking finding is that empirically, having a good maintenance of the rule of law matters less (if at all) for economic growth than *not having a bad one*.

However our main conclusion is more cautionary. What we find is a striking lack of evidence for (or a lack of confirmation of) a systematic, measurable, robust effect. We cannot infer from this that rule of law doesn’t matter, but rather that whatever is measured by the four indicators put to the test does not seem critical to economic growth.

As has been argued extensively elsewhere, synthetic indicators run the risk of being either gross oversimplifications (not accounting for all sorts of specific circumstances) or conceptually too vague and diluted. Therefore, great care should be exercised in drawing policy conclusions. This is meant to apply not only to the estimation results in this chapter, but equally to other results, particularly those that seem to reveal a tight connection between the publicly perceived quality of governance and economic growth. Claims that it should be a policy priority to improve one’s scores on rule of law indicators (or property rights indicators) are potentially overstated and cannot be sustained at present by econometric panel data analysis. Users should be warned against reliance on the indicators to rate the performance of governments or institutions.

Several explanations are thinkable why some studies have found more significant effects. One is that such studies selected the ‘right’ (most significant) indicator, or that they became more prominent due to ordinary publication bias. Secondly, a number of earlier studies looked at levels rather than growth, or cross-sectional rather than panel data, therefore attributing country heterogeneity to the variation in whichever indicators were included. The trend is now clearly towards growth in panel data models. A third possible reason is that many studies omitted a proper human capital measure, as argued by Glaeser *et al.* (2004). Clearly, more systematic research is needed on these issues. Meanwhile, empirical analysts should be careful not to overstate the case and policy makers should be careful not to weigh weak evidence too heavily.

References

- Acemoglu D., S. Johnson & J.A. Robinson (2001): "The Colonial Origins of Comparative Development: An Empirical Investigation". *The American Economic Review* 91 (5) 1369-1401
- Albouy D. (2006): "The Colonial Origins of Comparative Development: An Investigation of the Settler Mortality Data". Unpublished, University of California at Berkeley.
- Alcalá F. & A. Ciccone (2003): "Trade, Extent of the Market, and Economic Growth 1960-1996". Economics Working Papers 765, Department of Economics and Business, Universitat Pompeu Fabra.
- Alesina A. & D. Rodrik (1994): "Distributive Politics and Economic Growth". *Quarterly Journal of Economics* 109, 465-490.
- Arndt C.E. (2009): *Governance Indicators*. Maastricht University doctoral dissertation, Maastricht Graduate School of Governance dissertation series Nr. 3.
- Arndt C. & C. Oman (2006): *Uses and Abuses of Governance Indicators*. OECD Development Centre Studies, Paris.
- Aron J. (2000): "Growth and Institutions. A Review of the Evidence". *The World Bank Research Observer* 15 (1) 99-135.
- Banerjee A.V. & E. Duflo (2003): "Inequality and Growth: What can the data say?" *Journal of Economic Growth* 8 (3) 267-299.
- Barro R.J. (1991): "Economic Growth in a Cross-Section of Countries". *Quarterly Journal of Economics* 106 (2) 407-443.
- Barro R.J. (2000): "Inequality and Growth in a Panel of Countries". *Journal of Economic Growth* 8 (3) 267-299.
- Barro R.J. & J.W. Lee (1993): "International comparisons of educational attainment". *Journal of Monetary Economics* 32 (3) 363-394.
- Barro R.J. & J.W. Lee (2000): "International data on educational attainment: Updates and Implications". *Oxford Economic Papers* 3, 541-563.
- Barro R.J. & X. Sala-i-Martin (1990): "Public Finance in Models of Economic Growth". NBER Working Papers No. 3362.
- Barro R.J. & X. Sala-i-Martin (1995): *Economic Growth*, McGraw-Hill, New-York.
- Barro R.J. & X. Sala-i-Martin (2004): *Economic Growth*, 2nd edition. MIT Press, Cambridge.
- Benhabib J. & M.M. Spiegel (1994): "The Role of Human Capital in Economic Development: Evidence from aggregate cross-country data". *Journal of Monetary Economics* 34 (2) 143-173.
- Bhattacharya S. (2009): "Unbundled institutions, human capital and growth". *Journal of Comparative Economics* 37 (1) 106-120.
- Brock W. & S. Durlauf (2001): "Growth Empirics and Reality". *World Bank Economic Review* 15, 2, 229-272.
- Burnside C. & D. Dollar (2000): "Aid, Policies, and Growth". *American Economic Review* 90 (4) 847-868.
- Burnside C. & D. Dollar (2004): "Aid, Policies, and Growth: Revisiting the Evidence". World Bank Policy Research Working Paper, No. 3251 (March 2004).
- Carothers T. (2003): *Promoting the Rule of Law Abroad: The Problem of Knowledge*. Rule of Law Series, Democracy and Rule of Law Project No. 34. Retrieved May 2008, from:
<http://www.carnegieendowment.org/publications/index.cfm?fa=view&id=1169>

- Chamberlain G. (1980): "Analysis of Covariance with Qualitative Data". *Review of Economic Studies XLVII*, 225-38.
- Charles C., V. Djakovic, M. Kaempfer & M. Kozlova (2008): "Indicateurs de gouvernance. Une étude comparative des approches et des stratégies de la Banque mondiale, de l'Union européenne et du PNUD. Rapport final. - Atelier international 2007/2008". Institut d'Etudes Politiques de Paris.
- Cheibub J.A. & J. Gandhi (2004): "Classifying Political Regimes: A Sixfold Classification of Democracies and Dictatorships". Paper presented at the 2004 Annual Meeting of the American Political Science Association.
- Chong A. & C. Calderon (2000): "Causality and Feedback Between Institutional Measures and Economic Growth". *Economics and Politics 12* (1) 69-81.
- Cohen D. & M. Soto (2007): "Growth and human capital: good data, good results". *Journal of Economic Growth 12*: 51-76.
- Dam K. (2006): *The Law-Growth Nexus: The Rule of Law and Economic Development*. Brookings Institution Press.
- Davis K.E. (2004): "What Can the Rule of Law Variable Tell Us About Rule of Law Reforms?" NYU Law and Economics Research Paper No. 04-026
- Durham J.B. (1999): "Economic Growth and Political Regimes". *Journal of Economic Growth 4* (1) 81-111.
- Easterly W. & R. Levine (1997): "Africa's Growth Tragedy: Policies and Ethnic Divisions". *Quarterly Journal of Economics 112* (4) 1203-1250.
- Easterly W., R. Levine & D. Roodman (2000): "Aid, Policies, and Growth: Comment". *American Economic Review 94* (3) 774-780.
- Frankel J. & D. Romer (1999): Does Trade Cause Growth? *American Economic Review 89* (3) 379-399.
- Gallup J.L., A.D. Mellinger & J.D. Sachs (1999): Geography Datasets. Harvard Centre for International Development. Retrieved December 2007 from: <http://www.cid.harvard.edu/ciddata/geographydata.htm>
- Glaeser E.L., R. La Porta, F. Lopez-de-Silanes & A. Shleifer (2004): "Do Institutions Cause Growth?" *Journal of Economic Growth 9* (3) 271-303.
- Gray H. (2007): "Governance for Economic Growth and Poverty Reduction: Empirical Evidence and New Directions Reviewed". Available at: www.gsdr.org/docs/open/RET422.pdf
- Gwartney J. & R. Lawson (2006): Economic Freedom of the World: 2006 Annual Report. Vancouver: The Fraser Institute.
- Gwartney J. & R. Lawson, with S. Norton (2008): Economic Freedom of the World: 2008 Annual Report. The Fraser Institute.
- Heritage Foundation (1995,...,2006): *Index of Economic Freedom 1996,...,2007* (annual issues). Washington, D.C.: The Heritage Foundation and Dow Jones & Company, Inc.
- Heston A., R. Summers & B. Aten (2006): *Penn World Tables Version 6.2*. Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania, September 2006.
- Holmes K.R., E.J. Feulner, M.A. O'Grady, et al. (2008): *Index of Economic Freedom 2008*. Washington, D.C.: The Heritage Foundation and Dow Jones & Company, Inc.
- IDA (2007): *Selectivity and Performance: IDA's Country Assessment and Development Effectiveness*. International Development Association, Feb. 2007.

- Islam N. (1995): "Growth Empirics: A Panel Data Approach". *Quarterly Journal of Economics* 110 (4) 1127-1170.
- Jacob J. & T. Osang (2007): "*Institutions, Geography and Trade: A Panel Data Study*". Departmental Working Papers 0706, Southern Methodist University, Department of Economics.
- Kalaitzidakis P. , T. Mamuneas & T. Stengos (2000): "A Non-Linear Sensitivity Analysis of Cross-Country Growth Regressions". *Canadian Journal of Economics* 33, 3, 604-617.
- Kaufmann D. & A. Kraay (2002): "Growth without Governance". *Economía* 3 (1) 169-215. (Also available as World Bank Policy Research Working Paper 2928, Washington, D.C.)
- Kaufmann D., A. Kraay & M. Mastruzzi (2005): *Measuring Governance Using Cross-Country Perceptions Data* (August 2005). The World Bank. Available at: www.worldbank.org/wbi/governance/govdata/
- Kaufmann D., A. Kraay & M. Mastruzzi (2007a): *Governance Matters VI: Governance Indicators for 1996-2006* (July 2007). World Bank Policy Research Working Paper No. 4280. Available at SSRN: <http://ssrn.com/abstract=999979>
- Kaufmann D., A. Kraay & M. Mastruzzi (2007b): *The Worldwide Governance Indicators Project: Answering the Critics*. World Bank Policy Research Working Paper No. 4149.
- Kaufmann D., A. Kraay & M. Mastruzzi (2008): *Governance Matters VII: Aggregate and Individual Governance Indicators, 1996-2007* (June 24, 2008). World Bank Policy Research Working Paper No. 4654. Available at SSRN: <http://ssrn.com/abstract=1148386>
- Kaufmann D., A. Kraay & M. Mastruzzi (2009): *Governance Matters VIII: Aggregate and Individual Governance Indicators, 1996-2008* (June 2009). World Bank Policy Research Working Paper No. 4978.
- King R. & R. Levine (1993): Finance and growth: Schumpeter might be right, *Quarterly Journal of Economics* 108 (3) 717-737.
- Knack S. (2002): *Governance and Growth: Measurement and Evidence*. Forum Series on the Role of the Institutions in Promoting Economic Growth. Washington, D.C.
- Knack S. (2006): *Governance and Growth*. Retrieved May 2008, from: http://www.sdnpsbd.org/sdi/issues/governance/governance/summarynoteg_governance_and_growth.pdf
- Knack S. (2007): Measuring Corruption: A Critique of Indicators in Eastern Europe and Central Asia. *Journal of Public Policy* 27 (3) 255-291.
- Knack S. & P. Keefer (1995): "Institutions and Economic Performance: Cross-Country Tests Using Alternative Institutional Measures". *Economics and Politics* 7, 207-227.
- Knack S. & P. Keefer (1997): "Does Social Capital Have an Economic Payoff? A Cross-Country Investigation". *Quarterly Journal of Economics* 112 (4) 1251-1288.
- Kurtz M. & A. Schrank (2007): "Growth and Governance: Models, Measures, and Mechanisms". *The Journal of Politics* 69 (2) 538-554.
- Lee K., M. Pesaran & R. Smith (1997): "Growth and Convergence in Multi-Country Empirical Stochastic Solow Model". *Journal of Applied Econometrics* 12 (4) 357-392.

- Levine R. & D. Renelt (1992): "A Sensitivity Analysis of Cross-Country Growth Regressions". *The American Economic Review*, 82 (4) 942-963.
- Maddison A. (2003): *The World Economy: Historical Statistics*. OECD Development Centre, Paris 2003.
- Mankiw N.G., D. Romer, D. Weil (1992): "A Contribution to the Empirics of Economic Growth". *Quarterly Journal of Economics* 107 (2) 407-437.
- Mauro P. (1995): Corruption and Growth. *Quarterly Journal of Economics* 110 (3) 681-712.
- Mundlak Y. (1978): On the Pooling of Time Series and Cross-Section Data, *Econometrica* 46, 69-85.
- North D.C. (1990): *Institutions, Institutional Change and Economic Performance*. Cambridge University Press, Cambridge.
- North D.C. (2005): *Understanding the Process of Economic Change*. Princeton University Press, New Jersey.
- OECD (2008): *Handbook on Constructing Composite Indicators: Methodology and User Guide*. OECD Publishing, Paris.
- Olson M., N. Sarna & A.V. Swamy (2000): "Governance and growth: A simple hypothesis explaining cross-country differences in productivity growth". *Public Choice* 102, 341-364.
- Oman C.P. (2000): *Policy Competition for Foreign Direct Investment: A Study of Competition Among Governments to Attract FDI*. OECD Development Centre Studies, Paris.
- Persson T. & G. Tabellini (1994): "Is Inequality Harmful for Growth?" *The American Economic Review*, 84 (3) 600-621.
- Political Risk Services (2007): *International Country Risk Guide (ICRG)*. Retrieved December 15, 2007 from: <http://www.prsgroup.com/>
- Przeworski A., M.E. Alvarez, J.A. Cheibub & L. Fernando (2000): *Democracy and Development: Political Institutions and Material Well-Being in the World, 1950-1990*. Cambridge University Press, New York.
- Rigobon R. & D. Rodrik (2005): "Rule of Law, Democracy, Openness and Income: Estimating the Interrelationships", *Economics of Transition* 13 (3): 533-564.
- Rodrik D. (2007): "What does the 'rule of law' mean?" November 13, 2007 weblog. http://rodrik.typepad.com/dani_rodriks_weblog/
- Rodrik D. (2008): "Second-Best Institutions", *American Economic Review, Papers and Proceedings*, May 2008.
- Rodrik D., A. Subramanian & F. Trebbi (2004): "Institutions Rule: The Primacy of Institutions Over Geography and Integration in Economic Development". *Journal of Economic Growth* 9 (2) 131-165, 06..
- Sachs J. (2003): "Institutions Don't Rule: Direct Effects of Geography on Per Capita Income". NBER Working Paper No. 9490.
- Sachs J. & J.W. McArthur (2001): Institutions and Geography: Comment on Acemoglu, Johnson, and Robinson (2000): NBER Working Paper 8114, Cambridge, MA.
- Sachs J. & A. Warner (1995): "Natural Resource Abundance and Economic Growth". NBER Working Paper 5398
- Sala-i-Martin X. (1997): "I just ran 2 million regressions". *American Economic Review* 87 (2) 178-183.
- Solow R. (1956): "A Contribution to the Theory of Economic Growth". *Quarterly Journal of Economics* 70 (1) 65-94.

- Temple J. (1998): "Robustness tests of the augmented Solow model". *Journal of Applied Econometrics* 13 (4) 361-375.
- Teorell J., H. Sören & R. Bo (2008): *The Quality of Government Dataset. Version 15 May 2008*. University of Gothenburg: The Quality of Government Institute. Available at <http://www.qog.pol.gu.se>
- The Economist (13 March 2008): "Economics and the rule of law: Order in the jungle".
- Thomas M.A. (2007): "What do the Worldwide Governance Indicators measure?" Working Paper. Baltimore, Maryland: Johns Hopkins University.
- Trebilcock M. & R. Daniels (2008): *Rule of Law Reform and Development*. Edward Elgar.
- Van de Walle S. (2006): "The State of the World's Bureaucracies". *Journal of Comparative Policy Analysis* 8 (4): 439-450
- Wooldridge J. (2002): *Econometric Analysis of Cross Section and Panel Data*, MIT Press, Cambridge, MA.
- World Bank (2007): *World Development Indicators 2007*. World Bank, Washington, D.C.
- World Bank (2008): *World Development Indicators 2008*. World Bank, Washington, D.C.
- World Bank & EBRD (2007): *Business Environment and Enterprise Performance Survey*. Washington, D.C.: World Bank.
- Zak P.J. & S. Knack (2001): "Trust and Growth". *Economic Journal* 111 (470) 295-321.

Appendix 6.A: Tables of growth equations including Rule of Law indicators

Tables 6.A.01-6.A.08 in this appendix present estimation results for growth equations including rule of law indicators. The following notes are common to all these tables.

- All regressors are lagged by one five-year period with respect to growth.
- Fixed time effects are included.
- Robust standard errors are shown in parentheses.
- Significance tests: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.
- TA denotes Time Averages.
- A superscripted ^c indicates a time-constant regressor.
- Q1 and Q3 denote the first and third quartile of the indicator, respectively.

Table 6.A.01: Growth equations estimated by Pooled OLS

	(1)	(2)	(3)	(4)
	Fraser PR	Heritage PR	ICRG L&O	WGI RoL
Panel (a): Unbalanced samples				
Rule of Law \leq Q1	-0.958*** (0.356)	0.071 (0.528)	-0.834** (0.385)	-1.035** (0.453)
Rule of Law \geq Q3	0.198 (0.588)	-0.168 (0.413)	-0.650 (0.451)	0.242 (0.509)
ln GDP per cap.	-0.716** (0.337)	0.165 (0.329)	-0.681** (0.329)	-0.027 (0.338)
Investment rate	0.107*** (0.036)	0.078** (0.034)	0.095*** (0.036)	0.073** (0.033)
Population growth	-0.218* (0.115)	-0.278** (0.120)	-0.333*** (0.127)	-0.234** (0.118)
Trade openness	0.0010 (0.0043)	0.0046 (0.0047)	0.0015 (0.0043)	0.0042 (0.0046)
Dictatorship	-0.410 (0.355)	0.274 (0.421)	-0.471 (0.381)	0.420 (0.431)
Education	0.107 (0.093)	-0.041 (0.097)	0.112 (0.092)	-0.036 (0.095)
Health	0.015 (0.010)	0.005 (0.015)	0.014 (0.010)	-0.001 (0.014)
Temperate zone ^c	0.700 (0.576)	0.640 (0.576)	0.975* (0.546)	0.627 (0.557)
Observations	332	177	332	180
R-squared	0.713	0.763	0.708	0.769
RMS Residual	2.363	1.877	2.400	1.842
Panel (b): Balanced samples				
Rule of Law \leq Q1	-1.328*** (0.361)	-0.438 (0.557)	-0.731* (0.425)	-1.403*** (0.468)
Rule of Law \geq Q3	0.349 (0.484)	-0.447 (0.454)	-0.265 (0.477)	0.198 (0.509)
ln GDP per cap.	-0.168 (0.331)	-0.023 (0.331)	-0.086 (0.330)	-0.253 (0.353)
Investment rate	0.105** (0.043)	0.110*** (0.042)	0.094** (0.045)	0.102** (0.041)
Population growth	-0.318** (0.133)	-0.428*** (0.131)	-0.406*** (0.131)	-0.381*** (0.129)
Trade openness	0.0028 (0.0047)	0.0053 (0.0046)	0.0056 (0.0046)	0.0054 (0.0045)
Dictatorship	0.277 (0.420)	0.298 (0.452)	0.0887 (0.474)	0.337 (0.454)
Education	-0.022 (0.087)	-0.021 (0.088)	-0.055 (0.091)	-0.032 (0.087)
Health	0.003 (0.014)	0.004 (0.015)	0.005 (0.015)	0.001 (0.014)
Temperate zone ^c	0.255 (0.574)	0.857 (0.583)	0.696 (0.576)	0.739 (0.555)
Observations	160	160	160	160
R-squared	0.794	0.780	0.782	0.789
RMS Residual	1.741	1.802	1.791	1.763

Table 6.A.02: Growth equations estimated by GLS (RE-TA)

	(1)	(2)	(3)	(4)
	Fraser PR	Heritage PR	ICRG L&O	WGI RoL
Panel (a): Unbalanced samples				
Rule of Law \leq Q1	-0.729** (0.350)	-0.206 (0.566)	-0.603 (0.404)	-0.461 (0.451)
Rule of Law \geq Q3	0.290 (0.624)	-0.154 (0.392)	-0.466 (0.477)	0.249 (0.507)
ln GDP per cap.	-1.253*** (0.419)	-0.139 (0.440)	-1.192** (0.410)	-0.231 (0.411)
Investment rate	0.074** (0.037)	0.085** (0.039)	0.078** (0.037)	0.080** (0.037)
Population growth	0.130 (0.126)	-0.077 (0.136)	0.077 (0.147)	-0.054 (0.133)
Trade openness (TA) ^c	-0.0029 (0.0073)	0.0064 (0.0080)	-0.0034 (0.0075)	0.0067 (0.0076)
Dictatorship (TA) ^c	-1.114 (0.688)	0.126 (0.787)	-1.292* (0.709)	0.127 (0.808)
Education (TA) ^c	0.169 (0.128)	0.111 (0.157)	0.190 (0.130)	0.102 (0.150)
Health (TA) ^c	0.043** (0.018)	-0.002 (0.023)	0.035* (0.019)	-0.004 (0.021)
Temperate zone ^c	1.034 (0.766)	0.854 (0.767)	1.369* (0.734)	0.789 (0.762)
Observations	344	184	341	187
Number of countries	88	94	87	95
RMS Residual	2.078	1.233	2.120	1.240
Panel (b): Balanced samples				
Rule of Law \leq Q1	-0.900** (0.417)	-0.704 (0.649)	-0.985* (0.525)	-0.827* (0.486)
Rule of Law \geq Q3	0.581 (0.448)	-0.563 (0.393)	0.208 (0.420)	0.240 (0.517)
ln GDP per cap.	-0.337 (0.442)	-0.304 (0.499)	-0.452 (0.473)	-0.392 (0.466)
Investment rate	0.111** (0.049)	0.109** (0.050)	0.102** (0.051)	0.105** (0.049)
Population growth	-0.240 (0.151)	-0.307** (0.146)	-0.294* (0.151)	-0.288* (0.147)
Trade openness (TA) ^c	0.0070 (0.0084)	0.0103 (0.0089)	0.0104 (0.0087)	0.0098 (0.0082)
Dictatorship (TA) ^c	0.054 (0.775)	0.143 (0.819)	-0.254 (0.874)	0.090 (0.821)
Education (TA) ^c	0.060 (0.141)	0.100 (0.157)	0.037 (0.156)	0.064 (0.147)
Health (TA) ^c	-0.002 (0.023)	0.001 (0.026)	-0.001 (0.025)	0.000 (0.023)
Temperate zone ^c	0.341 (0.800)	1.053 (0.831)	0.602 (0.842)	0.748 (0.816)
Observations	163	163	163	163
Number of countries	83	83	83	83
RMS Residual	1.212	1.174	1.175	1.214

Table 6.A.03: Growth equations estimated by GLS (RE)

	(1)	(2)	(3)	(4)
	Fraser PR	Heritage PR	ICRG L&O	WGI RoL
Panel (a): Unbalanced samples				
Rule of Law \leq Q1	-0.555 (0.347)	-0.153 (0.599)	-0.508 (0.406)	-0.510 (0.446)
Rule of Law \geq Q3	0.386 (0.617)	-0.188 (0.381)	-0.393 (0.465)	0.218 (0.504)
ln GDP per cap.	-1.328*** (0.434)	-0.181 (0.491)	-1.379*** (0.440)	-0.335 (0.476)
Investment rate	0.047 (0.039)	0.080* (0.041)	0.046 (0.039)	0.072* (0.039)
Population growth	-0.022 (0.123)	-0.154 (0.152)	-0.081 (0.147)	-0.121 (0.147)
Trade openness	0.0074 (0.0055)	0.0043 (0.0053)	0.0074 (0.0055)	0.0046 (0.0052)
Dictatorship	-0.492 (0.404)	0.311 (0.539)	-0.542 (0.444)	0.308 (0.538)
Education	0.275** (0.124)	0.061 (0.147)	0.293** (0.128)	0.061 (0.140)
Health	0.018 (0.011)	0.004 (0.019)	0.018* (0.011)	0.004 (0.018)
Temperate zone ^c	1.431* (0.755)	0.997 (0.738)	1.809** (0.741)	0.930 (0.737)
Observations	332	177	332	180
Number of countries	87	91	86	92
RMS Residual	2.025	1.241	2.035	1.237
Panel (b): Balanced samples				
Rule of Law \leq Q1	-0.819** (0.406)	-0.698 (0.650)	-0.785 (0.537)	-0.745 (0.473)
Rule of Law \geq Q3	0.638 (0.442)	-0.589 (0.396)	0.214 (0.434)	0.146 (0.510)
ln GDP per cap.	-0.539 (0.474)	-0.416 (0.517)	-0.565 (0.501)	-0.547 (0.509)
Investment rate	0.104** (0.050)	0.108** (0.051)	0.098* (0.054)	0.099* (0.051)
Population growth	-0.384** (0.150)	-0.463*** (0.152)	-0.423*** (0.155)	-0.429*** (0.148)
Trade openness	0.0046 (0.0051)	0.0057 (0.0051)	0.0056 (0.0051)	0.0061 (0.0049)
Dictatorship	0.429 (0.547)	0.506 (0.582)	0.204 (0.631)	0.396 (0.576)
Education	0.057 (0.547)	0.090 (0.582)	0.051 (0.631)	0.069 (0.576)
Health	0.007 (0.019)	0.002 (0.020)	0.006 (0.019)	0.006 (0.019)
Temperate zone ^c	0.426 (0.744)	1.174 (0.765)	0.674 (0.791)	0.870 (0.760)
Observations	160	160	160	160
Number of countries	82	82	82	82
RMS Residual	1.183	1.155	1.159	1.189

Table 6.A.04: Growth equations estimated by FE-LS (LSDV)

	(1)	(2)	(3)	(4)
	Fraser PR	Heritage PR	ICRG L&O	WGI RoL
Panel (a): Unbalanced samples				
Rule of Law \leq Q1	-0.372 (0.315)	-0.721 (0.868)	-0.022 (0.419)	-0.689 (0.489)
Rule of Law \geq Q3	-0.081 (0.499)	-0.352 (0.555)	-0.365 (0.428)	-0.758* (0.415)
ln GDP per cap.	-7.621*** (1.140)	-12.510*** (1.692)	-7.693*** (1.151)	-12.450*** (1.728)
Investment rate	-0.067** (0.033)	0.068* (0.040)	-0.047 (0.036)	0.054 (0.039)
Population growth	0.135 (0.153)	0.196* (0.113)	0.198 (0.212)	0.204 (0.113)
Trade openness	0.0340*** (0.0119)	0.0145 (0.0124)	0.0348*** (0.0122)	0.0142 (0.0124)
Dictatorship	-0.131 (0.484)	0.009 (0.518)	-0.119 (0.516)	0.007 (0.569)
Education	0.140 (0.273)	-0.119 (0.822)	0.187 (0.297)	0.287 (0.790)
Health	0.009 (0.012)	-0.002 (0.021)	0.014 (0.012)	0.013 (0.021)
Observations	351	187	341	191
R-squared	0.444	0.561	0.423	0.554
RMS Residual	1.398	0.702	1.458	0.711
Panel (b): Balanced samples				
Rule of Law \leq Q1	-0.465 (0.517)	-1.028 (0.973)	-1.514* (0.890)	-0.428 (0.454)
Rule of Law \geq Q3	0.506 (0.393)	-0.555 (0.567)	0.217 (0.340)	-0.643 (0.467)
ln GDP per cap.	-13.21*** (1.842)	-12.97*** (1.835)	-13.78*** (1.845)	-13.28*** (1.838)
Investment rate	0.093** (0.045)	0.085* (0.043)	0.086** (0.041)	0.088** (0.044)
Population growth	-0.021 (0.197)	0.004 (0.201)	-0.038 (0.176)	0.026 (0.186)
Trade openness	0.0168* (0.0100)	0.0126 (0.0103)	0.0171* (0.0095)	0.0129 (0.0106)
Dictatorship	0.307 (0.460)	0.133 (0.524)	-0.587 (0.686)	0.130 (0.544)
Education	0.332 (0.776)	0.265 (0.811)	0.535 (0.855)	0.452 (0.772)
Health	0.012 (0.021)	-0.001 (0.018)	0.013 (0.019)	0.008 (0.020)
Observations	165	165	165	165
R-squared	0.624	0.636	0.644	0.623
RMS Residual	0.654	0.643	0.637	0.655

Table 6.A.05: Growth equations estimated by Pooled IV

	(1)	(2)	(3)	(4)
	Fraser PR	Heritage PR	ICRG L&O	WGI RoL
Panel (a): Unbalanced samples				
Rule of Law \leq Q1	-0.960*** (0.358)	0.042 (0.531)	-0.859** (0.385)	-1.133** (0.463)
Rule of Law \geq Q3	0.341 (0.583)	0.0360 (0.422)	-0.527 (0.459)	0.510 (0.497)
ln GDP per cap.	-1.101*** (0.371)	-0.232 (0.318)	-1.100*** (0.360)	-0.426 (0.327)
Investment rate	0.113*** (0.037)	0.076** (0.035)	0.102*** (0.036)	0.073** (0.034)
Population growth	-0.208* (0.117)	-0.275** (0.121)	-0.327** (0.128)	-0.232* (0.120)
Trade openness	0.0011 (0.0043)	0.0045 (0.0047)	0.0014 (0.0043)	0.0041 (0.0046)
Dictatorship	-0.531 (0.360)	0.217 (0.424)	-0.607 (0.389)	0.373 (0.436)
Education	0.177* (0.095)	0.040 (0.093)	0.191** (0.095)	0.041 (0.091)
Health	0.018* (0.010)	0.010 (0.015)	0.017 (0.011)	0.003 (0.015)
Temperate zone ^c	0.831 (0.582)	0.732 (0.583)	1.124** (0.552)	0.677 (0.562)
Observations	332	177	332	180
RMS Residual	2.371	1.886	2.408	1.852
Panel (b): Balanced samples				
Rule of Law \leq Q1	-1.338*** (0.364)	-0.481 (0.560)	-0.736* (0.426)	-1.501*** (0.480)
Rule of Law \geq Q3	0.519 (0.476)	-0.245 (0.458)	-0.120 (0.475)	0.456 (0.499)
ln GDP per cap.	-0.489 (0.316)	-0.398 (0.327)	-0.426 (0.315)	-0.618* (0.347)
Investment rate	0.108** (0.042)	0.111*** (0.043)	0.097** (0.045)	0.104** (0.041)
Population growth	-0.314** (0.134)	-0.428*** (0.130)	-0.404*** (0.131)	-0.380*** (0.129)
Trade openness	0.0027 (0.0046)	0.0052 (0.0046)	0.0055 (0.0045)	0.0053 (0.0044)
Dictatorship	0.221 (0.419)	0.245 (0.453)	0.023 (0.477)	0.289 (0.456)
Education	0.033 (0.083)	0.041 (0.085)	0.007 (0.087)	0.023 (0.084)
Health	0.007 (0.014)	0.009 (0.015)	0.010 (0.015)	0.005 (0.014)
Temperate zone ^c	0.352 (0.580)	0.977* (0.589)	0.806 (0.583)	0.818 (0.559)
Observations	160	160	160	160
RMS Residual	1.748	1.811	1.799	1.771

Table 6.A.06: Growth equations estimated by RE-TA-IV

	(1)	(2)	(3)	(4)
	Fraser PR	Heritage PR	ICRG L&O	WGI RoL
Panel (a): Unbalanced samples				
Rule of Law \leq Q1	-0.758** (0.331)	-0.229 (0.502)	-0.648* (0.370)	-0.534 (0.436)
Rule of Law \geq Q3	0.353 (0.531)	-0.0817 (0.439)	-0.430 (0.457)	0.377 (0.596)
ln GDP per cap.	-1.600*** (0.337)	-0.367 (0.401)	-1.549*** (0.342)	-0.456 (0.389)
Investment rate	0.085*** (0.030)	0.086*** (0.028)	0.087*** (0.030)	0.082*** (0.027)
Population growth	0.141 (0.132)	-0.0698 (0.121)	0.086 (0.143)	-0.046 (0.118)
Trade openness (TA) ^c	-0.0023 (0.0067)	0.0065 (0.0075)	-0.0031 (0.0068)	0.0066 (0.0071)
Dictatorship (TA) ^c	-1.299** (0.657)	0.026 (0.804)	-1.480** (0.682)	0.040 (0.752)
Education (TA) ^c	0.221 (0.143)	0.153 (0.171)	0.247* (0.145)	0.140 (0.165)
Health (TA) ^c	0.047*** (0.016)	0.001 (0.018)	0.039** (0.016)	-0.001 (0.018)
Temperate zone ^c	1.154 (0.730)	0.903 (0.788)	1.491** (0.732)	0.811 (0.773)
Observations	344	184	341	187
Number of countries	88	94	87	95
Panel (b): Balanced samples				
Rule of Law \leq Q1	-0.950** (0.413)	-0.715 (0.546)	-0.995** (0.489)	-0.885* (0.485)
Rule of Law \geq Q3	0.599 (0.569)	-0.525 (0.461)	0.217 (0.509)	0.329 (0.608)
ln GDP per cap.	-0.467 (0.397)	-0.404 (0.445)	-0.559 (0.423)	-0.555 (0.431)
Investment rate	0.114*** (0.033)	0.111*** (0.034)	0.105*** (0.035)	0.107*** (0.034)
Population growth	-0.226 (0.148)	-0.300** (0.153)	-0.284* (0.152)	-0.277* (0.148)
Trade openness (TA) ^c	0.0069 (0.0078)	0.0102 (0.0085)	0.0103 (0.0083)	0.0098 (0.0081)
Dictatorship (TA) ^c	-0.019 (0.751)	0.094 (0.844)	-0.321 (0.803)	0.023 (0.777)
Education (TA) ^c	0.080 (0.164)	0.113 (0.178)	0.053 (0.175)	0.085 (0.168)
Health (TA) ^c	0.000 (0.019)	0.003 (0.021)	0.000 (0.020)	0.002 (0.020)
Temperate zone ^c	0.393 (0.825)	1.087 (0.884)	0.655 (0.875)	0.791 (0.851)
Observations	163	163	163	163
Number of countries	83	83	83	83

Table 6.A.07: Growth equations estimated by RE-IV

	(1)	(2)	(3)	(4)
	Fraser PR	Heritage PR	ICRG L&O	WGI RoL
Panel (a): Unbalanced samples				
Rule of Law \leq Q1	-0.543 (0.337)	-0.179 (0.500)	-0.519 (0.371)	-0.559 (0.446)
Rule of Law \geq Q3	0.472 (0.526)	-0.107 (0.433)	-0.333 (0.449)	0.354 (0.591)
ln GDP per cap.	-1.719*** (0.367)	-0.487 (0.419)	-1.726*** (0.382)	-0.636 (0.413)
Investment rate	0.050 (0.031)	0.078*** (0.029)	0.051* (0.031)	0.072** (0.028)
Population growth	-0.016 (0.138)	-0.155 (0.121)	-0.086 (0.151)	-0.121 (0.119)
Trade openness	0.0079 (0.0056)	0.0043 (0.0053)	0.0074 (0.0057)	0.0047 (0.0052)
Dictatorship	-0.565 (0.403)	0.277 (0.480)	-0.616 (0.424)	0.268 (0.459)
Education	0.355*** (0.130)	0.129 (0.148)	0.363*** (0.134)	0.125 (0.145)
Health	0.020* (0.010)	0.007 (0.015)	0.020* (0.011)	0.007 (0.015)
Temperate zone ^c	1.627** (0.744)	1.105 (0.739)	1.966*** (0.747)	1.001 (0.739)
Observations	332	177	332	180
Number of countries	87	91	86	92
Panel (b): Balanced samples				
Rule of Law \leq Q1	-0.833** (0.396)	-0.716 (0.520)	-0.796* (0.474)	-0.773* (0.466)
Rule of Law \geq Q3	0.685 (0.541)	-0.552 (0.447)	0.238 (0.491)	0.232 (0.592)
ln GDP per cap.	-0.715* (0.402)	-0.561 (0.439)	-0.730* (0.421)	-0.743* (0.433)
Investment rate	0.105*** (0.033)	0.108*** (0.034)	0.099*** (0.036)	0.099*** (0.033)
Population growth	-0.383*** (0.148)	-0.464*** (0.151)	-0.422*** (0.152)	-0.430*** (0.148)
Trade openness	0.0046 (0.0054)	0.0058 (0.0057)	0.0056 (0.0056)	0.0061 (0.0055)
Dictatorship	0.398 (0.478)	0.491 (0.505)	0.173 (0.506)	0.372 (0.487)
Education	0.090 (0.142)	0.118 (0.152)	0.082 (0.152)	0.103 (0.146)
Health	0.009 (0.015)	0.004 (0.016)	0.008 (0.015)	0.008 (0.015)
Temperate zone ^c	0.504 (0.773)	1.245 (0.807)	0.758 (0.815)	0.943 (0.795)
Observations	160	160	160	160
Number of countries	82	82	82	82

Table 6.A.08: Growth equations estimated by FE-IV

	(1)	(2)	(3)	(4)
	Fraser PR	Heritage PR	ICRG L&O	WGI RoL
Panel (a): Unbalanced samples				
Rule of Law \leq Q1	-0.428 (0.315)	-0.762 (0.645)	-0.034 (0.363)	-0.751 (0.518)
Rule of Law \geq Q3	-0.101 (0.506)	-0.304 (0.417)	-0.300 (0.423)	-0.732 (0.631)
ln GDP per cap.	-9.002*** (0.873)	-14.38*** (3.086)	-9.147*** (0.929)	-13.40*** (3.093)
Investment rate	0.071* (0.036)	-0.052 (0.041)	0.055 (0.036)	-0.033 (0.041)
Population growth	0.158 (0.149)	0.140 (0.138)	0.153 (0.178)	0.205 (0.146)
Trade openness	0.0346*** (0.0086)	0.0177 (0.0125)	0.0353*** (0.0091)	0.0159 (0.0127)
Dictatorship	-0.064 (0.469)	-0.141 (0.861)	-0.098 (0.503)	-0.088 (0.883)
Education	0.147 (0.291)	-0.011 (0.863)	0.211 (0.301)	0.360 (0.876)
Health	0.009 (0.011)	0.003 (0.027)	0.015 (0.012)	0.015 (0.026)
Observations	351	187	341	191
Number of countries	93	97	89	98
Panel (b): Balanced samples				
Rule of Law \leq Q1	-0.434 (0.431)	-0.998 (0.635)	-1.445** (0.672)	-0.377 (0.495)
Rule of Law \geq Q3	0.534 (0.577)	-0.608 (0.438)	0.224 (0.464)	-0.665 (0.646)
ln GDP per cap.	-12.09*** (2.690)	-11.70*** (2.683)	-12.92*** (2.625)	-12.27*** (2.774)
Investment rate	0.091** (0.043)	0.083* (0.0436)	0.087* (0.045)	0.087** (0.043)
Population growth	-0.049 (0.218)	-0.023 (0.215)	-0.057 (0.211)	0.001 (0.221)
Trade openness	0.0151 (0.0120)	0.0105 (0.0120)	0.0158 (0.0116)	0.0113 (0.0124)
Dictatorship	0.396 (0.801)	0.235 (0.791)	-0.480 (0.881)	0.226 (0.823)
Education	0.249 (0.822)	0.173 (0.808)	0.471 (0.796)	0.365 (0.840)
Health	0.008 (0.027)	-0.004 (0.026)	0.010 (0.026)	0.005 (0.027)
Observations	165	165	165	165
Number of countries	85	85	85	85

Appendix 6.B: Tables of investment equations including Rule of Law indicators

Tables 6.B.01-6.B.04 in this appendix present estimation results for investment equations including rule of law indicators. The following notes are common to all these tables.

- All regressors are lagged by one five-year period with respect to growth.
- Fixed time effects are included.
- Robust standard errors are shown in parentheses.
- Significance tests: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.
- TA denotes Time Averages.
- A superscripted ^c indicates a time-constant regressor.
- Q1 and Q3 denote the first and third quartile of the indicator, respectively.

Table 6.B.01: Investment equations estimated by Pooled OLS

	(1)	(2)	(3)	(4)
	Fraser PR	Heritage PR	ICRG L&O	WGI RoL
Panel (a): Unbalanced samples				
Rule of Law \leq Q1	-0.123 (0.565)	2.066*** (0.663)	-0.343 (0.630)	1.252* (0.642)
Rule of Law \geq Q3	-0.355 (0.666)	-0.569 (0.687)	-0.432 (0.603)	0.760 (0.712)
GDP growth	-0.037 (0.115)	-0.028 (0.150)	-0.018 (0.105)	-0.013 (0.127)
Lagged investment	0.689*** (0.055)	0.659*** (0.041)	0.667*** (0.052)	0.677*** (0.040)
Population growth	-0.256 (0.210)	-0.415** (0.195)	-0.343 (0.224)	-0.456** (0.180)
Trade openness	-0.0058 (0.0078)	-0.0109* (0.0056)	-0.0056 (0.0076)	-0.0136** (0.0059)
Dictatorship	-0.178 (0.536)	-0.273 (0.585)	0.0538 (0.554)	0.174 (0.630)
Education	-0.087 (0.115)	-0.082 (0.115)	-0.150 (0.109)	-0.178 (0.126)
Health	0.023* (0.013)	0.015 (0.021)	0.025* (0.014)	0.007 (0.021)
Temperate zone ^c	-0.369 (0.805)	0.780 (0.966)	-0.409 (0.784)	0.153 (0.824)
Observations	332	177	332	180
R-squared	0.972	0.983	0.971	0.982
RMS Residual	3.631	2.898	3.685	2.932
Panel (b): Balanced samples				
Rule of Law \leq Q1	-0.037 (0.663)	2.286*** (0.748)	-0.230 (0.681)	0.846 (0.795)
Rule of Law \geq Q3	0.964 (0.755)	-0.551 (0.761)	0.975 (0.889)	0.776 (0.747)
GDP growth	-0.069 (0.175)	-0.025 (0.168)	-0.057 (0.178)	-0.051 (0.176)
Lagged investment	0.688*** (0.068)	0.653*** (0.064)	0.685*** (0.067)	0.677*** (0.068)
Population growth	-0.557** (0.225)	-0.567*** (0.210)	-0.540** (0.240)	-0.592*** (0.222)
Trade openness	-0.0143** (0.0068)	-0.0115* (0.0065)	-0.0145** (0.0068)	-0.0141** (0.0068)
Dictatorship	0.283 (0.759)	-0.270 (0.661)	0.221 (0.764)	0.199 (0.737)
Education	-0.202 (0.135)	-0.073 (0.129)	-0.198 (0.133)	-0.180 (0.137)
Health	-0.002 (0.022)	0.016 (0.022)	-0.005 (0.022)	0.005 (0.023)
Temperate zone ^c	-0.152 (0.938)	0.468 (1.041)	-0.257 (1.002)	-0.127 (0.902)
Observations	160	160	160	160
R-squared	0.981	0.982	0.981	0.981
Number of countries	3.068	2.980	3.066	3.061

Table 6.B.02: Investment equations estimated by GLS (RE-TA)

	(1)	(2)	(3)	(4)
	Fraser PR	Heritage PR	ICRG L&O	WGI RoL
Panel (a): Unbalanced samples				
Rule of Law \leq Q1	-0.156 (0.548)	1.846** (0.775)	-0.111 (0.598)	0.576 (0.780)
Rule of Law \geq Q3	-0.280 (0.666)	-0.560 (0.734)	-0.158 (0.609)	1.100 (0.804)
GDP growth	-0.023 (0.109)	0.031 (0.158)	0.004 (0.100)	0.016 (0.133)
Lagged investment	0.701*** (0.053)	0.632*** (0.045)	0.672*** (0.050)	0.632*** (0.043)
Population growth	-0.176 (0.195)	-0.335* (0.192)	-0.261 (0.210)	-0.443** (0.190)
Trade openness (TA) ^c	-0.0025 (0.0094)	0.0010 (0.0096)	-0.0032 (0.0092)	-0.0024 (0.0097)
Dictatorship (TA) ^c	0.247 (0.714)	-0.098 (1.066)	0.523 (0.724)	0.873 (1.201)
Education (TA) ^c	-0.021 (0.133)	-0.032 (0.182)	-0.087 (0.126)	-0.096 (0.192)
Health (TA) ^c	0.0198 (0.018)	-0.005 (0.025)	0.025 (0.018)	-0.022 (0.027)
Temperate zone ^c	-0.306 (0.821)	1.051 (1.092)	-0.446 (0.809)	0.426 (0.966)
Observations	344	184	341	187
Number of countries	88	94	87	95
RMS Residual	3.652	2.841	3.707	2.796
Panel (b): Balanced samples				
Rule of Law \leq Q1	-0.485 (0.683)	1.791** (0.886)	-0.381 (0.732)	-0.006 (0.817)
Rule of Law \geq Q3	1.234 (0.872)	-0.425 (0.798)	1.156 (0.941)	1.294 (0.838)
GDP growth	-0.025 (0.183)	0.021 (0.177)	0.000 (0.186)	-0.022 (0.184)
Lagged investment	0.572*** (0.071)	0.575*** (0.069)	0.574*** (0.071)	0.559*** (0.072)
Population growth	-0.574** (0.248)	-0.531** (0.236)	-0.572** (0.257)	-0.656*** (0.239)
Trade openness (TA) ^c	-0.0072 (0.0107)	-0.0038 (0.0101)	-0.0061 (0.0106)	-0.0064 (0.0106)
Dictatorship (TA) ^c	1.192 (1.233)	0.210 (1.124)	1.059 (1.223)	1.260 (1.247)
Education (TA) ^c	-0.143 (0.210)	-0.0720 (0.198)	-0.149 (0.203)	-0.141 (0.209)
Health (TA) ^c	-0.007 (0.026)	0.014 (0.027)	-0.008 (0.026)	-0.005 (0.027)
Temperate zone ^c	-0.196 (1.119)	0.476 (1.211)	-0.184 (1.161)	-0.221 (1.070)
Observations	163	163	163	163
Number of countries	83	83	83	83
RMS Residual	2.590	2.625	2.613	2.573

Table 6.B.03: Investment equations estimated by GLS (RE)

	(1)	(2)	(3)	(4)
	Fraser PR	Heritage PR	ICRG L&O	WGI RoL
Panel (a): Unbalanced samples				
Rule of Law \leq Q1	0.068 (0.498)	0.726 (0.668)	-0.408 (0.526)	0.041 (0.774)
Rule of Law \geq Q3	-0.492 (0.415)	-0.563 (0.492)	-0.853* (0.487)	0.053 (0.517)
GDP growth	-0.030 (0.086)	-0.013 (0.094)	0.002 (0.080)	-0.056 (0.086)
Lagged investment	0.695*** (0.042)	0.572*** (0.044)	0.652*** (0.063)	0.607*** (0.054)
Population growth	-0.357*** (0.130)	-0.528*** (0.147)	-0.444*** (0.137)	-0.612*** (0.145)
Trade openness	-0.0045 (0.0052)	-0.0083 (0.0056)	-0.0031 (0.0057)	-0.0089 (0.0070)
Dictatorship	-0.456 (0.472)	0.061 (0.628)	-0.175 (0.501)	0.619 (0.733)
Observations	428	277	442	294
Number of countries	119	152	130	161
RMS Residual	3.741	2.813	3.752	2.971
Panel (b): Balanced samples				
Rule of Law \leq Q1	-0.537 (0.639)	0.905 (0.708)	-0.290 (0.637)	0.380 (0.704)
Rule of Law \geq Q3	-0.128 (0.509)	-0.182 (0.501)	0.222 (0.573)	0.330 (0.524)
GDP growth	0.045 (0.122)	0.089 (0.120)	0.057 (0.120)	0.066 (0.122)
Lagged investment	0.567*** (0.063)	0.580*** (0.062)	0.571*** (0.062)	0.575*** (0.062)
Population growth	-0.610*** (0.178)	-0.600*** (0.170)	-0.587*** (0.188)	-0.617*** (0.177)
Trade openness	-0.0104* (0.0055)	-0.0087* (0.0052)	-0.0109** (0.0053)	-0.0104** (0.0052)
Dictatorship	0.483 (0.746)	0.186 (0.691)	0.460 (0.738)	0.461 (0.733)
Observations	205	205	205	205
Number of countries	108	108	108	108
RMS Residual	2.604	2.629	2.616	2.616

Table 6.B.04: Investment equations estimated by FE-LS

	(1)	(2)	(3)	(4)
	Fraser PR	Heritage PR	ICRG L&O	WGI RoL
Panel (a): Unbalanced samples				
Rule of Law \leq Q1	-0.234 (0.680)	0.877 (0.977)	-0.051 (0.682)	-0.154 (1.352)
Rule of Law \geq Q3	-0.938 (0.879)	-0.525 (0.722)	-0.060 (0.673)	1.653* (0.898)
GDP growth	-0.0760 (0.110)	0.067 (0.113)	-0.006 (0.097)	-0.025 (0.111)
Lagged investment	0.263** (0.118)	0.227** (0.113)	0.183** (0.079)	0.221** (0.102)
Population growth	0.188 (0.314)	-0.217 (0.297)	0.006 (0.370)	-0.459 (0.410)
Trade openness	-0.0315 (0.0195)	-0.0199 (0.0216)	-0.0220 (0.0205)	-0.0112 (0.0211)
Dictatorship	-1.700 (1.051)	0.677 (1.384)	-2.079* (1.194)	0.662 (1.313)
Observations	428	277	442	294
Number of countries	119	152	130	161
R-squared	0.088	0.088	0.058	0.091
RMS Residual	2.902	1.816	2.816	1.902
Panel (b): Balanced samples				
Rule of Law \leq Q1	-1.418** (0.636)	-0.887 (1.053)	-1.124 (0.892)	-0.748 (0.639)
Rule of Law \geq Q3	0.729 (1.147)	-0.108 (0.954)	-0.756 (1.054)	2.421 (1.462)
GDP growth	0.063 (0.114)	0.088 (0.121)	0.057 (0.114)	0.088 (0.120)
Lagged investment	0.042 (0.110)	0.046 (0.110)	0.040 (0.113)	0.035 (0.105)
Population growth	-0.791** (0.369)	-0.854** (0.402)	-0.943** (0.406)	-0.909** (0.381)
Trade openness	-0.0288 (0.0253)	-0.0332 (0.0248)	-0.0339 (0.0257)	-0.0253 (0.0232)
Dictatorship	-0.718 (1.323)	-0.730 (1.319)	-1.083 (1.423)	-0.910 (1.337)
Observations	205	205	205	205
Number of countries	108	108	108	108
R-squared	0.125	0.102	0.110	0.129
RMS Residual	1.500	1.519	1.513	1.496

Chapter 7

Conclusion

7.1 Looking back

The aim of this dissertation was to assemble a number of studies involved with the analysis of multi-unit and panel data. This final chapter provides an opportunity to reflect on that encompassing exercise and to draw some general lessons from it, as well as to indicate some directions for the future. I begin with summarising the stories once again.

The areas of application were diverse. At the micro-economic level, one chapter looked at working mothers in Germany, and estimated how much their wage rates suffered from a prolonged maternal leave. The panel data methods used were designed specifically to take into account the selection of working women, the censored and endogenous character of working hours, and unobserved heterogeneity among panel members. We saw that in terms of pay per hour, working mothers can catch up even after a prolonged maternity leave. That doesn't make the maternity costless in terms of lifetime earnings, of course. There is a substantial wage loss during and immediately after the leave, and there is an opportunity cost in terms of on-the-job experience. Nevertheless, it is a positive finding that a mother is not condemned to be paid forever less than a comparable non-mother woman, or than a comparable mother who took only a minimal maternity leave.

A second chapter at the micro-economic level was concerned with the vulnerability of Russian households to income shocks, and evaluated their ability to smooth consumption expenditures, i.e., to maintain living standards in the recent period of transition from a centrally planned to a market based economic system. Attention was paid to the dynamics of adjustment, and different GMM methods were applied to deal with the complications of dynamics in panel data. We saw that households do smooth consumption expenditures to a large extent although, certainly in the Russian transition context, not as much as they would probably like. Their longer-term behaviour is not easily reconciled with the patterns evident in the cross-sectional data (or with a lifetime budget constraint, for that matter).

At the macro-economic level, one chapter estimated bilateral import functions for three small European countries. The principal aim was to assess the advantage of linking the estimation of comparable functions by an exchangeable parameter distribution hypothesis. The method belonged to the class of random coefficient regression models, however the interest was directed to estimating individual-specific bilateral parameters rather than common hyperparameters. We found that bilateral imports are sensitive to relative prices, even though this may not be visible in individual data sets. Linking comparable or collateral data sets, which contain the same kind of information for different observation units, provides more statistical power to detect tenuous relationships and distinguish their

details.

A separate chapter proposed a technique for estimating the matrix variance components of random coefficient models in a way ensuring their positive-semidefiniteness; this is necessary for the implied matrix-weighted averages to make sense. Although not widely used, the technique is not hard to apply and can deliver substantial gains at least in terms of mean squared estimation errors. We demonstrated the feasibility of this constrained estimation method and saw how it performed in a simulation experiment.

A final chapter in the global macro-economic sphere explored a collection of data bases comprising economic and institutional indicators for the widest possible panel of countries. The objective was to identify the effect of indicators designed to measure the observance and maintenance of the rule of law in medium-term growth equations. The maintenance of the rule of law is a fundamental aspect of what is understood by ‘good governance’ in current discussions on development policy, and it figures prominently among the factors thought to affect the effectiveness of aid and investments. Rule of law indicators are indeed used by public and private decision-makers as criteria in the allocation of development funds. Several estimation methods were applied, allowing for different degrees of nonlinearity, heterogeneity and endogeneity, in order to evaluate the sensitivity, or lack of robustness, of the estimated impact of the indicators to different model assumptions.

In fact, we did not find robust evidence of the crucial link, generally taken for granted, between rule of law measures and economic growth. It is hard to say whether this is due to the weakness of the theory, the flaws in the data, or the inadequate modelling. But clearly, it is easy to abuse governance indicators and other institutional data. The underlying concepts tend to be vague and diluted by aggregation; theoretical insights are tentatively general and open-ended; cause and effect remain ambiguous. Hence, this is an area of empirical research particularly vulnerable to the threat of data mining. One will screen alternative indicators and watch world maps, scatter plots or regression relationships till she finds recognisable patterns, and sometimes interpret them as confirmation of her expectations. Being under pressure to produce an interesting, convincing or simply salable result, the temptation may be hard to resist.

7.2 Pondering on the assumptions

I would like to close the circle here by referring the reader to the typology sketched at the end of Chapter 1. Allowing for the necessary extensions and generalisations, we can situate the various micro and macro multi-unit applications collected in this dissertation in the roster of Table 1.1. The degree of heterogeneity taken into account increases as we move in the table from left to right and from top to bottom. The first micro study, that of working mothers in Chapter 4, is a classical panel study using random-effects assumptions for the intercepts and common slope coefficients. Some refinements are introduced, in particular the Mundlak terms that make the random effects assumption more plausible, and interaction effects that make the common slope assumption more palatable. Nevertheless the model is firmly of class 4, in the middle of the left-most column, and the obvious way forward is to the right, on the grounds that behavioural parameters other than the intercepts are likely to be subject to inter-individual variation as well. In view of current time and space limits, such progress is left for further research.

The second micro study, concerning Russian households in the midst of institutional transition, moves one step down in the table by avoiding the random-effects assumptions for the intercepts. This is a safe move since it allows for correlation between the intercepts

and all sorts of individual characteristics, but it is of course more demanding on the data. Refinements are introduced mainly to deal with the presence of dynamics of adjustment and errors of measurement. The model belongs to class 2 and the way forward is, again, to the right. As a small step in the right direction, namely towards heterogeneity in the slopes, the model is re-estimated in Chapter 5 for subsamples of households with different levels of standard of living and poverty risk.

The first macro study used a random coefficient (or parameter distribution) assumption to link a set of bilateral import equations. This approach clearly belongs to class 5 and future progress will probably remain within that class. It should aim at better ways to estimate variance and covariance components coherently, as done here in Chapter 3; at dealing much more thoroughly with dynamics and liminal time series properties; at linking the systems of bilateral import equations across importers, etc. Technical intricacy mingles here with more philosophical and judgmental aspects.

The final macro study, on a world-wide collection of countries, has been exploring a relatively standard growth equation from a standpoint hesitating between classes 2 (FE models) and 4 (RE models). In theory class 2 is the safest choice, but the problem of a very shallow history of good-governance statistics means that very little data variation is left to base inference on after allowing for fixed effects. It seems desirable if not indispensable to find ways of exploiting the wide cross-country variation as well as the relatively minor within-country changes. Arguably, the future also lies to the right, once again because technical and behavioural parameters other than the intercepts are most likely to be subject to inter-country variation. This is an exciting direction for further theoretical and empirical research.

In principle, it is very clear that the choice of a method must be made in the light of the applicable model assumptions. In practice, a clear understanding of the nature of the data helps a lot to see what assumptions might be realistic, but there nevertheless remains a range of uncertainty with substantial room for experimenting and testing. One becomes painfully aware of this when pursuing a diversity of empirical investigations, and the last chapter is certainly a case in point. Thus we are reminded of the essential constituent parts of a good empirical model: (1) well-developed and well-formulated theoretical grounds, with due attention paid to the choice, the interpretation and the permanency of parameters; (2) carefully formulated hypotheses within this theoretical framework; (3) a measurement system producing reasonably accurate approximations of the relevant theoretical concepts; etc.

Nonetheless, as we accumulate experience with different data sets, models and methods, it becomes evident that one way to reduce the specification uncertainty is to exploit all the possible levels of variation found in data. We need modelling approaches able to reconcile data features and patterns of variation in different dimensions: cross-sectional similarity and heterogeneity; common variation and dependence between units; long-run adjustment and short-run fluctuations.

By the same token, one should not ignore the similarities that link the various (seemingly unrelated) observation units. As Mme de Staël pointed out in our opening quote, true insight follows from recognising both similitudes and dissimilitudes around. We can gain data intelligence by modelling the unity in the diversity — building both the similarity and the heterogeneity into statistical models. The higher level of complexity of the models and techniques may make specific inferences harder to obtain; yet in the end sound technique will help to unmask deep data features.

7.3 Looking ahead

Like any branch of social science, empirical economic research is made out of little steps, sometimes up and sometimes down, on the irregular staircase of progress and discovery. The real world is too complex for our analyses of it ever to be complete, definitive or exhaustive. Every issue being resolved opens the way for more unanswered questions. It is not surprising then that the complications which slowed down the analyses in the preceding chapters also made room and created perspectives for further work. Based on what has just been said, let me venture in some ideas for further research.

In micro-economic applications, an interesting guideline for further work is not only to allow for more general heterogeneity in the slope coefficients and for possibly correlated random effects, but also and ultimately to account for the features found in more aggregate data. Concretely, in the Russian household panel, we remain puzzled by the difficulty of reconciling long-term theoretical constraints with the observed behavioural patterns. We consider introducing more subtle possibilities of nonlinear adjustment and heterogeneous groups of households. Another, relatively technical concern is that we should pay more systematic attention to different sources of bias and the possibility of weak instruments. Besides, there are many critiques of the available RLMS data that we are essentially powerless to remedy. Unfortunately, access to the most recent waves of the RLMS has become prohibitively expensive.

Access to recent GSOEP waves is relatively affordable and since the surveys are done in far less remote places and under far better conditions the quality and coherency of the data base is understandably better. Given good support at the level of the data management and local knowledge, GSOEP offers excellent potential for systematic investigations into the different sources of bias, the treatment of dynamics and heterogeneity, and the consequences of weak instruments in panel data. Chapter 4 focussed on the details of one specific issue concerning an important subpopulation and there is no shortage of other issues that would lead to estimation and testing problems of methodological interest.

On the subject of linking comparable units, the main question left unattended was probably the integration in the linked estimation framework of recent developments in time series econometrics. In this respect, the synthesis that is currently under way between convergence theories in the cross-sectional and temporal dimensions is fundamental, as is the related work in the domain of nonstationary and cointegrated panels with interdependent units. The bilateral imports equations implicitly also raise questions of aggregation and adding-up, since total imports are by definition a sum of bilateral imports. An ideal model formulation and data analysis should be able to deal with all these issues.

The most concrete plans we have concern the governance and growth issue. First of all, we feel we need to understand better why the main estimated parameters (those of initial income and those of the production inputs) are what they are, and why they are fluctuating so much. Once more, it is important to gain a balanced view of the different biasing forces likely to be at work.

There are different likely sources of bias with potentially very different effects in the various specifications. At the level of the data base, probably one of the main weaknesses is the measurement of human capital; it is one of the few measures for which we did not have several alternatives available. In spite of all its pioneering merit, the original data set of Barro & Lee (1993, 2000; see references to Chapter 6) is known to have severe limitations. An improved data set due to Cohen & Soto (2007; see references to Chapter 6) has now been made available and we intend to include it in our database. In terms of econometric methodology, much more sophisticated dynamic panel methods could and perhaps should

be considered; however, they deliver few extra benefits unless strong instrumental variables are found that avoid the weak instrument trap. It is not obvious that this condition can be met.

At a more conceptual level, we do not seem to have a good understanding, much less a consensus, about the channels through which institutions in general, and rule of law in particular, may determine cross-country differences in growth and development. We would like to introduce an aspect of endogenous growth by explaining the circumstances favourable to the adoption of new technologies in the developing world. Some authors in the recent literature argue that the main impact of governance on output is via factor accumulation; Subsection 6.5.4 is a crude attempt at exploring this possibility. In this approach, whereas the production inputs may remain the immediate or proximate determinants of output and growth, the quality of governance would belong to a 'deeper' level of explanatory factors. A good strategy to test such a multi-layered theory of growth and development will without doubt be a tremendous contribution.

Samenvatting in het Nederlands

Eenheid in verscheidenheid: Micro-panels en Macro-panels

De mogelijkheid om gegevens met een panel- of panel-achtige structuur te verzamelen en te analyseren is tegenwoordig eerder regel dan uitzondering. Dank zij de vorderingen in de informatica is het bijhouden van de gegevens van een groot aantal eenheden zelfs een kwestie van routine geworden. Marktanalisten houden bestanden bij van potentiële klanten, werkgevers van werknemers, verzekeraars van verzekerden, beleggers van bedrijven, overheden van belastingbetalers, internationale organisaties van lidstaten, enz. De meeste hoofdstukken van dit proefschrift presenteren analyses van panel-gegevens, of althans soortgelijke gegevens, die betrekking hebben op een aantal vergelijkbare eenheden. De reden om ze samen te brengen is dat een gemeenschappelijke methodologie wordt gebruikt, die erop gericht is gegevensvariatie op verschillende vlakken tegelijk te onderzoeken, met name: variatie in de doorsnededimensie en variatie in de tijdsdimensie. Mijn bedoeling is om, mede door de bespreking van een aantal empirische toepassingen, de methodologische samenhang duidelijk te maken, enkele oplossingen bij te dragen tot theoretische en praktische problemen die zich stellen, en hopelijk ook de toegankelijkheid van de techniek te verbeteren voor potentiële gebruikers van panel of 'multi-unit' data sets.¹

Het oorspronkelijke begrip 'panel' betrof een groep individuele respondenten die een tijd lang werden gevolgd. De structuur van modellen en methoden geconcipeerd met het oog op individuen bleek echter eveneens van toepassing te zijn voor de studie van groepen of entiteiten zoals huishoudens, bedrijven, overheden, financiële activa, enz. De term 'panel' is dus generiek geworden en moet in brede zin worden begrepen. Als wij nu spreken van een panel of een panel-achtige structuur, dan bedoelen we eigenlijk dat de steekproeftrekking plaats heeft gevonden in verschillende dimensies, waarvan er één de tijd is.

Een essentieel aspect van panel-gegevens is dat de waargenomen entiteiten (ook wel individuen of doorsnee-eenheden genoemd) structureel van elkaar (kunnen) verschillen. Er is over de jaren heen een progressie geweest in de mate waarin het mogelijk was rekening te houden met heterogeniteit in de doorsnee-dimensie van de data. Een scala van modellen

¹Prof. Barten had dit wellicht in het Leuvens 'veel-éénheid-gegevensverzamelingen' genoemd.

en methoden heeft het licht gezien, te beginnen met het eenvoudig samenvoegen ('poolen') van waarnemingen voor de verschillende entiteiten. Vervolgens werden de 'pooling' technieken uitgebreid door rekening te houden met al dan niet systematische verschuivingen in coëfficiënten, heteroskedasticiteit, variantiecomponenten, en verscheidene vormen van stochastische afhankelijkheid. De meest algemene modellen bevatten verscheidene stochastische parameters, al dan niet gecorreleerd met de regressoren en afhankelijk van individuele karakteristieken. Op het eerste gezicht lijken de verschillen tussen al die modellen soms uiterst theoretisch of zelfs filosofisch. Zij kunnen echter ingrijpende praktische gevolgen hebben, en leiden niet zelden tot uiteenlopende conclusies. Het is daarom van belang om zich bewust te zijn van de voorwaarden waaronder specifieke methoden betrouwbare uitkomsten kunnen geven, en om in de praktijk na te gaan welke voorwaarden realiteitswaarde (zouden kunnen) bezitten.

In de opeenvolgende hoofdstukken van dit proefschrift worden verschillende typen panel-gegevens bekeken. We bespreken de gemeenschappelijke en tegengestelde karakteristieken van micro-panels en macro-panels, de methodologische aspecten, en de praktische oplossingen voor een aantal frequente problemen. Twee toepassingen in de micro-economische en twee in de macro-economische sfeer worden gepresenteerd; de micro-economische betreffen een grote steekproef van huishoudens of personen, de macro-economische een kleine of minder kleine verzameling landen. In alle gevallen is slechts een beperkt aantal jaren beschikbaar; de tijdsdimensie wordt nog een stuk kleiner wanneer de jaren ook nog eens geaggregeerd worden tot vijf- of tienjarige periodes, zoals in de analyse van de lange-termijngroei.

In elke toepassing besteden wij aandacht aan het oplossen of neutraliseren van een of meer problemen die zich voordoen in onze data: het schatten van de componenten van een samengestelde variantie-covariantie matrix; selectiviteit, heterogeniteit en endogeniteit in inkomensgegevens van werkende moeders; persistentie en dynamiek in de consumptie-uitgaven van huishoudens die zijn blootgesteld aan grote inkomensschokken; fragiliteit van groeivergelijkingen, en van het effect van de kwaliteit van het staatsbestel voor de groei van een land op lange termijn. De inhoudelijke conclusies die men kan trekken zijn doorgaans gevoelig voor uitbreiding van de data in de doorsnee-breedte. Soms moeten wij zelfs leren leven met het feit dat eerder getrokken conclusies niet worden versterkt of verscherpt, maar integendeel worden afgezwakt of ontkracht.

De rode draad door de hoofdstukken heen, en onze voornaamste doelstelling, is het ontwikkelen en beproeven van methoden die gegevensvariatie op meerdere niveaus verenigen om zo de inhoudelijke conclusies aan te kunnen scherpen. Door informatie op verschillende niveaus te combineren wordt het mogelijk om het gedrag van economische agenten vanuit nieuwe invalshoeken en in groter detail te bekijken, te beschrijven en te verklaren. Bovendien wordt het maken van een opportunistische keuze op het methodologische vlak ontmoedigd. Zo wordt bijvoorbeeld vermeden dat men de modelkeuze laat bepalen door eenmalige waarnemingsuitkomsten.

* * *

Hoofdstuk 2 is van macro-economische aard. De bedoeling is een praktische methode te ontwikkelen om de spreiding te verkleinen van de puntschattingen voor vergelijkbare parameters. Standaard-schatters overdrijven deze spreiding en door de puntschattingen naar elkaar toe te trekken kan hun precisie verbeterd worden. De stroomlijning van effecten kan eveneens van belang zijn voor de constructie van bruikbare simulatiemodellen. Om het doel te bereiken worden stochastische parameters opgevoerd die in statistische zin 'uitwisselbaar' zijn tussen landen onderling, en wordt een verbonden schattingsprocedure

voorgesteld. De steekproefverdeling van de verbonden schatters wordt beschreven, alsmede een toets van de veronderstelde parameterverdeling. Deze benadering is verwant met een bekend principe voor de schatting van stochastische effecten, namelijk ‘Beste Lineaire Onvertekende Voorspelling’ (‘Best Linear Unbiased Prediction’; onvertekendheid wordt in het Nederlands ook wel ‘zuiverheid’ genoemd). De nieuwigheid zit dus niet zozeer in de modelformulering noch in de manier om schatters te wegen en te verbinden, dan wel in de uitbreiding van het toepassingsbereik. Bij het onderzoek van doorsnee-eenheden die niet identiek zijn maar wel van dezelfde aard, valt het aan te bevelen een gemeenschappelijk theoretisch kader te gebruiken en de analyse op gezamenlijke, ‘verbonden’ wijze uit te voeren. Door een combinatie van verwante schattingsproblemen te bestuderen wordt de basis voor de statistische inductie verbreed en versterkt.

Internationale handel is niet meer weg te denken uit meerlandenmodellen. Een bilaterale-handelsmodel wordt gebruikt om de verbonden-schattingstechniek empirisch te illustreren. Een eenvoudige bilaterale invoervergelijking met inkomens- en prijseffecten wordt gespecificeerd en geschat voor drie kleine, open economieën, nl. België, Nederland en Denemarken. Vanwege de eis van continuïteit in de definitie en constructie van de grootheden was slechts een korte steekproefperiode beschikbaar. Onder die omstandigheid kan men, zelfs in een heel eenvoudig model, verwachten geconfronteerd te worden met deels onaanneembare, zonderlinge of afwijkende schattingen. Dit is ook hier het geval. Men kan zien hoe de afwijkende schattingen met grote standaardfouten weer ‘in het gareel’ worden gebracht ten opzichte van de grotere groep schattingen met kleinere standaardfouten. Om de proef op de som te nemen worden voorspellingen gemaakt en met de feitelijke uitkomsten vergeleken. De voorspelfouten blijken ongeveer een kwart kleiner te zijn als we verbonden schatters gebruiken in plaats van de traditionele, afzonderlijke schatters (een kwart kleiner, in termen van de wortel van het kwadratische gemiddelde van de voorspelfout; d.i. minstens 40% in termen van het kwadratische gemiddelde zelf).

Een praktische moeilijkheid in deze toepassing was de schatting van de variantiecomponenten of, om precies te zijn, de componenten van de variantie-covariantiematrix. Dit is een bekend probleem dat zich in het algemeen voordoet bij modellen met variantiecomponenten of met stochastische coëfficiënten. In die modellen zijn schatters nodig van de covariantiematrix-componenten die niet alleen statistisch consistent zijn, maar ook logisch verenigbaar met de modelstructuur. Dit is het onderwerp van Hoofdstuk 3, geschreven in samenwerking met Geert Dhaene.

Zelfs in een correct gespecificeerd model zal steekproefvariatie leiden tot een proportie gevallen waar de gebruikelijke zuivere variantieschatters negatieve waarden impliceren voor een variantiecomponent. Dit geldt ook meer in het algemeen: schattingen van additieve componenten van variantie-covariantiematrices zullen vaak niet positief semidefinit zijn. In het geval van een scalaire variantiecomponent (bv. bij één enkele stochastische coëfficiënt) ligt het voor de hand om een negatieve variantieschatting te vervangen door nul. Het geval van een covariantiematrix-component (bv. bij een stochastische coëfficiënten-vector) is ingewikkelder. We bestuderen dit geval en stellen voor dit probleem een nieuwe oplossing voor.

De basisgedachte is de eenvoudige behandeling van het scalaire geval te veralgemenen door aan een covariantiematrix-schatting de restrictie op te leggen dat zij positief semidefinit moet zijn. Onvertekendheid wordt slechts opgelegd in de mate dat zij niet in conflict komt met deze beperking. We leiden een schattingsformule af die dit dilemma oplost. De uitkomst ervan blijkt nooit op een grotere afstand van de populatie waarde te liggen (in een aangepaste metriek) dan de gebruikelijke onvertekende schatter. Zij wordt verkregen door nullen te substitueren voor de negatieve wortels in een veralgemeende

spectrale decompositie van de onvertekende schatter. Zij kan ook berekend worden door een eenvoudige iteratieve procedure. Wij presenteren experimentele gegevens over de werking van de nieuwe formule. Het mag verrassend worden genoemd dat in recente jaren niet meer gebruik is gemaakt van deze benadering voor het schatten van covariantiematrix-componenten, en overigens voor andere schattingsproblemen met nevenvoorwaarden in de econometrie.

In Hoofdstuk 4 gaan we over tot de analyse van een micro-economisch panel, de Duitse SOEP of 'GSOEP'. De focus van deze analyse is de loonachterstand die jonge werkende moeders op de Duitse arbeidsmarkt oplopen, vooral wanneer zij een lang zwangerschapsverlof nemen. Tot op zekere hoogte kunnen jonge moeders zelf beslissen of zij willen werken en hoeveel uren, en hoelang zij met moederschapsverlof blijven. Om de loonachterstand met enige betrouwbaarheid te kunnen meten was het dus nodig een correctie in te bouwen voor zelfselectie (die ook weer afhankelijk is van het loonniveau). Bestaande schattingen veronachtzamen doorgaans de keuze van de werkduur en laten achterstanden zien die oplopen tot wel 30%, bovendien zonder dat er nadien een inhaalbeweging optreedt. Geprikkeld door deze hoge penaliteit, en tevens door de vloed aan evidentie over blijvende loondiscriminatie, hebben wij willen nagaan of dergelijke bevindingen een voorzichtige toets aan de GSOEP data zouden doorstaan.

In dit hoofdstuk passen we daarom recente panel-data methoden toe om de waarschijnlijke problemen van selectiviteit, heterogeniteit en endogeniteit aan te pakken. Selectiviteit treedt op omdat geen loon wordt waargenomen voor (vrouwelijke) werknemers met (moederschaps- of ander) verlof. Heterogeniteit nemen wij op in de vorm van niet-waargenomen individuele effecten die mogelijk zijn gecorreleerd met verklarende variabelen. Endogeniteit is te wijten aan de simultane bepaling van lonen en participatiebeslissingen. De klassieke behandeling van selectiviteit, waarvoor Heckman beroemd is geworden, kan worden uitgebreid door rekening te houden met heterogeniteit en simultaneiteit. Wij presenteren zo'n uitbreiding voor het geval van een 'tobit-censored' model van de participatie, en passen het toe op de feitelijke werktijdgegevens die in GSOEP beschikbaar zijn. We onderzoeken eveneens de gevoeligheid van de resultaten voor de keuze van de methode. Wat wij vinden is dat het loontekort inderdaad bestaat maar bij lange na niet zo groot is als volgens eerder gepubliceerde schattingen. Bovendien is dit tekort niet even persistent als andere studies impliceren. Vijf jaar na de loopbaanonderbreking hebben Duitse moeders andere werknemers weer ingehaald. Omdat deze betrekkelijk optimistische conclusies verschillen van die van andere auteurs, verwachten wij dat dit onderwerp in de toekomst verder aandacht zal blijven trekken. Dit onderzoek werd uitgevoerd in samenwerking met Bianca Buligescu, Gülçin Menteşoğlu en Raymond Montizaan.

Met de volgende dataset reizen wij verder naar het oosten, tot in Rusland. In dit onderzoek beschouwen wij huishoudens als verbruikers van goederen en diensten eerder dan als aanbieders van arbeid. Het is in Rusland na 1990 een tijd van politieke omwenteling en transitie van het hele economische systeem. Behalve de nieuwe systemische onzekerheden moesten Russische huishoudens bovendien leren leven met een lange reeks van eigen risico's: ziekte, invaliditeit of overlijden van een gezinslid, criminaliteit, werkloosheid en mislukte oogsten... Armoedestatistieken geven aan dat zeer veel huishoudens niet in staat waren hun eerdere consumptiepatroon te handhaven, en dat ontbering geen zeldzame ervaring was. Hoe trokken zij hun plan in deze volatiele omstandigheden? Wij interesseren ons hier voor de aanpassingstrategie die reeds impliciet aanwezig was in de permanente-inkomenshypothese van Milton Friedman in de jaren '50: "consumption smoothing", de intertemporale afvlakking of demping van de consumptie-uitgaven. Wij willen onderzoeken in welke mate Russische huishoudens, geconfronteerd met inkomensschokken tij-

dens de transitieperiode, in staat waren daar het hoofd aan te bieden en hun consumptie-standaarden (tijdelijk) te handhaven; hoe zij hun uitgaven aanpasten op de langere termijn; en welke specifieke factoren invloed kunnen hebben gehad op hun dempingvermogen.

Een flexibel en veelzijdig aanpassingsmechanisme dat eerder typisch is voor macro-economische modellen staat bekend als het 'Equilibrium Correction Model' (ECM; door Prof. Barten in het Nederlands 'bijstuurmodel' genoemd). Er is geen fundamentele reden om hetzelfde mechanisme niet toe te passen wanneer men dynamische aanpassingen wil modelleren in een micro-economische context. Dit was de opzet van een gezamenlijk onderzoek met Geranda Notten van een groot huishoudpanel, de zogenaamde Russian Longitudinal Monitoring Survey (RLMS). Hoofdstuk 5 behandelt de specifieke vraagstelling tot op welke hoogte het Russische huishoudens is gelukt om hun consumptie-uitgaven af te schermen voor inkomensschokken tijdens een groot deel van de transitieperiode. Tevens gaan wij na hoe de mogelijkheid om schokken in de consumptie-uitgaven te dempen gerelateerd is aan het armoederisico. Het ECM staat ons toe om de korte- en de langetermijndynamiek van de uitgaven voor een deel te ontwarren. Wij besteden tevens aandacht aan problemen van heterogeniteit tussen huishoudens en imputatiefouten in de inkomensgegevens.

Onze bevinding is dat huishoudens tijdens de transitieperiode in staat waren het grootste deel van hun *voedingsuitgaven* te stabiliseren, vooral dan op de korte termijn. We kunnen echter geen stellige conclusies trekken wat betreft de overige ('nonfood') uitgaven door ongunstig uitvallende specificatietoetsen. Er zijn slechts beperkte aanwijzingen van systematische verschillen in dempingvermogen tussen huishoudens. We stellen in het bijzonder vast dat een laag dempingvermogen niet noodzakelijk gepaard gaat met een verhoogd armoederisico.

Tenslotte vestigen wij de aandacht op één van de voornaamste macro-economische vraagstukken van vandaag. Er blijkt in de internationale politiek een consensus tot stand gekomen te zijn, dat de kwaliteit van de staatsinstellingen en van het staatsbestuur van wezenlijk belang is voor de economische groei en ontwikkeling van een land. Bovendien handelt men alsof die kwaliteit objectief waarneembaar en meetbaar zou zijn. Er is hierover sedert 1999 een zeer bekende en invloedrijke reeks van Wereldbankrapporten samengesteld door Daniel Kaufmann, Aart Kraay en medewerkers. Hoe solide is dit soort bevindingen met betrekking tot de keuze van bestuursindicatoren, databronnen, aggregatiemethodologie, en statistische specificaties? Kan een landenpanel met waarnemingen over enkele tientallen jaren, en een aantal strijdige bestuursindicatoren die enkel betrekking hebben op de allerlaatste jaren, volstaan om robuuste conclusies te trekken? Kan het begrip van recursiviteit in de tijd helpen om het endogeniteitsprobleem op te lossen?

Hoofdstuk 6, de vrucht van een samenwerkingsverband met Christiane Arndt, heeft tot doel de interactie te meten tussen de kwaliteit van het overheidsbestel en de economische groei in een wereldwijd panel van landen. Een databank werd opgesteld met gegevens uit een groot aantal bronnen, waaronder de bekende Heston-Summers Penn World Tables (PWT 6.1) en de Quality of Governance collectie (QoG Institute, University of Gothenburg), beide vrij on-line te verkrijgen. Wij verzamelden in het bijzonder een waaier van indicatoren om het respect voor de wet ('Rule of Law') te meten. We probeerden een aantal manieren uit om het effect hiervan in groeivergelijkingen in te bouwen. We gingen de fragiliteit van de resultaten na, hun gevoeligheid voor de keuze van de indicatoren, de tijdperiode, en de details van de specificatie. Met heterogeniteit werd rekening gehouden in de vorm van niet-waargenomen stochastische effecten, mogelijk gecorreleerd met regressoren, volgens de suggestie van Mundlak zo'n 30 jaar terug. Er werd recursiviteit in de tijd

aangenomen om het effect van de kwaliteit van het staatsbestuur in de laatste decennia te kunnen identificeren. We vonden dat de resultaten uiterst fragiel bleven. De keuze van de institutionele indicatoren blijkt de significantie te bepalen, niet enkel van de indicatoren die betrekking hebben op het respect voor de wet, maar ook van de bijkomende controlevariabelen.

Dit was de laatste empirische studie van het proefschrift. Hoewel het een gebied is waarin de laatste jaren enorm veel werk is verricht, lijkt de stroom van publicaties over de determinanten van groei en ontwikkeling alleen maar verder aan te zwellen. Wij hopen met dit en ander onderzoek hieraan een bijdrage te leveren.

* * *

Om de verschillende gebruikte modellen in een samenhangende context te plaatsen wordt een modeltypologie voor 'multi-unit' studies voorgesteld in het begin van dit proefschrift; zie Tabel 1.1. Mits voor de hand liggende uitbreidingen en veralgemeningen zijn de net geschetste toepassingen goed in die typologie te situeren. Hoe meer we ons in de tabel verplaatsen van boven naar beneden en van links naar rechts, hoe meer heterogeniteit deel gaat uitmaken van het model.

De eerste micro-economische studie, betreffende het effect van moederschapsverlof op lonen in Hoofdstuk 4, is een klassiek panelonderzoek gebaseerd op de zogenaamde 'Random Effects' (RE) aanname. Deze aanname laat het bestaan toe van stochastische individuele effecten (eigenlijk intercepten) die aan welbepaalde onafhankelijkheidsvoorwaarden voldoen. Dit betekent dat de intercepten van de geschatte verbanden een kansverdeling hebben die niet samenhangt met de waargenomen individuele karakteristieken opgenomen in het model. Alle overige coëfficiënten (hellingscoëfficiënten) zijn bij veronderstelling identiek voor alle panel-leden. Er wordt nog een maatregel getroffen om de RE voorwaarde aannemelijker te helpen maken. Naast een aantal interactietermen worden ook zogenaamde Mundlak-termen opgenomen om mogelijke correlatie tussen de individuele stochastische effecten en het initiële of gemiddelde niveau van bepaalde grootheden op te vangen. Het model is te plaatsen in type 4, middenin de linkse kolom van Tabel 1.1. De richting voor vooruitgang is waarschijnlijk naar rechts te zoeken, want andere gedragsparameters dan de intercepten zouden ook van individu tot individu kunnen verschillen. Deze mogelijkheid wordt voor toekomstig onderzoek bestemd.

De tweede micro-economische studie betreft Russische huishoudens tijdens de transitieperiode van een plangestuurd naar een marktgericht systeem. De RE aanname (of onafhankelijkheidsvoorwaarde) omtrent de intercepten wordt hier vermeden en we dalen een rij af in de typologie van Tabel 1.1, naar modellen van het 'Fixed Effects' (FE) type. Dit laat de mogelijkheid toe van correlatie tussen de intercepten en waargenomen karakteristieken, maar is veeleisender in termen van de gegevens. Verfijningen zijn mogelijk door rekening te houden met dynamische aanpassingsprocessen en tevens met meetfouten. Het model behoort tot het type 2 in de tabel en de vooruitgang moet in rechtse richting worden gezocht, naar meer heterogeniteit in de hellingscoëfficiënten. In Hoofdstuk 5 wordt een kleine beweging in die richting gedaan, door het consumptiemodel te herschatten voor deelpopulaties van huishoudens met verschillende levensstandaarden en armoedekansen.

De eerste macro-economische studie veronderstelt de geldigheid van een parameterverdeling om een stelsel bilaterale invoerfuncties met elkaar te verbinden. Deze aanpak behoort duidelijk tot het type 5 in Tabel 1.1, en toekomstige verbeteringen zullen waarschijnlijk binnen deze klasse plaats blijven vinden. Eén punt van verbetering betreft het schatten van coherente variantie- en covariantiecomponenten, en een concrete oplossing

hiervoor wordt (zoals reeds gemeld) uitgewerkt in Hoofdstuk 3 van dit proefschrift. Andere noemenswaardige punten zijn een betere behandeling van de dynamiek en liminale tijdreeks eigenschappen, een verbinding van de stelsels van bilaterale invoervergelijkingen over de invoerlanden onderling, enz.

De tweede en laatste macro-economische studie bestudeert een traditionele groeivergelijking voor een zo volledig mogelijk landenbestand. Modellen van het type 2 (FE) en modellen van het type 4 (RE) worden naast elkaar geschat. In theorie is type 2 de veiligste keuze. De zeer korte geschiedenis van gegevensverzameling omtrent 'goed bestuur' ('good governance') stelt echter een probleem, daar er slechts weinig variatie in de data overblijft na het afrekenen met FE. Het lijkt wenselijk, zoniet onontbeerlijk, om naast de beperkte binnenlandse schommelingen ook en vooral de grote internationale verschillen te benutten en te verklaren. Toekomstige ontwikkelingen zullen wellicht ook in rechtse richting gaan, aangezien behalve de intercepten ook andere technische en gedragsparameters hoogstwaarschijnlijk onderhevig zijn aan internationale verschillen. Dit is een actief en veelbelovend terrein voor verder theoretisch en empirisch onderzoek.

* * *

Hopelijk zal de thematische diversiteit van de hoofdstukken niet verhinderen dat de eenheid van beginselen en doelstellingen in dit proefschrift goed zichtbaar blijft. Het overheersende principe bestaat uit het verenigen, gelijktijdig benutten en gezamenlijk verklaren van de aanwezige variatie in metingsgegevens op verschillende vlakken: de doorsnede en de tijdsdimensie. Deze gedachte ligt aan de basis van de methodologische ontwikkelingen voor de analyse van panels of panel-achtige gegevens.

De realiteitswaarde van de gemaakte specifieke veronderstellingen, niet de algemene principes of de algebra, maken de micro- en macro-panels verschillend van elkaar. In principe is het duidelijk dat de keuze van een methode moet volgen uit de toepasselijke modelveronderstellingen. In de praktijk zal een goed begrip van de aard van de gegevens helpen om realistische aannamen te herkennen, maar niettemin blijft er een grote marge van onzekerheid over en veel ruimte voor aftastende experimenten. Daarbij zal men zich moeten hoeden voor de neiging om allerlei beschikbare formules toe te passen zonder grondig de assumpties te evalueren en zo mogelijk te toetsen waarop ze (vaak impliciet) gebaseerd zijn. Door de veelheid van voorgeprogrammeerde opties en de overvloed aan reken capaciteit kan soms in een mum van tijd een grote hoeveelheid computerooutput automatisch worden geproduceerd. Dit kan er wel eens toe leiden dat het moeilijk wordt een geloofwaardige keuze te maken tussen de talrijke schattingen en toetsen.

Om goed om te gaan met de onzekerheid in de specificatie zijn de essentiële ingrediënten: theorievorming, een overwogen keuze van interpreteerbare en stabiele parameters, zorgvuldig geformuleerde hypothesen, een adequate meting van de theoretische concepten, enz. De in dit proefschrift toegepaste technieken laten zien dat er complementaire mogelijkheden bestaan wanneer verschillende niveaus van variatie in gegevens kunnen worden waargenomen. Het is daarom van belang om bij modelbouw aandacht te besteden aan het vergelijken en verzoenen van de variatiepatronen en kenmerken op verschillende vlakken: gelijksoortigheid en heterogeniteit; gemeenschappelijke variatiefactoren en afhankelijkheid tussen doorsnee-eenheden; lange-termijn aanpassingen en korte-termijn schommelingen.

Gelijksoortigheid is bruikbaar. Wij kunnen inzichten winnen uit gegevens door de éénheid in de verscheidenheid te modelleren — door gelijksoortigheid en heterogeniteit allebei in te bouwen in statistische modellen. De hogere graad van complexiteit in de modellen en in de technieken zal het analytische werk verzwaren; maar uiteindelijk zal de degelijke techniek helpen om de verborgen structuur in de data te ontdekken.

Curriculum vitae

Denis de Crombrugge was born on 17th April 1954 in Genk, Belgium. He studied classical humanities (St-Andriesabdij, Brugge, 1967-1972), economic and social sciences (F.U.N.D.P., Namur, 1972-1974, Candidat), economics (K.U. Leuven 1974-1976, Licentiaat; 1976-1977, Doctorandus, magna cum laude), and business administration (University of Chicago, 1977-1978, M.B.A., straight A).

In August 1978, he became a research associate and teaching assistant to Prof. Anton P. Barten at the department of economics of the K.U. Leuven. In this position, he was responsible for the development of COMET III, a macro-economic, multi-country, medium-term model used for policy evaluation purposes by the European Commission (at the time the D.G. II-B). Further, until 1986, he fulfilled various undergraduate and graduate teaching duties in econometrics, and research tasks on methodological and empirical issues in multi-country model building. During 1981-1982, he was on leave as a British Council scholar at the London School of Economics and Political Science (LSE, London). For some months in 1986, he was employed by the European Institute for Advanced Studies in Management (E.I.A.S.M., Brussels) to start up the European leg of a Harvard University project entitled "The Competitive Advantage of Nations", under the direction of Prof. Michael E. Porter (Harvard Business School).

In December 1986, he was appointed as an assistant professor and lecturer in econometrics and statistics at the Maastricht University School of Business and Economics (then still called *Faculteit der Economische Wetenschappen, Rijksuniversiteit Limburg*). He has been teaching various courses and giving counsel in theoretical as well as applied econometrics (probability theory, mathematical statistics, empirical econometrics, forecasting, market research, empirical analysis of financial markets, econometric methods for cross-sectional and panel data, econometric analysis of survey data, etc.) As a researcher, he has continued studying methodological aspects of statistical inference and model building for data concerning multiple units and containing both geographical and temporal dimensions of variation.



