10.14754/CEU.2017.05

THE ECONOMETRICS OF LINEAR MODELS FOR MULTI-DIMENSIONAL PANELS

LASZLO BALAZSI

Submitted in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy at
Central European University

Budapest, Hungary

Supervisor: Laszlo Matyas

March, 2017

10.14754/CEU.2017.05

CENTRAL EUROPEAN UNIVERSITY
DEPARTMENT OF ECONOMICS

The undersigned hereby certify that they have read and recommend to the Department of Economics for acceptance a thesis entitled **"The Econometrics of Linear Models for Multi-dimensional Panels"** by **László Balázsi**.
Dated: March 28, 2017

I certify that I have read this dissertation and in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.
Chair of the Thesis Committee:

Julius Horváth

I certify that I have read this dissertation and in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.
Advisor:

László Mátyás

I certify that I have read this dissertation and in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.
Internal Examiner:

Gábor Kézdi

I certify that I have read this dissertation and in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.
External Examiner:

Patrick Sevestre

I certify that I have read this dissertation and in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.
Internal Member:

Miklós Koren

I certify that I have read this dissertation and in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.
External Member:

Szentpéteri Ádám

Ádám Szentpéteri

10.14754/CEU.2017.05

I certify that I have read this dissertation and in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.
External Member:

András Sugár

10.14754/CEU.2017.05

CENTRAL EUROPEAN UNIVERSITY
DEPARTMENT OF ECONOMICS

Author: **Laszlo Balazsi**
Title: **The Econometrics of Linear Models for Multi-dimensional Panel Data**
Degree: **Ph.D.**
Date: **March 28, 2017**

Hereby I testify that this thesis contains no material accepted for any other degree in any other institution and that it contains no material previously written and/or published by another person except where appropriate acknowledgement is made.

Signature of the author: ______________________
Laszlo Balazsi

10.14754/CEU.2017.05

DISCLOSURE OF CO-AUTHORS CONTRIBUTION

Title of the work: The Estimation of Multi-dimensional Fixed Effects Panel Data Models (Chapter 1)
Co-authors: Laszlo Matyas and Tom Wansbeek

The nature of the cooperation and the roles of the individual co-authors and the approximate share of each co-author in the joint work are the following. The paper was developed in cooperation with professors Laszlo Matyas and Tom Wansbeek. My contribution was the derivation of most of the formulas, estimators, properties, and some of the writing. Prof. Matyas lead the work, worked out the overall methodology and did some of the writing, while Prof. Wansbeek elaborated the dynamic models, and did some of the matrix representations. Some sections in this chapter, as indicated in the dissertation, are solely my own work.

Title of the work: Modelling Multi-dimensional Panel Data: A Random Effects Approach (Chapter 2)
Co-authors: Laszlo Matyas, Badi H. Baltagi, Daria Pus, Mark Harris and Felix Chan

The nature of the cooperation and the roles of the individual co-authors and the approximate share of each co-author in the joint work are the following. The paper was developed in cooperation professors Laszlo Matyas, Badi Baltagi, Mark Harris and Felix Chan, and Daria Pus. My contribution was the derivation of most of the formulas, estimators, properties as well as providing some of the methodology and some of the writing. Prof. Matyas worked out the overall methodology and did some part of the writing, Daria Pus worked on earlier versions of the study and derived some starting formulas, while Prof. Badi Baltagi gave useful comments and reviewed the paper at various stages. Professors Harris and Chang's contribution is limited to the issue of endogeneity (as it is coming from a separate work), where they were more pronounced in working out the overall methodology, and providing some of the formulas, while I was more pronounced in constructing the estimators, tests, deriving the formulas and doing some of the writing of the study. Some sections in this chapter are also my sole works.

Title of the work: Contemporaneous and Lagged Wage Returns to Foreign-Firm Experience – Evidence from Linked Employer-Employee Data (Chapter 4)
Co-authors: Istvan Boza and Janos Kollo

10.14754/CEU.2017.05

The nature of the cooperation and the roles of the individual co-authors and the approximate share of each co-author in the joint work are the following. The paper was developed in cooperation with Istvan Boza and Dr. Kollo. My contribution was the identification and calculation of the contemporaneous wage gap, and the discussion of the estimation issues with fixed effects models and the related technical issues. Istvan Boza assisted with the data management, worked out some part of the methodology, and identified and calculated the spillover effect, while Dr. Kollo managed the data, worked out the structure of the paper, the methodology and the identification of the various effects, calculated the lagged wage effect, and wrote most of the text.

10.14754/CEU.2017.05

Abstract

Recent advances in information technology have been constantly bringing down the barriers of collecting and managing data sets with sizes and representativeness unimaginable before. These data sets are typically arranged in the forms of panels, comprising tens of thousands, perhaps millions of entities, observed over a long time span. The new ways of data management, the comprehensive registry of transactions and other activities, and the attempts at the international harmonization of the data lead to the massive presence and direct accessibility of multi-dimensional panels.

The econometrics of standard, two-dimensional panel data is well-developed: it has been the subject of practically limitless research in the past fifty-sixty years. As much as efforts devoted to two-dimensional panels are admirable, multi-dimensional panels challenge analysts in several new ways. First, two-way models and toolsets are usually insufficient to fully describe and address problems in this three-dimensional context, where the unobserved heterogeneity can take on several new and interesting forms. Second, various new or existent, but increasingly present, data-related issues emerge, like feasibility of the estimators due to the sheer size of the data, incompleteness of observations, variable index deficiencies, or the large number of economically feasible model specifications.

Despite the massive presence of multi-dimensional data sets, the econometrics of three-dimensional panels remains grossly underdeveloped. Luckily, an increasing number of econometricians understand its importance, and aid empiricists with menus of modelling techniques and estimators capable of extracting the excess information embedded in the data. This thesis contributes to the literature by collecting several appealing model formulations, fixed effects, random effects and varying coefficients models, and proposing suitable estimation techniques. The comprehensiveness of the results lies in the diversity of issues discussed (both theoretical and data-related), and the fact that most techniques are feasible in practice and so have a strong potential for empirical use.

10.14754/CEU.2017.05

Chapter 1: The Estimation of Multi-dimensional Fixed Effects Panel Data Models

Sections 1.2–1.6 are joint works with Laszlo Matyas and Tom Wansbeek, Sections 1.7 and 1.8 are solely my own.

The first chapter of the thesis formulates the excess heterogeneity in the data with fixed, observable parameters. Several such three-dimensional fixed effects models are collected from the literature, all of which correspond to empirically relevant cases. The models are estimated with Least Squares Dummy Variable (LSDV) estimator. In order to prevent the joint estimation of possibly (hundreds of) thousands of parameters, the estimators are also expressed separately for each model parameter. It is also shown that the so-called *Within estimator*, which first wipes out the fixed effect parameters with a linear transformation, then performs a Least Squares on the transformed model, is numerically equivalent to the LSDV. The Within estimator reaches estimates at no costs, as long as the data at hand is complete. Typically, however, the data contains "holes". It is discussed how the Within estimator alleviates the dimensionality issue (the high cost of the estimation) completely, for structured incompleteness (like the no self-flow phenomenon), and partially, when it comes to handling incompleteness in general. This chapter also contributes to the literature by considering dynamic autoregressive specifications with fixed effects, first, by showing how the presence of various lags of the dependent variable violates the consistency of the Within estimator, then, by proposing Arellano-Bond-type instrumental variable estimators to correct for the arising inconsistency. Somewhat surprisingly, not all three-way model specifications carry this asymptotic bias. Eventual heteroscedasticity and the cross-correlation of the disturbance terms are also accounted for by proposing appropriate Feasible Generalized Least Squares (FGLS) estimators. The chapter ends with a generalization to four- and higher-dimensional fixed effect models, and intuitively argues that the results of the study can easily be generalized to any fixed effects specifications in any dimensions.

Chapter 2: Modelling Multi-dimensional Panel Data: A Random Effects Approach

Sections 2.2–2.4 are joint works with Badi H. Baltagi, Laszlo Matyas and Daria Pus, Sections 2.5 and 2.6.2–2.6.3 are joint works with Mark N. Harris, Felix Chan and Maurice Bun, Sections 2.6.1 and 2.7 are solely my own.

The second chapter of the thesis proposes several random effects model specifications. The chapter first assumes that the strict exogeneity assumption holds for the regressors, and derives optimal (F)GLS estimators for all models accordingly, discussing the estimation processes in depth. This is utterly important, as with the proposed methods the performed spectral decompositions and variance components estimations, needed for feasibility reasons and to complete the estimation process, can be easily generalized to any random effects model specification in any dimension. As the data can now grow in not only two, but three dimensions at the same time, it is crucial to collect the exact properties under which the FGLS estimator is consistent. Some of the consistency properties also carry a *convergence property*, which means that the FGLS estimator of a model converges to that model's specific Within estimator. For some models, consistency even *implies* convergence. While this phenomenon by itself does not violate the feasibility of the estimators or their properties, the parameters of some fixed regressors – just like in case of fixed effects models – become unidentified, rendering the estimation of such parameters impossible. Apart from this identification problem, inconsistency in many of the several semi-asymptotic cases persists. To correct for this, so-called *mixed models* are proposed, combining both fixed and random components. One of the main reasons why random effects lag behind in popularity, is that the strict exogeneity assumption is hard to fulfill. The chapter also considers the case of endogenous regressors, and proposes Hausman-Taylor IV estimators to reach a full set of parameter estimates. The main results of the chapter are also extended to higher dimensions and to incomplete data, to argue for their wide applicability and easy generalizability. Finally, some basic insights on testing for random effects model specifications, for exogeneity, and for instrument validity are considered.

Chapter 3: The Estimation of Varying Coefficients Multi-dimensional Panel Data Models

The third chapter of the thesis considers several new varying coefficients models, and derives appropriate Least Squares estimators for them. The varying slope coefficients are assumed to be fixed, rather than random, and the slope parameters are assumed to comprise a universal part, common for all entities and time periods, as well as a varying component, which can be individual and/or time specific. In order to disentangle these two effects in these under-identified models, some parameter restrictions are to be assumed. As it turns out, the Least Squares estimation of the restricted model is simple theoretically, but cumbersome in practice due to the many complex functional forms and large matrices to work with. Further, as alternative parameter restrictions mean the full repetition of the calculation, alternative solutions are proposed. Luckily, the so-called *Least Squares of incomplete rank*, on the other hand, is easy to implement even in practice, and derives the part of the estimator which is model-specific before arriving at the restriction. In this way, the flexible exchange of various parameter restrictions is guaranteed. Some

10.14754/CEU.2017.05

insights on the identification issues, and on the interpretation of models with variables with index deficits are considered, as well as some preliminary results on varying coefficient autoregressive models. Mixed coefficients models, having both fixed and random coefficients, are also briefly visited, and some of their estimation issues considered.

Chapter 4: Empirical Applications for Multi-dimensional Panels

Section 4.3 is joint work with Janos Kollo and Istvan Boza, Sections 4.1–4.2 are solely my own.

The fourth chapter of the thesis merges two distinct empirical studies employed on three-way data: an international trade application, "*Regularities of Panel Estimators: A Trade Application*", and a study on wage returns, "*Contemporaneous and Lagged Wage Returns to Foreign-Firm Experience – Evidence from Linked Employer-Employee Data*". The former contributes to the literature by (i) comparing several fixed and random effects estimators, reflecting the typical estimation issues and some further regularities detailed in Chapters 1 and 2; (ii) by considering a new data set and taking into account data related issues, such as incompleteness, improving the results of several earlier papers which measured the effect of trade membership on real trade activity. The second study falls in line with several international studies capturing the (contemporaneous and lagged) wage returns of foreign experience on workers and on their colleagues. Foreign capital in emerging economies is subject to many criticisms, such as displacing local businesses, expatriating profits, or reducing tax liabilities. It is not clear, however, to what extent the domestic market gains from FDI. Apart from the fact that foreign wages are spent in the host country, and that domestic firms can imitate foreign-owned enterprises, workers of foreign-owned firms are usually more productive and are paid higher (contemporaneous effects). This wage premium can then be preserved when the worker re-enters the domestic market (lagged effect). Further, the presence of the accumulated knowledge of ex-foreign workers can also raise the productivity of their colleagues with no foreign experience (spillover effect). These advantages of FDI may in fact outweigh its losses. To elaborate on these ideas, several, mostly fixed effects models are formulated and regressed on a matched employer-employee data set covering half of the Hungarian working-age population.

10.14754/CEU.2017.05

Acknowledgements

First and foremost I owe my deepest gratitude to my supervisor and mentor, Laszlo Matyas for his constant, great help and support he has shown me over the past years. His continuous advising, inspiring criticism, and high standards have been guiding me through my PhD years, have greatly improved my professional as well as my personal skills, and have resulted in several successful works along with this thesis. I can not be grateful enough for his endless efforts, his strenuousness, and his directive: never to be satisfied with the second-best.

I thank my excellent examiners, Patrick Sevestre and Gabor Kezdi for their valuable comments and suggestions which have brought this thesis into much better shape and have improved its contribution to the literature greatly. I also thank all members of the Thesis Committee for their useful comments, questions, remarks.

Several academic visits and joint works have helped the promotion of this thesis. I am most grateful to Tom Wansbeek (University of Groningen), Mark Harris (Curtin University), Felix Chan (Curtin University), Jaya Krishnakumar (University of Geneva), Monika Avila Marquez (University of Geneva) and Janos Kollo (Eotvos Lorand University) for the fruitful discussions, useful critics and afternoons spent in bringing the papers and chapters into life.

I am indebted to my Professors and fellow students for their remarks on the Research Seminar presentations and on the early versions of this thesis, especially to Adam Szeidl, Sergey Lychagin, Botond Koszegi, Miklos Koren, Peter Zsohar, Miklos Farkas, Peter Farkas, Balazs Krusper, Istvan Szabo, Mirjam Szillery, and many others.

Comments received from the participants of the Panel Data Conferences in Paris, London and Budapest, as well as of the EEA-ESEM Annual Meeting in Geneva are acknowledged. I am thankful for the financial support from the CEU foundation and COEURE FP7.

Last, but not least, I am forever grateful for the constant support of my family: Laszlo Balazsi, Katalin Balazsine Farkas and Tamas Balazsi, Jozsef Farkas and

Jozsefne Farkas. I thank Virag Voros for all her love and support. Their encouragement has always given me strength to complete the next challenge and has never let me believe for one minute that I can fail.

10.14754/CEU.2017.05

Contents

List of tables *page* xv
List of illustrations xvii
Introduction 1

1 The Estimation of Multi-dimensional Fixed Effects Panel Data Models 10
1.1 Introduction 10
1.2 Models with Different Types of Heterogeneity 11
1.3 Least Squares Estimation of the Models 15
1.4 Incomplete Panels 20
1.5 The Within Estimator 22
1.5.1 The Equivalence between LSDV and the Within Estimator 22
1.5.2 Incomplete Panels with the Within Estimator 24
1.6 Dynamic Models 29
1.6.1 Nickell Biases 29
1.6.2 Arellano–Bond Estimation 31
1.7 Heteroscedasticity and Cross-correlation 33
1.7.1 The New Covariance Matrices and the GLS Estimator 34
1.7.2 Estimation of the Variance Components and the Cross Correlations 35
1.8 Extensions to Higher Dimensions 38
1.8.1 Different Forms of Heterogeneity 38
1.8.2 Least Squares and the Within Estimators 39
1.8.3 Some Data Problems 40
1.9 Conclusion 41
A Background Calculations – Obtaining M_D from D 42

B Background Calculations – Derivations of the No Self-flow Transformations 42
B.1 No Self-flow Derivation for the Pure Cross-sectional Panel Model 42
B.2 No Self-flow Transformation for Model (1.2) 44
B.3 No Self-flow Transformation for Model (1.6) 46
B.4 No Self-flow Transformation for Model (1.7) 47

2 Modelling Multi-dimensional Panel Data: A Random Effects Approach 50
2.1 Introduction 50
2.2 Different Model Specifications 51
2.2.1 Various Heterogeneity Formulations 52
2.2.2 Spectral Decomposition of the Covariance Matrices 54
2.3 FGLS Estimation 58
2.4 Incomplete Data 64
2.4.1 Structure of the Covariance Matrices 64
2.4.2 The Inverse of the Covariance Matrices 67
2.4.3 Estimation of the Variance Components 69
2.5 Endogenous Regressors 71
2.5.1 The Hausman-Taylor-like Instrumental Variable Estimator 72
2.5.2 Time Varying Individual Specific Effects 78
2.5.3 Properties 81
2.5.4 Incomplete Data 82
2.5.5 Using External Instruments 83
2.6 Various Tests for Random Effects Models 83
2.6.1 Testing for Model Specification 83
2.6.2 Testing for Exogeneity 86
2.6.3 Testing for Instrument Validity 87
2.7 Extensions 88
2.7.1 4D and beyond 89
2.7.2 Mixed Fixed-Random Effects Models 90
2.8 Conclusion 93
A Rationale Behind the Normalization Factors 95
A.1 Example for Normalizing with 1 95
A.2 Example for Normalizing with $\sqrt{N_1 N_2}/A$ 95
B Proof of Formula (2.17) 96
C Inverse of (2.49), and the Estimation of the Variance Components 97

10.14754/CEU.2017.05

3 The Estimation of Varying Coefficients Multi-dimensional Panel Data Models 101

3.1 Introduction 101

3.2 Fixed Coefficients Models and their Estimation 104

3.2.1 The Benchmark Model and its Estimation with Least Squares 104

3.2.2 The Least Squares of Incomplete Rank 106

3.2.3 Incomplete Data 108

3.2.4 Alternative Model Specifications 108

3.3 Extensions 114

3.3.1 Varying Coefficients as Functions of Observables 114

3.3.2 Index Deficiency in the Variables 116

3.3.3 Mixed Models 118

3.4 Some Thoughts on Dynamic Models 119

3.5 Conclusion 121

A Detailed Estimation Strategy 122

4 Empirical Applications for Multi-dimensional Panels 126

4.1 Introduction 126

4.2 Regularities of Panel Estimators: A Trade Application 127

4.2.1 Introduction and Previous Results 127

4.2.2 The Data and Model Specifications 128

4.2.3 Results 129

4.2.4 Discussion 137

4.3 Contemporaneous and Lagged Wage Returns to Foreign-Firm Experience – Evidence from Linked Employer-Employee Data 140

4.3.1 Introduction 140

4.3.2 FDI in Hungary 146

4.3.3 Data 147

4.3.4 Estimation Strategies 148

4.3.5 Results 155

4.3.6 Conclusion 164

A Supplementary Tables 166

B Data Appendix 168

10.14754/CEU.2017.05

Tables

1.1	Model specific D matrices	15
1.2	Different forms of M_D after simplification	16
1.3	Trace calculations for the bias of different models	31
1.4	Semi-asymptotic bias of each model formulation	31
2.1	Structure of the Ω^{-1} matrices	57
2.2	Asymptotic conditions when the models' FGLS converges to a Within estimator	58
2.3	Sample conditions for the consistency of the FGLS Estimator	62
2.4	Normalization factors for the finiteness of $\hat{\beta}_{OLS}$	63
2.5	Asymptotic results when the OLS should be used	63
2.6	Sources of endogeneity on the level of partitions of the regressors for model (2.31)	73
2.7	Translation of matrix operations into scalar	75
2.8	Sources of endogeneity on the level of partitions of the regressors for model (2.37)	78
2.9	Pairs of variables needed to be instrumented jointly	79
2.10	Proposed instruments H_p for each endogenous variable	80
2.11	Order conditions for Model (2.37)	82
2.12	Specific functional forms of the ANOVA F-test	85
2.13	Mixed effects model formulations	93
3.1	Monte Carlo simulation for assessing optimality against efficiency loss	112
3.2	Orders of the largest matrix to be inverted during estimation	114
3.3	Feasible model restrictions in response to various forms of index deficiencies	117
4.1	Fixed effects and random effects model specifications used in estimating (4.1)	130
4.2	Pooled OLS estimate of model (4.1)	131
4.3	The source of variation needed for identification, and the interpretation of the coefficient of main interest	132

10.14754/CEU.2017.05

4.4	Within estimates of various fixed effects model formulations	133
4.5	FGLS estimates of various random effects model formulations	135
4.6	Asymptotic conditions under which the FGLS estimator converges to a Within, and conditions needed for consistency	136
4.7	Least squares estimates of various random effects specifications	138
4.8	Foreign ownership in the estimation sample in 2003	147
4.9	Firms and workers by type of mobility in the estimation sample	150
4.10	Average wages of firms and workers by type of mobility	151
4.11	Estimates of the foreign-domestic wage gap by skill levels	156
4.12	Comparing incumbent workers in newly created incumbent firms	158
4.13	The wage advantage of ex-foreign workers, worker fixed effects	159
4.14	The wage advantage of ex-foreign workers, worker and firm fixed effects	160
4.15	The wages of workers arriving from/leaving for foreign and other domestic enterprises	161
4.16	Spillover effects estimated with worker and firm fixed effects, 2005-2009	163
A.1	Descriptive statistics for the estimation sample of (4.2)	166
A.2	Pooled OLS results for (4.2)	167

10.14754/CEU.2017.05

Illustrations

2.1 Possible Instrumental Variables 79

10.14754/CEU.2017.05

Introduction

In the last decade or so, we have experienced a data revolution of unbelievable size and scale. The rapid explosion of information technology – and its effects on computer performance and computational limits – opened the way to easily storing, collecting, and managing data sets with thousands of variables, and (possibly several) millions of observations. Users of both cross-sectional and time-series data have gained from the increased size through several channels, including the representativeness of the data, the higher precision of the estimates, or the use of a larger subset of the observations for testing for the validity of model assumptions. None of the improvements on the two data types, however, can be compared to the fundamental developments on panel data.

From traditional two-way panels, forming clusters on individuals as an augmentation of the individual index, or collecting data involving new indices, three- and higher-dimensional panels emerge. We see several good examples for such data, *e.g.*, linked employer-employee panels of nearly all advanced economies, world trade datasets (which can also embrace industry- or even product levels), the EU KLEMS industry level data, data on academic research performance, and many others.

Two dimensional (2D) models are not always suitable to describe phenomena based on multi-dimensional data. Although by defining pairs of individuals with a composite, single index, any two-dimensional model can be casted in the three-dimensional context, such models are unable to fully deplete the true richness in three-dimensional panels. This is so, as the underlying excessive heterogeneity of the data now takes on several complex forms impossible to be represented by models formed on 2D data. In order to successfully deal with such multi-dimensional heterogeneity, new multi-dimensional models, together with new, or adjusted estimation techniques should be constructed. Model building and estimation under three- or higher-way panel data are subject, however, to four key difficulties in general, which I refer to as the *four regularities of multi-dimensional panels*.

First, the number of possible model specifications increases dramatically. Let's consider fixed effects models for the moment. Unlike with two-way models, the decision is not whether to include time effects or individual effects (maybe both), but which fixed effect(s) to include from the many. As an illustration for the complexity of the problem, three-way data allows 63, while four-way data 16 383 fixed effects model specifications (as opposed to 3 in case of two-way panels). Although it comes as no surprise that the majority of model formulations is hardly useful economically, the remaining number of empirically relevant specifications is still high, further, it grows exponentially with the dimensions.

Second, the size of the data can make some estimators unfeasible for practical use. While the mere (stored) size of the data is rarely of any concern (hard disc spaces are usually well beyond raw data sizes), several calculations involve operations (*e.g.*, multiplication, inversion) with matrices of extreme orders. If this is the case, the derived estimators are of no practical use, and the efforts put into the derivations of the methods are wasted.

Third, the covariates are likely to suffer from index deficits. This phenomenon is also present for variables on 2D data, fixed over time or entities, like age, gender, and educational attainment for individuals. Index deficiency, however, is incomparably more significant for variables on three-way data, where it is not uncommon to exclusively have such variables, which show no variation in some of the three (or higher) dimensions. While this deficiency of the variables seems harmless, it can lead to possibly severe identification issues, or, in worst cases, may even invalidate the model specification.

Lastly, but not less importantly, multi-dimensional data is almost exclusively of an incomplete nature. Incompleteness can be the consequence of, for example, data unavailability, non-reporting, or individuals dropping out of the sample for various reasons, but can also be in the data by construction. One of the leading examples for such "unbalancedness by construction" corresponds to flow-type data, where self-flows are naturally unobserved, and so are left out of the data set. While incompleteness does not affect some estimators, techniques assuming complete data become biased and inconsistent in general, upon using on incomplete data.

When it comes to modelling on multi-dimensional panels, it is crucial to constantly keep track of the above four regularities, in order not to reduce the value of the results. Each chapter of this thesis recognizes these problems: the regularities together with suggested solutions are discussed thoroughly.

In contrast with data constructed by the researcher, panel data for economics use (and in general, data in social sciences) are usually less transparent, the data generating processes (DGPs) are harder to identify or detect. Nerlove et al. (2008) argue that in such cases forming estimators and constructing parameter tests are only one part of the job: identifying and learning about the DGP is not less important.

10.14754/CEU.2017.05

Essential knowledge on how the data was generated should be part of the model specification. Do I observe all trade flows between countries, or just the ones being non-zero? Can it be that small trade flows are uniformly non-reported or considered zero? Do I have a pool of firms coming from random sampling, representative enough for the universe of Hungarian companies, or are they selective in one way or another? What about the individuals in the sample? Can I consider them as random draws from a large population, or should they be addressed as fixed entities, like with states or countries? In case of an employer-employee matched panel, it is usually reasonable to assume that workers or firms are drawn randomly: exchanging two will not alter the distribution of the observables, unlike with time, where periods can not be switched without consequences. In any such scenarios, individual effects should be considered random, while time effects should be thought of as fixed. In typical applications (perhaps heavy) *a priori* assumptions have to be made on the DGP, whose validity in turn can be assessed with testing.

The most direct way to capture the relationship between left hand side and right hand side variables is done with *linear* econometric models, which this thesis exclusively focuses on. Fixed effects models are dealt with in Chapter 1, where the unobserved heterogeneity is represented by different intercept parameters for different entities and/or time periods, while Chapter 2 discusses the case when individual and/or time variation is random. Due to their popularity, and the tremendous work dedicated to them, numerous extensions of the traditional fixed and random effects models exist, like varying slope coefficients models (Chapter 3), simultaneous equation models, models with random regressors, just to mention a few.

One of the virtues, which is also a curse of panel data, is that the entities, pairs of entities are followed over time. While this enables to control for individual characteristics, and by that to compare individuals with similar demographics, observations in economics panel data almost surely have some path dependence. Individual histories matter in present decisions, and as such, no perfectly exogenous regressors exist. Regardless of that the econometric model is dynamic (Chapter 1), that is, has past values of the dependent variable on the right hand side, or some regressors are endogenous (Chapter 2), and correlated with the disturbance, fixed effects and random effects estimators are generally biased and inconsistent. Issues with endogeneity therefore must be taken into account rigorously when dealing with panel data. While we will see how different transformations on the data remove part or all of that endogeneity in some lucky cases, how asymptotics can wipe the bias out, or even how transformed data can serve as its own instrument, the need for clever Instrumental Variable (IV) and Generalized Method of Moments (GMM) techniques is constant.

Although the thesis concentrates on linear models, non-linear models also have a determinant role in dealing with latent variables or variables describing probabil-

10.14754/CEU.2017.05

ities of the occurrence of some event. Incorporating individual and/or time heterogeneity to non-linear models in this multi-dimensional context is however much less trivial, than it is for linear models. Furthermore, failure to account for the proper form of heterogeneity results in severe biases, not only inefficiencies, which might have been the case for linear models (Nerlove et al., 2008). Non-linear models, however, would take me away from the goal of this thesis and would open so many new and interesting questions, that a separate thesis could be devoted to their discussion.

As computers are more and more heavily involved in following and registering everyday transactions, data sets can cover entire populations and can grow almost without bounds, giving rise to the concept of *Big Data*. These data sets may not only consist of billions of individual transactions, but might as well comprise several thousands of variables: De Mol et al. (2017) bring the example of how linking administrative data can tremendously increase their number. Varian (2014) argues that as much as various econometric- or machine learning techniques associated with information extraction from Big Data work with more or less success, Big Data challenges researchers in at least two distinct ways. One, the sheer size of the data demands high-end computation techniques and resources, and two, the availability of the excessively many predictors requires some variable selection tool in order to enhance the estimates. As most Big Data analytic tools originate from machine learning techniques, Varian (2014) and several other economists undoubtedly think that panel data methods have a lot to offer for the better understanding of Big Data and for the better predictions formed on that. Bringing closer computer-originated learning techniques to traditional econometric tools is the joint interest of all Big Data analysts then. Clearly these ideas, and the field of Big Data itself is a lot bigger than what could be covered, or at least meaningfully addressed by this thesis. Instead, most estimators are inspected from the side of computational feasibility, and wherever such burdens are expected to persist, alternative, much less computationally heavy techniques are proposed. This, in some way, can be thought of as efforts dedicated to dealing with Big Data.

Chapter 1, "*The Estimation of Multi-dimensional Fixed Effects Panel Data Models*", formulates the excess heterogeneity in the data with fixed, observable parameters. In such cases, the heterogeneous parameters are in fact splits of the regression constant. Several such three-dimensional fixed effects models are collected from the literature, all of which correspond to empirically relevant cases. The models are estimated with Least Squares Dummy Variable (LSDV) estimator, and in order to circumvent the joint estimation of possibly (hundreds of) thousands of parameters, the estimators are also expressed separately for each model parameter. It is also shown that the so-called *Within estimator*, which first wipes out the fixed effect parameters with a linear transformation, then performs a Least Squares on the transformed model, is numerically equivalent to the LSDV. The Within estimator reaches estimates at no costs, as long as the data at hand is complete. Typically, however, the data contains "holes", which can either correspond to observations missing "randomly", or are there by construction. It is shown that conveniently, the Within estimator completely alleviates the dimensionality issue (the high cost of the estimation), for structured incompleteness (like the no self-flow phenomenon, which, to our knowledge, had never been explored), and partially, when it comes to handling incompleteness in general. As the incompleteness-robust Within estimator can still be cumbersome to perform for some models, an iterative way converging to the Within estimator is also suggested. While this iteration usually takes a tremendous amount of time, it almost fully eliminates computational burdens.

This chapter also contributes to the literature by considering dynamic autoregressive fixed effects specifications, first, by showing how the presence of various lags of the dependent variable violates the consistency of the Within estimator, generalizing the so-called *Nickell-bias*, then, by offering Arellano-Bond-type instrumental variable estimators to correct for the arising inconsistency. Somewhat surprisingly, however, not all three-way model specifications carry this asymptotic bias. Eventual heteroscedasticity and the cross-correlation of the disturbance terms are also accounted for by proposing appropriate Feasible Generalized Least Squares (FGLS) estimators. The chapter ends with a generalization to four- and higher-dimensional fixed effect models, and intuitively argues that the results of the study, especially Within estimators of any nature, can easily be generalized to any dimensions and for any fixed effects model specifications.

The contribution of the chapter is (i) collecting 3D fixed effects model formulations and deriving estimators universally; (ii) deriving incompleteness-robust estimators in case of general incompleteness (extensions of 2D results), and in case of no self-flow data (complete novelty); (iii) extending the Nickell-bias and proposing proper IV/GMM estimators; (iv) taking into account cross-section dependence in the presence of fixed effects; and, (v) arguing for the wide applicability of the results by discussing four-way extensions. My respective contribution involves the derivation and explanation of estimators and transformations in (i) and (ii), estimators and bias-formulas in (iii), while points (iv) and (v) are my own work.

The formulation of fixed, observable parameters, however, is not the only way to incorporate heterogeneity into multi-dimensional panel models. By assuming that the unobserved heterogeneity is random, that is, can be described with a set of random variables, we arrive at random effects models.

Chapter 2, "*Modelling Multi-dimensional Panel Data: A Random Effects Approach*", proposes several appealing random effects model specifications. Interestingly, only a subset of these models has been used in the literature, most probably due to the unavailability of estimators and their unknown properties. The chapter first assumes that the strict exogeneity assumption holds for the regressors, and derives optimal (F)GLS estimators for all models accordingly, discussing the estimation processes in depth. This is utterly important, as with the proposed methods the performed spectral decompositions and variance components estimations, needed for feasibility reasons and to complete the estimation process, can be easily generalized to any random effects model specification and to any dimension.

As the data can now grow in not only two, but three dimensions at the same time, it is crucial to collect the exact properties under which the FGLS estimator is consistent. Some of the consistency properties also carry a 'convergence property', which means that the FGLS estimator of a model converges to that model's specific Within estimator. For some models, consistency even *implies* convergence. While this phenomenon by itself does not violate the feasibility of the estimators or their properties, the parameters of some fixed regressors – just like in case of fixed effects models – become unidentified, rendering the estimation of such parameters impossible. Apart from this identification problem, inconsistency in many of the several semi-asymptotic cases persists. To correct for this, so-called *mixed models* are proposed, combining both fixed and random components.

One of the main reasons why random effects lag behind in popularity, is that the strict exogeneity assumption is hard to fulfill. The chapter also considers the case of endogenous regressors, and proposes Hausman-Taylor instrumental variable estimators to reach a full set of parameter estimates. As endogeneity can come from many different sources, further, variables with various index deficiencies are affected differently, covering the relevant cases and formulating the proper IV estimators with the order conditions are real challenges here. The main results of the chapter are also extended to higher dimensions and to incomplete data, to argue for their wide applicability and easy generalizability. Finally, some basic insights on testing for random effects model specifications, for exogeneity, and for instrument validity are considered. These tests are essential to collect some evidence on which model to choose from the many (by using an extended version of Fisher's ANOVA test), and on where to go with our random effects model (are the regressors exogenous or not?). Hausman tests are developed first to test for the existence of endogeneity among the regressors, and for the identification of the sources of endogeneity, then to test for the validity of the collected instruments if endogeneity is the case.

The contribution of the chapter is (i) collecting new random effects formulations and extending (F)GLS estimators giving a technical know-how; (ii) extending 2D results on incomplete-robust estimators, mixed models, and further to four-way panels; (iii) extending the Hausman-Taylor instrumental variable estimator to tackle the case of endogenous regressors; and, (iv) constructing various tests for model selection, exogeneity and instrument validity to verify various hypotheses formed on the models. My respective contribution is the derivation of decompositions, estimators, properties in (i), (ii) and (iii), the construction of tests in (iv). Results concerning mixed models, four-dimensional extensions and tests for model selection are my own work.

Chapters 1 and 2 considered models of *average* effects. An individual fixed effect, for example, encompasses all effects specific to the given individual. Its parameter estimate is then interpreted, as the *average of all observed and unobserved effects, specific to that individual*. One of the most important statistical features of multi-dimensional panel data sets is, however, that heterogeneity is likely to take more complicated forms, which begs for more complex heterogeneity formulations. One appealing way is to incorporate heterogeneity *marginally*, that is, through individual (and/or time) varying slope parameters.

Chapter 3, "*The Estimation of Varying Coefficients Multi-dimensional Panel Data Models*" considers several new varying coefficients models, and derives appropriate Least Squares estimators for them. The varying slope coefficients are assumed to be fixed, rather than random, as it would be the case for random coefficients models, further, the slope parameters are also assumed to comprise a universal part, common for all entities and time periods, as well as a varying component, which can be individual and/or time specific. In order to disentangle these two effects in these under-identified models (economically speaking, to identify the model parameters), some parameter restrictions are to be assumed. As it turns out, Least Squares estimation of the restricted model is simple theoretically, but cumbersome in practice due to the many complex functional forms and large matrices to work with. Further, as alternative parameter restrictions mean the full repetition of the calculation, alternative solutions are proposed. Luckily, the so-called *Least Squares of incomplete rank*, on the other hand, is easy to implement even in practice, and derives the part of the estimator which is model-specific before arriving at the restriction. In this way, the flexible exchange of various parameter restrictions is guaranteed. Some insights on the identification issues with, and on the interpretation of models with variables with index deficits are considered, as well as some preliminary results on varying coefficient autoregressive models. Mixed coefficients models, having both fixed and random coefficients, are briefly visited, and some of their estimation issues are considered, while the idea of expressing the varying coefficients as functions of observables, and by that highly reducing the number of parameters to estimate, is also noted.

The contribution of this chapter is (i) proposing new fixed variable coefficients models; (ii) applying the concept of Least Squares of incomplete rank to these models, which, to the best of my knowledge, has never been done; (iii) visiting

10.14754/CEU.2017.05

various extensions, like incomplete data, variables with index deficiency, dynamic autoregressive models and mixed models. The concept of mixed models within the context of varying coefficient models is existent, but to my knowledge no real efforts had been dedicated towards its proper estimation.

Together with the theoretical results, it is important to show how the main models and estimators of the thesis fare empirically. After all, the ultimate goal of any theoretical work, apart from motivating other theoretical works, is to form the foundations of empirical efforts. Chapter 4, "*Empirical Applications for Multidimensional Panels*", merges two distinct empirical studies employed on three-way data: an international trade application, "*Regularities of Panel Estimators: A Trade Application*", and a study on wage returns, "*Contemporaneous and Lagged Wage Returns to Foreign-Firm Experience – Evidence from Linked Employer-Employee Data*". The former contributes to the literature by (i) comparing several fixed and random effects estimators, reflecting the typical estimation issues and some further regularities detailed in Chapters 1 and 2; (ii) by considering a new data set and taking into account data related issues, such as incompleteness, improving the results of several earlier papers which measured the effect of trade membership on real trade activity. This part of Chapter 4 is my own work.

The second study falls in line with several international studies capturing the (contemporaneous and lagged) wage returns of foreign experience on workers and on their colleagues. Foreign capital in emerging economies is subject to many criticisms, such as displacing local businesses, expatriating profits, or reducing tax liabilities. It is not clear, however, to what extent the domestic market gains from FDI. Apart from the fact that foreign wages are spent in the host country, and that domestic firms can imitate foreign-owned enterprises, workers of foreign-owned firms are usually more productive and are paid higher (contemporaneous effects). This wage premium can then be preserved when the worker re-enters the domestic market (lagged effect). Further, the presence of the accumulated knowledge of ex-foreign workers can also raise the productivity of their colleagues with no foreign experience (spillover effect). These advantages of FDI may in fact outweigh its losses. To elaborate on these ideas, several, mostly fixed effects models are formulated and regressed on a matched employer-employee data set covering half of the Hungarian working-age population. The contribution of this second part of the thesis is (i) taking international efforts devoted to uncovering these three effects into account and applying them to the case of the Hungarian economy; (ii) turning to "true" 3D fixed effects models to get a better grip on issues like selectivity. My respective contributions in this section are the discussion of estimation issues with multiple fixed effects and the identification and calculation of the contemporaneous wage gap.

10.14754/CEU.2017.05

Ongoing work on random coefficient models (Krishnakumar et al., 2017), and on models with more complex functions of the dependent variable on the right hand side are excluded from this thesis due to size limits.

10.14754/CEU.2017.05

1

The Estimation of Multi-dimensional Fixed Effects Panel Data Models

Sections 1.2–1.6 are joint works with Laszlo Matyas and Tom Wansbeek and are forthcoming in Econometric Reviews. Sections 1.7 and 1.8 are solely my own.

1.1 Introduction

Multi-dimensional panel data sets are becoming more readily available. They are used to study phenomena like international trade, capital flow between countries or regions, the trading volume across several products and stores over time, employee-employer matches over time (three panel dimensions), the air passenger numbers between multiple hubs by different airlines, research performance (four panel dimensions) and so on. Models on multi-dimensional panels have the exceptional advantage (over two-dimensional (2D) ones) of incorporating excessive heterogeneity in several newly attainable forms.

Model formulations in which the individual and/or time heterogeneity factors are considered observable parameters, rather than random variables are called fixed effects models. In the basic, most frequently used models, these heterogenous parameters are in fact splits of the regression constant. They can take different values in different sub-spaces of the original data space, while the slope parameters remain the same. In this chapter we propose estimation mechanisms to deal with three-dimensional (3D) fixed effects models, and generalize the results using numerous extensions to widen the applicability of the study.

In Section 1.2, we line up various fixed effects model specifications proposed in the literature for three-dimensional data. For each of these models, we pay special attention to the structure of the intercept parameters. In Section 1.3, we show the Least Squares estimation procedure along with some of its finite and asymptotic properties, also taking an insightful glimpse into parameter testing. The data at hand is often incomplete, either by construction (like the lack of within-country ob-

10.14754/CEU.2017.05

servations in case of flow-type data, hereafter *no self-flow*), or simply due to existing 'gaps' in the data (*general incompleteness*). Section 1.4 describes how to adjust Least Squares to handle incomplete data, together with the caveats of this estimator, and also presents an intuitive way to fix the arising dimensionality issue. Section 1.5 introduces the so-called *Within Estimator*, and shows its numerical equivalence to Least Squares. As the Within Estimator employs linear transformations on the data, some of the dimensionality problem is alleviated, even in case of incomplete data. Up to this point, the models considered were static. In Section 1.6, we show how the presence of the lagged dependent variable may render Least Squares on the transformed data inconsistent, thus generalizing the well-known Nickell (1981) bias. Somewhat surprisingly, however, with three-way data, inconsistency does not occur in all models. For the cases with inconsistency, we present the appropriate generalization of the Arellano-Bond estimator. We also account in Section 1.7 for eventual heteroscedasticity and cross-correlation in the disturbance terms, while Section 1.8 extends the chapter's results to four- and higher dimensions, to argue for their easy and wide generalizability. Finally, section 1.9 concludes.

1.2 Models with Different Types of Heterogeneity

In three-dimensional panel data, the dependent variable of a model is observed along three indices, such as y_{ijt}, $i = 1,\ldots,N_1$, $j = 1,\ldots,N_2$, and $t = 1,\ldots,T$, and the observations have the same ordering: index i goes the slowest, then j, and finally t the fastest,[1] such as

$$(y_{111},\ldots,y_{11T},\ldots,y_{1N_21},\ldots,y_{1N_2T},\ldots,y_{N_111},\ldots,y_{N_11T},\ldots,y_{N_1N_21},\ldots,y_{N_1N_2T})'.$$

We assume in general that the index sets, $i \in \{1,\ldots,N_1\}$ and $j \in \{1,\ldots,N_2\}$ are (completely or partially) different. When dealing with economic flows, such as trade, capital, investment (FDI), *etc.*, there is some kind of reciprocity, in such cases it is assumed, that $N_1 = N_2 = N$.

The main question is how to formalize the individual and time heterogeneity — in our case, the fixed effects. In standard 2D panels, there are only two effects, individual and time, so in principle 2^2 model specifications are possible (if we also count the model with no fixed effects). The situation is fundamentally different in three-dimensions. Strikingly, the 6 unique fixed effects formulations enable a great variety, precisely 2^6, of possible model specifications. Of course, only a subset of these are used, or make sense empirically, so in this chapter we are only considering the empirically most meaningful ones.

Throughout the chapter, we follow standard ANOVA notation, that is I and J

[1] Please note, that the N_1, N_2 notation does not mean, by itself, that the data is unbalanced.

denote the identity matrix, and the square matrix of ones respectively, with the size indicated in the index, $\bar{J}$ denotes the normalized J (each element is divided by the number in the index), and ι denotes the column vector of ones, with size in the index. Furthermore, an average over an index for a variable is indicated by a bar on the variable and a dot on the place of that index. When discussing unbalanced data, a plus sign at the place of an index indicates summation over that index. The matrix M with a subscript denotes projection orthogonal to the space spanned by the subscript.

The models can be casted in the general form

$$y = X\beta + D\pi + \varepsilon \tag{1.1}$$

with y and X being the vector and matrix of the dependent and explanatory variables (covariates) respectively of size $(N_1N_2T \times 1)$ and $(N_1N_2T \times K)$, β being the vector of the slope parameters of size $(K \times 1)$, π the composite fixed effects parameters, D the matrix of dummy variables, and finally, ε the vector of the disturbance terms.

The first attempt to properly extend the standard fixed effects panel data model to a multi-dimensional setup was proposed by Matyas (1997) (see also Baltagi, 2005 and Balestra and Krishnakumar, 2008). The specification of this model is

$$y_{ijt} = \beta' x_{ijt} + \alpha_i + \gamma_j + \lambda_t + \varepsilon_{ijt} \tag{1.2}$$

where the α_i, γ_j, and λ_t parameters are the individual and time-specific fixed effects (picking up the notation of (1.1), $\pi = (\alpha' \ \gamma' \ \lambda')'$), and ε_{ijt} are the $i.i.d.(0, \sigma_\varepsilon^2)$ idiosyncratic disturbance terms. We also assume that the x_{ijt} covariates and the disturbance terms are uncorrelated.

Matyas (1997) and Matyas et al. (1997) applies model (1.2) to predict foreign trade flows: with local country i, target country j and year t, y_{ijt} denotes real export, while x'_{ijt} are various measures to affect the intensity of trade, like GDP, distance, bilateral dummies. In the present context α_i and γ_j are local and target country effects, while λ_t is the time (business-cycle) effect. The local country parameter shows the efficiency of country i in exporting, relative to other countries and to characteristics x'_{ijt}, while the target country parameter γ_j is interpreted as trade openness relative to other target countries and to characteristics x'_{ijt}. Then, the focus parameter for local GDP, for example, captures the increase in y_{ijt} in response to a unit increase in local GDP, controlled for the average GDP of the local country over time, the average target GDP over time, and the average local and target GDPs over countries. The effect β_k of a general x'_{ijtk} regressor is identified from (i) variation of x'_{ijtk} within group i, (ii) variation of x'_{ijtk} within group j and (iii) variation of x'_{ijtk} within group t. For example for GDP_{it}, the coefficient β is identified if $\mathrm{Var}(GDP_{it}) \neq 0$ within group i, for some $i = 1, \ldots, N_1$, and at the same time, $\mathrm{Var}(GDP_{it}) \neq 0$ within group t, for some $t = 1, \ldots, T$.

A model has been proposed by Egger and Pfaffermayr (2003), popular in the trade literature, which takes into account bilateral interaction effects. The model specification is

$$y_{ijt} = \beta' x_{ijt} + \gamma_{ij} + \varepsilon_{ijt}\,, \tag{1.3}$$

where the γ_{ij} are the bilateral specific fixed effect.

A variant of model (1.3), proposed by Cheng and Wall (2005), used in empirical studies is

$$y_{ijt} = \beta' x_{ijt} + \gamma_{ij} + \lambda_t + \varepsilon_{ijt}\,. \tag{1.4}$$

It is worth noticing that models (1.3) and (1.4) are in fact straight 2D panel data models, where the individuals are now the (ij) pairs.

As noted by Cheng and Wall, a country would still export different quantities to two target countries with the same GDP or distance simply due to having different cultural, political, ethnic relations affecting the level of trade. Bilateral fixed effects are then introduced to control for these (possibly) unobserved factors. Clearly to identify the focus parameters we need x'_{ijt} to have non-zero variation over t for at least one ij-pair for model (1.3), and we need (i) x'_{ijt} to have non-zero variation over t for at least one ij-pair and (ii) x'_{ijt} to have non-zero variation over i or j for at least one t year, for model (1.4). The coefficient for GDP, for example, is interpreted as "increasing *GDP* by one unit, export is increased by β_k units, controlling for other factors in x'_{ijt} and for unobserved country-pair factors", for model (1.3), and is interpreted as "increasing *GDP* by one unit, export changes by β_k units, controlling for other factors in x'_{ijt}, and for unobserved country-pair and business-cycle characteristics", for model (1.4).

Baltagi et al. (2003), Baldwin and Taglioni (2006) and Baier and Bergstrand (2007) suggest several other forms of fixed effects. A simpler model is

$$y_{ijt} = \beta' x_{ijt} + \alpha_{jt} + \varepsilon_{ijt}\,, \tag{1.5}$$

where we allow the individual effect to vary over time. It is reasonable to present the symmetric version of this model (with α_{it} fixed effects); however, as it has exactly the same properties, we consider the two models together. Grogger and Hanson (2011) use a somewhat similar gravity setup in analysing the selection and sorting of international migrants between many host and target countries. While they consider the left hand side variable to be the utility of worker i moving from country j to t, and the main explanatory variables are the wages to be paid and costs of migrating, we can easily modify the framework to focus on a single target country with worker i migrating from country j at year t. Although the utility of moving (wage minus costs) for two workers with similar individual and source country characteristics is most likely very similar, we can not rule out differences which can be attributed to unobserved but existing *source country–year* factors, such as the source country's political or cultural relationship with the host country, barriers of leaving the country of origin, etc. Clearly these unobserved factors might vary over years, especially if migration patterns are followed for a long enough period. For the identification of β we need x'_{ijt} to show variation over i for at least one jt-pair. The focus parameter β_k is then interpreted as the response to a unit jump in x'_{kijt}, controlling for observable and home country-year unobservable characteristics.

A variation of model (1.5) is

$$y_{ijt} = \beta' x_{ijt} + \alpha_{it} + \alpha^*_{jt} + \varepsilon_{ijt}\,, \tag{1.6}$$

where, using the same application, a *worker–time* fixed effect is also added to control for any (unobserved) time-varying worker characteristic, such as personal costs of leaving the home country, some measurement of current family status, attitude and motivation towards upcoming employment. Now, along with variation over i for at least one jt-pair, x'_{ijt} also has to show variation over j for some it-pairs as well in order to identify β, and further, the parameter for the cost of leaving, for example, is interpreted as the "change in utility in response to a unit change in cost of leaving, when other observables, as well as unobservable worker-time and home country-time effects are controlled for".

Lastly, the model that encompasses all the above effects is

$$y_{ijt} = \beta' x_{ijt} + \gamma_{ij} + \alpha_{it} + \alpha^*_{jt} + \varepsilon_{ijt}\,, \tag{1.7}$$

where y_{ijt} could stand for Hungary's Foreign Direct Investment (FDI) to sector i from country j at year t, as explained by x'_{ijt}, like distance, factor endowments, trade barriers, etc. In the present context, β is not only hard to interpret, but is difficult to identify as well. To get it identified we need x'_{ijt} to show variation over i for all jt-pairs, variation over j-for all it-pairs, and finally, variation over t for all ij-pair. In other words, parameters associated with regressors showing non-zero variation in *all* three dimensions are identified only under specification (1.7). A β_k is then interpreted as "the change in y_{ijt} in response to a unit change in x'_{ijtk}, controlling for other characteristics in x'_{ijt} as well as all fixed unobserved characteristics".

Each model with its specific D matrix from formulation (1.1) is summarized in Table 1.1.

Table 1.1 *Model specific D matrices*

Model	D
(1.2)	$((I_{N_1}\otimes \iota_{N_2T}),\ (\iota_{N_1}\otimes I_{N_2}\otimes \iota_T),\ (\iota_{N_1N_2}\otimes I_T))$
(1.3)	$(I_{N_1N_2}\otimes \iota_T)$
(1.4)	$((I_{N_1N_2}\otimes \iota_T),\ (\iota_{N_1N_2}\otimes I_T))$
(1.5)	$(I_{N_1}\otimes \iota_{N_2}\otimes I_T)$
(1.6)	$((I_{N_1}\otimes \iota_{N_2}\otimes I_T),\ (\iota_{N_1}\otimes I_{N_2T}))$
(1.7)	$((I_{N_1N_2}\otimes \iota_T),\ (I_{N_1}\otimes \iota_{N_2}\otimes I_T),\ (\iota_{N_1}\otimes I_{N_2T}))$

1.3 Least Squares Estimation of the Models

Let us assume, along with their independence from the disturbance terms, that the vector of regressors x_{ijt} is non-stochastic, and further, that none of the x_{ijt} variables is perfectly collinear with the fixed effects. In this case, if the matrix $(X,\ D)$ has full column rank, the Ordinary Least Squares (OLS) estimation of model (1.1), also called the Least Squares Dummy Variables (LSDV) estimator

$$\begin{pmatrix} \hat{\beta} \\ \hat{\pi} \end{pmatrix} = \begin{pmatrix} X'X & X'D \\ D'X & D'D \end{pmatrix}^{-1} \begin{pmatrix} X'y \\ D'y \end{pmatrix},$$

is the Best Linear Unbiased Estimator (BLUE). This joint estimator, however, in some cases is cumbersome to implement, for example for model (1.3), as one has to invert a matrix of order $(K+N_1N_2)$, which can be quite difficult for large N_1 and/or N_2. Nevertheless, following the Frisch-Waugh-Lovell theorem, or alternatively, applying partial inverse methods, the estimators can be expressed as

$$\begin{aligned} \hat{\beta} &= (X'M_DX)^{-1}X'M_Dy \\ \hat{\pi} &= (D'D)^{-1}D'(y-X\hat{\beta}), \end{aligned} \tag{1.8}$$

where the idempotent and symmetric matrix $M_D = I - D(D'D)^{-1}D'$ is the so called *within projector*. In the usual panel data context, we call $\hat{\beta}$ in (1.8) the optimal Within estimator (due to its BLUE properties mentioned above). The LSDV estimator for each specific model is then obtained by substituting in the concrete form of D and M_D, specific to that given model. Table 1.2 captures these different projection matrices for all models discussed. Appendix A gives some insights on how to obtain M_D from D. Also, it is important to define the actual degrees of freedom to work with, so the third column of the table captures this. By using M_D, instead of possibly large matrices, we only have to invert a matrix of size $(K\times K)$ to get $\hat{\beta}$.

Estimation of the fixed effects parameters are captured by the second part of

10.14754/CEU.2017.05

Table 1.2 *Different forms of M_D after simplification*

Model	M_D	Degrees of Freedom
(1.2)	$I-(I_{N_1}\otimes\bar{J}_{N_2T})-(\bar{J}_{N_1}\otimes I_{N_2}\otimes\bar{J}_T)-(\bar{J}_{N_1N_2}\otimes I_T)+2\bar{J}_{N_1N_2T}$	$N_1N_2T-N_1-N_2-T+1-K$
(1.3)	$I-(I_{N_1N_2}\otimes\bar{J}_T)$	$N_1N_2(T-1)-K$
(1.4)	$I-(I_{N_1N_2}\otimes\bar{J}_T)-(\bar{J}_{N_1N_2}\otimes I_T)+\bar{J}_{N_1N_2T}$	$(N_1N_2-1)(T-1)-K$
(1.5)	$I-(I_{N_1}\otimes\bar{J}_{N_2}\otimes I_T)$	$N_1(N_2-1)T-K$
(1.6)	$I-(I_{N_1}\otimes\bar{J}_{N_2}\otimes I_T)-(\bar{J}_{N_1}\otimes I_{N_2T})+(\bar{J}_{N_1N_2}\otimes I_T)$	$(N_1-1)(N_2-1)T-K$
(1.7)	$I-(I_{N_1}\otimes\bar{J}_{N_2}\otimes I_T)-(\bar{J}_{N_1}\otimes I_{N_2T})-(I_{N_1N_2}\otimes\bar{J}_T)$ $+(\bar{J}_{N_1N_2}\otimes I_T)+(\bar{J}_{N_1}\otimes I_{N_2}\otimes\bar{J}_T)+(I_{N_1}\otimes\bar{J}_{N_2T})-\bar{J}_{N_1N_2T}$	$(N_1-1)(N_2-1)(T-1)-K$

(1.8). So far we have assumed that D has full column rank. Unfortunately, this is only true for models of one fixed effect, that is, for (1.3) and (1.5). These model-specific estimators read as

$$\hat{\gamma} = \frac{1}{T}(I_{N_1N_2} \otimes \iota_T')(y - X\hat{\beta})$$

for model (1.3), and

$$\hat{\alpha} = \frac{1}{N_2}(I_{N_1} \otimes \iota_{N_2}' \otimes I_T)(y - X\hat{\beta})$$

for model (1.5). For the other models, the fixed effects are not identified, since the D matrix of such models has no full column rank. This is easy to see, as for example for model (1.2) the column-wise sums of $(I_{N_1} \otimes \iota_{N_2T})$, $(\iota_{N_1} \otimes I_{N_2} \otimes \iota_T)$ and $(\iota_{N_1N_2} \otimes I_T)$, the dummy matrices associated with the fixed effects parameters, all give the column of ones, $\iota_{N_1N_2T}$. To make them identified, we have to use extra information in form of some restrictions over the fixed effects parameters. The two most widely used ones are either to normalize the fixed effects, or to leave the parameters belonging to the last (or first) individual or time period out. We will follow this latter approach. Let us illustrate the idea on model (1.2). For model (1.2), D has a rank deficiency of 2, but for the sake of symmetry, we leave out all three last fixed effects parameters, α_{N_1}, γ_{N_2}, and λ_T from the model, and add back a general constant term c. That is, for a given (ijt) observation $(i, j, t \neq N_1, N_2, T)$, the intercept is $c + \alpha_i + \gamma_j + \lambda_t$, but for example for $i = N_1$, it is only $c + \gamma_j + \lambda_t$. Let us denote this modified D dummy matrix by D^*, to stress that now it contains the restriction. As D^* has full column rank, estimator (1.8) works flawlessly with D^*:

$$\hat{\pi}^* = (D^{*'}D^*)^{-1}D^{*'}(y - X\hat{\beta}),$$

where now $\pi^* = (c', \ \alpha' \ \gamma' \ \lambda')'$. We may have a better understanding on these estimators, if we express them separately for each fixed effects parameters. This step, however, requires the introduction of complex matrix forms, and nontrivial manipulations, but as it turns out, using scalar notation, they can easily be represented. For model (1.2), this is

$$\begin{aligned}
\hat{c} &= (\bar{y}_{N_1..} + \bar{y}_{.N_2.} + \bar{y}_{..T} - 2\bar{y}_{...}) - (\bar{x}'_{N_1..} + \bar{x}'_{.N_2.} + \bar{x}'_{..T} - 2\bar{x}'_{...})\hat{\beta} \\
\hat{\alpha}_i &= (\bar{y}_{i..} - \bar{y}_{N_1..}) - (\bar{x}'_{i..} - \bar{x}'_{N_1..})\hat{\beta} \\
\hat{\gamma}_j &= (\bar{y}_{.j.} - \bar{y}_{.N_2.}) - (\bar{x}'_{.j.} - \bar{x}'_{.N_2.})\hat{\beta} \\
\hat{\lambda}_t &= (\bar{y}_{..t} - \bar{y}_{..T}) - (\bar{x}'_{..t} - \bar{x}'_{..T})\hat{\beta}\,.
\end{aligned}$$

Notice, that as we excluded α_{N_1} from the model, its estimator is indeed $\hat{\alpha}_{N_1} = (\bar{y}_{N_1..} - \bar{y}_{N_1..}) - (\bar{x}'_{N_1..} - \bar{x}'_{N_1..})\hat{\beta} = 0$, similarly for $\hat{\gamma}_{N_2}$, and $\hat{\lambda}_T$. For model (1.4), where we have left out $\gamma_{N_1N_2}$ and λ_T, and, as before, added back some c general constant,

$$\begin{aligned}
\hat{c} &= (\bar{y}_{N_1N_2.} + \bar{y}_{..T} - \bar{y}_{...}) - (\bar{x}'_{N_1N_2.} + \bar{x}'_{..T} - \bar{x}'_{...})\hat{\beta} \\
\hat{\gamma}_{ij} &= (\bar{y}_{ij.} - \bar{y}_{N_1N_2.}) - (\bar{x}'_{ij.} - \bar{x}'_{N_1N_2.})\hat{\beta} \\
\hat{\lambda}_t &= (\bar{y}_{..t} - \bar{y}_{..T}) - (\bar{x}'_{..t} - \bar{x}'_{..T})\hat{\beta} \,.
\end{aligned}$$

For model (1.6), and (1.7), the rank deficiency, however, is not 2 but T, and $(N_1 + N_2 + T - 1)$, respectively. This means, that the restriction above can not be used. Instead, let us leave out the α_{it} parameters for $i = N_1$, that is, that last T from model (1.6). In this way, the estimators for the intercept parameters are

$$\begin{aligned}
\hat{\alpha}_{it} &= (\bar{y}_{i.t} - \bar{y}_{N_1.t}) - (\bar{x}'_{i.t} - \bar{x}'_{N_1.t})\hat{\beta} \\
\hat{\alpha}^*_{jt} &= (\bar{y}_{.jt} + \bar{y}_{N_1.T} - \bar{y}_{..t}) - (\bar{x}'_{.jt} + \bar{x}'_{N_1.T} - \bar{x}'_{..t})\hat{\beta} \,.
\end{aligned}$$

For model (1.7), we leave out γ_{ij} for $i = N_1$, α_{it} for $t = T$, and α^*_{jt} for $j = N_2$, and add back a general constant c. In this way, exactly $N_2 + N_1 + T - 1$ intercept parameters are eliminated, so the dummy matrix D^* has full rank. The estimators, with this D^* reads in scalar form

$$\begin{aligned}
\hat{c} &= (\bar{y}_{N_1N_2.} + \bar{y}_{N_1.T} + \bar{y}_{.N_2T} - \bar{y}_{N_1..} - \bar{y}_{.N_2.} - \bar{y}_{..T} + \bar{y}_{...}) \\
&\quad - (\bar{x}'_{N_1N_2.} + \bar{x}'_{N_1.T} + \bar{x}'_{.N_2T} - \bar{x}'_{N_1..} - \bar{x}'_{.N_2.} - \bar{x}'_{..T} + \bar{x}'_{...})\hat{\beta} \\
\bar{\gamma}_{ij} &= (\bar{y}_{ij.} - \bar{y}_{N_1j.} + \bar{y}_{i.T} - \bar{y}_{N_1.T} - \bar{y}_{i..} + \bar{y}_{N_1..}) \\
&\quad - (\bar{x}'_{ij.} - \bar{x}'_{N_1j.} + \bar{x}'_{i.T} - \bar{x}'_{N_1.T} - \bar{x}'_{i..} + \bar{x}'_{N_1..})\hat{\beta} \\
\bar{\alpha}_{it} &= (\bar{y}_{i.t} - \bar{y}_{i.T} + \bar{y}_{.N_2t} - \bar{y}_{.N_2T} - \bar{y}_{..t} + \bar{y}_{..T}) \\
&\quad - (\bar{x}'_{i.t} - \bar{x}'_{i.T} + \bar{x}'_{.N_2t} - \bar{x}'_{.N_2T} - \bar{x}'_{..t} + \bar{x}'_{..T})\hat{\beta} \\
\bar{\alpha}^*_{jt} &= (\bar{y}_{.jt} - \bar{y}_{.N_2t} + \bar{y}_{N_1j.} - \bar{y}_{N_1N_2.} - \bar{y}_{.j.} + \bar{y}_{.N_2.}) \\
&\quad - (\bar{x}'_{.jt} - \bar{x}'_{.N_2t} + \bar{x}'_{N_1j.} - \bar{x}'_{N_1N_2.} - \bar{x}'_{.j.} + \bar{x}'_{.N_2.})\hat{\beta}
\end{aligned}$$

Now, that we have derived appropriate estimators for all models, it is time to assess their properties. In finite samples, the OLS assumptions imposed guarantee that all estimators derived above are BLUE, with finite sample variances

$$\mathrm{Var}(\hat{\beta}) = \sigma^2_\varepsilon (X' M_D X)^{-1}$$

with the appropriate M_D, and

$$\mathrm{Var}(\hat{\pi}^*) = \sigma^2_\varepsilon (D^{*\prime} D^*)^{-1} + (D^{*\prime} D^*)^{-1} D^{*\prime} X \,\mathrm{Var}(\hat{\beta}) X' D^* (D^{*\prime} D^*)^{-1} \,.$$

As σ^2_ε is usually unknown, we have to replace σ^2_ε by its estimator

$$\hat{\sigma}^2_\varepsilon = \frac{1}{\mathrm{rank}(M_D) - K} \sum_{i,j,t} \hat{\tilde{\varepsilon}}^2_{ijt} \,,$$

where

$$\hat{\tilde{\varepsilon}}^2_{ijt} = (\tilde{y}_{ijt} - \tilde{x}'_{ijt}\hat{\beta})^2 \tag{1.9}$$

is the transformed residual square and $(\text{rank}(M_D) - K)$ is collected for all models in the last column of Table 1.2.

As multi-dimensional panel data are usually large in one or more directions, it is important to have a closer look at the asymptotic properties as well. Unlike cross-sectional, or time series data, panels can grow in multiple dimensions at the same time. As a matter of fact, three-way panel data can fall in one of the following seven asymptotic cases:

- $N_1 \to \infty$, N_2, T fixed; $N_2 \to \infty$, N_1, T fixed; $T \to \infty$, N_1, N_2 fixed
- $N_1, N_2 \to \infty$, T fixed; $N_1, T \to \infty$, N_2 fixed; $N_2, T \to \infty$, N_1 fixed
- $N_1, N_2, T \to \infty$.

It can be shown, that $\hat{\beta}$ is consistent in all of the asymptotic cases for all models (if some weak properties hold). In order to make the models feasible for inference (*i.e.*, for testing), we have to normalize the variances according to the asymptotics considered. When, for example, N_1 goes to infinity, and N_2 and T are fixed, $N_1 \operatorname{Var}(\hat{\beta})$ is finite in the limit, as

$$\text{plim}_{N_1 \to \infty} N_1 \operatorname{Var}(\hat{\beta}) = \sigma_\varepsilon^2 \, \text{plim}_{N_1 \to \infty} \left(\frac{X' M_D X}{N_1} \right)^{-1} = \sigma_\varepsilon^2 Q_{XMX}^{-1} ,$$

where Q_{XMX} is assumed to be a finite, positive semi-definite matrix. The estimators of fixed effects are consistent only if at least one of the indices with which they are fixed with, is growing. For example, for model (1.2), $\hat{\alpha}_i$ is consistent only if N_2 and/or T is going to infinity, and its variance is finite, and in addition, if it is pre-multiplied by N_2, in the case of $N_2 \to \infty$, by T, in the case of $T \to \infty$, and by $N_2 T$, when $N_2, T \to \infty$.

Testing for parameter values or restrictions is done in the usual way, using standard t-tests or F-tests. Typically, when one seeks to test $\beta_k = 0$ or $\alpha_i = 0$, for example in model (1.2), the t-statistic is

$$\hat{\beta}_k / \sqrt{\hat{\operatorname{Var}}(\hat{\beta}_k)} \quad \text{and} \quad \hat{\alpha}_i / \sqrt{\hat{\operatorname{Var}}(\hat{\alpha}_i)} ,$$

where $\operatorname{Var}(\hat{\beta}_k)$ is the k-th diagonal element of $\operatorname{Var}(\hat{\beta})$, and $\operatorname{Var}(\alpha_i)$ is the diagonal element from $\operatorname{Var}(\hat{\pi}^*)$ corresponding to α_i. The degrees of freedom has to be adjusted accordingly, for each model, as Table 1.2 shows. The same degrees of freedom should be used when testing for the slope parameters and/or for the fixed effects of a given a model. It is not typical, however, to test for the significance of one particular α_i, unless that individual plays some specific role in the model. Usually we are more concerned with the *joint* existence of the individual parameters, in other words, with testing for $\alpha_1 = \alpha_2 = \ldots = \alpha_{N_1}$. Keep using model (1.2) for the illustration, and assuming normality, the statistic for the F-test is obtained as in

$$F = \frac{(R^2_{UR} - R^2_R)/(N_1 - 1)}{(1 - R^2_{UR})/(N_1 N_2 T - N_1 - N_2 - T + 1 - K)}$$

where R^2_{UR} is the R^2 of the *unrestricted model* (that is the full model (1.2)), while R^2_R is the R^2 of the restricted model, that is model (1.2) without the α_i individual effects. The null hypothesis puts $(N_1 - 1)$ restrictions on the parameters, while the degrees of freedom of the unrestricted model is simply $(N_1 N_2 T - N_1 - N_2 - T + 1 - K)$. This statistic then has an F-distribution with $(N_1 - 1, N_1 N_2 T - N_1 - N_2 - T + 1 - K)$ degrees of freedom.

1.4 Incomplete Panels

As in the case of the usual 2D panel data sets (see Wansbeek and Kapteyn (1989) or Baltagi (2005), for example), just more frequently, one may be faced with situations in which the data at hand is unbalanced. In our framework of analysis this means that $t \in T_{ij}$, for all (ij) pairs, where T_{ij} is a subset of the index set $t \in \{1, \ldots, T\}$, with T being chronologically the last time period in which we have any (i, j) observations. Note that two T_{ij} and $T_{i'j'}$ sets are usually different. A special case of incompleteness, which typically characterizes flow-type data, is the so-called no self-flow. In such data sets the individual index sets i and j are the same, so $N_1 = N_2 = N$ holds. Formally, this means, that, for all t, there are no observations when $i = j$, that is, we are missing a total NT of data points. In this section, however, we only consider general incompleteness, and take the no self-flow issue under lenses in Section 1.5.

In the case of incomplete data, the models can still be casted as in (1.1), but now D can not be represented nicely by kronecker products, as done in Table 1.1. However, with the incompleteness adjusted dummy matrices, $\tilde{D}$ (which we obtain from D by leaving out the rows corresponding to missing observations), the LSDV estimator of β and the fixed effects can still be worked out, maintaining its BLUE properties, following (1.8). There is, however, one practical obstacle in the way. Remember, that to reach $\hat{\beta}$ conveniently, we needed the exact form of M_D, which we collected for complete data in Table 1.2. As $\tilde{D}$ has a known form only if we know which observations are missing exactly, $M_{\tilde{D}} = I - \tilde{D}(\tilde{D}'\tilde{D})^{-}\tilde{D}'$ can not be defined element-wise analytically in general, where "$-$" stands for any generalized inverse. Instead, we have to invert $(\tilde{D}'\tilde{D})$ directly, or use partitioned matrix inversion. Either way, we usually can not avoid large computational burdens when carrying out (1.8) in case of incompleteness (as opposed to no computational burden when the data is complete).[2] Nevertheless, the estimators and the covariance matrices are obtained in the same way as for complete data (of course, after adjusting the matrices to incompleteness), and the properties of the estimators are the same as in the complete data case. Notice the crucial difference between $\tilde{D}$ and D^*: while $\tilde{D}$ has usually no full column rank, as we left out some rows from D (which also in general has no full column rank), D^* is simply designed to have full column rank (more precisely, to fix the rank deficiency in D). That is why we have to turn to generalized inverses for the former, but is enough to work with "simple" inverses for the latter dummy matrices.

Incompleteness is less of an issue in case of 2D models, where T is usually small, and N_1 is large (that is when we have one high dimensional fixed effects), but is generally present in case of 3D data, where typically along with N_1, N_2 is also large. In practice, to alleviate this issue with the large dimensions, the best solution seems to be to turn to iterative solutions. One of the most widely used is based on the work of Carneiro et al. (2008), and later on Guimaraes and Portugal (2009). Let us show the procedure on model (1.2), the rest is a direct consequence. Model (1.2) in matrix form reads as

$$y = X\beta + \tilde{D}_1\alpha + \tilde{D}_2\gamma + \tilde{D}_3\lambda + \varepsilon\,, \tag{1.10}$$

where $\tilde{D}_k$ meant to stress, that the data is possibly incomplete: from the original $D_1 = (I_{N_1} \otimes \iota_{N_2T})$, $D_2 = (\iota_{N_1} \otimes I_{N_2} \otimes \iota_T)$, and $D_3 = (\iota_{N_1T} \otimes I_T)$, the rows matching with the missing observations are deleted. The normal equations from (1.10) are

$$\begin{aligned}
\beta &= (X'X)^{-1}X'(y - \tilde{D}_1\alpha - \tilde{D}_2\gamma - \tilde{D}_3\lambda) \\
\alpha &= (\tilde{D}_1'\tilde{D}_1)^{-}\tilde{D}_1'(y - X\beta - \tilde{D}_2\gamma - \tilde{D}_3\lambda) \\
\gamma &= (\tilde{D}_2'\tilde{D}_2)^{-}\tilde{D}_2'(y - X\beta - \tilde{D}_1\alpha - \tilde{D}_3\lambda) \\
\lambda &= (\tilde{D}_3'\tilde{D}_3)^{-}\tilde{D}_3'(y - X\beta - \tilde{D}_1\alpha - \tilde{D}_2\gamma)\,,
\end{aligned}$$

which suggests the so-called Gauss-Seidel, also called the "zigzag" algorithm, that is, we alternate between the estimation of β, and the fixed effects parameters, starting from some arbitrary initial values β^0, and $(\alpha^0, \gamma^0, \lambda^0)$. The computational improvement is clear: $(\tilde{D}_k'\tilde{D}_k)^{-}\tilde{D}_k$ defines a simple group average ($k = 1,2,3$) of the residuals, so the dimensionality issue is no longer a concern. Specifically, $(\tilde{D}_1'\tilde{D}_1)^{-}\tilde{D}_1'$ is translated into an average over (jt), $(\tilde{D}_2'\tilde{D}_2)^{-}\tilde{D}_2'$ an average over (it), and $(\tilde{D}_3'\tilde{D}_3)^{-}\tilde{D}_3'$ an average over (ij). Furthermore, $\tilde{D}_1\alpha$, etc. are just the columns of the current estimates of α, etc. After the sufficient number of steps, the iterative estimators all converge to the true LSDV.[3]

[2] Actually the sparsity of $(\tilde{D}'\tilde{D})$ can help to reduce the computation. The study of sparse matrices has grown into a separate field in the past years offering numerous tools to go around (or at least attenuate) the "curse of dimensionality". This topic is however beyond the expertise of the authors, and the scope of the text.

[3] The STATA program command *reg2hdfe* implements these results and can be found in the STATA

10.14754/CEU.2017.05

1.5 The Within Estimator

1.5.1 The Equivalence between LSDV and the Within Estimator

As seen, LSDV estimates all parameters of the fixed effects models in one step. Usually, however, the individual/time parameters are outside of the center of attention, and we only care about the estimates of the focus local parameters. There is, luckily, an other appealing way to approach the estimation problem. The idea is that using orthogonal projections the slope parameters (and if needed the fixed effects) are estimated separately. First, with a projection orthogonal to D, we transform the model, in fact y and X, in such a way that clears the fixed effects. Then, we carry out an OLS estimation on the transformed variables $\tilde{y}$ and $\tilde{X}$. We have to point out, however, that unlike in the case of 2D models, there are usually multiple such Within transformations, which eliminate the fixed effects. Nevertheless, only the Within estimator based on the Within transformation originating from the LSDV conserves the BLUE properties and therefore is called the optimal one. To show this, notice, that as M_D is idempotent, the first part of formula (1.8) is equivalent to performing an OLS on

$$M_D y = M_D X\beta + \underbrace{M_D D}_{0}\pi + M_D\varepsilon\,,$$

where $M_D = I - D(D'D)^{-}D'$, as before. In the case of complete data, M_D can be translated into scalar notation, so we can fully avoid the dimensionality issue. Let us now go through all the models, and present the scalar form of the optimal Within transformation $M_D y$.

For model (1.2), the optimal transformation is

$$\tilde{y}_{ijt} = y_{ijt} - \bar{y}_{i..} - \bar{y}_{.j.} - \bar{y}_{..t} + 2\bar{y}_{...}\,. \tag{1.11}$$

As mentioned above, the uniqueness of the Within transformation is not guaranteed: for example transformation

$$\tilde{y}_{ijt} = y_{ijt} - \bar{y}_{ij.} - \bar{y}_{..t} + \bar{y}_{...} \tag{1.12}$$

also eliminates the fixed effects from model (1.2). For model (1.3), the transformation is simply

$$\tilde{y}_{ijt} = y_{ijt} - \bar{y}_{ij.}\,. \tag{1.13}$$

For model (1.4), the optimal Within transformation is in fact (1.12). Notice, that model (1.2) is a special case of model (1.4) (with the restriction $\gamma_{ij} = \alpha_i + \gamma_j$),

Documentation. The code is designed to tackle two high dimensional fixed effects, however, it can be improved to treat three, or even more fixed effects at the same time.

so while transformation (1.12) is optimal for (1.4), it is clear why it is not for the former: it 'over-clears' the fixed effects, by not using the extra piece of information.

For model (1.5), the transformation is

$$\tilde{y}_{ijt} = y_{ijt} - \bar{y}_{.jt}\,, \tag{1.14}$$

while for models (1.6) and (1.7), they are

$$\tilde{y}_{ijt} = y_{ijt} - \bar{y}_{.jt} - \bar{y}_{i.t} + \bar{y}_{..t}\,, \tag{1.15}$$

and

$$\tilde{y}_{ijt} = y_{ijt} - \bar{y}_{ij.} - \bar{y}_{.jt} - \bar{y}_{i.t} + \bar{y}_{..t} + \bar{y}_{.j.} + \bar{y}_{i..} - \bar{y}_{...}\,, \tag{1.16}$$

respectively.

It can be seen, that the Within transformation work perfectly in wiping out the fixed effects. However, frequently in empirical applications, some explanatory variables, (*i.e.*, some elements of the vector x'_{ijt}) do not span the whole (ijt) data space, that is, has some kind of '*index deficiency*'. This means, that sometimes one (or more) of the regressors are perfectly collinear with one of the fixed effects. In such cases, we can consider that regressor as fixed, as it is wiped out along with the fixed effects. For example, for model (1.3), if we put an individual's gender among the regressors, $x_{ijt} \equiv x_i$ holds, and so is eliminated by the Within transformation (1.11). Clearly, parameters associated with such regressors then can not be estimated. This is most visible for model (1.8), as in that case all regressors fixed at least in one dimension are excluded from the model automatically after the Within transformation (1.16).

As seen, Within transformations eliminate *all* fixed effects parameters with linear combinations of some group means on the underlying data. In fact, all transformations (1.11)–(1.16) are similar in being "complete", as opposed to "partial" transformations, which only eliminate some fixed effects parameters. Take model (1.2), for example, and assume that the number of individual i-s is extensive (large N_1), but the panel is short in the other two dimensions (N_2 and T are small). While we can proceed directly with the Within transformation (1.11) and wipe out *all* fixed effects, we can alternatively transform out α_i only with

$$\tilde{y}_{ijt} = y_{ijt} - \bar{y}_{i..}\,, \tag{1.17}$$

and incorporate the rest, γ_j and λ_t, explicitly to the model, ending up with

$$\tilde{y}_{ijt} = \tilde{x}'_{ijt}\beta + \tilde{\gamma}_j + \tilde{\lambda}_t + \tilde{\varepsilon}_{ijt}\,, \tag{1.18}$$

directly estimable with OLS. Clearly estimator 1.18, based on the *partial* transformation gives numerically the same estimator for β than the Within estimator, yet the Within estimator is less computationally complex in case of complete data, as only a number of K parameters is estimated, not $K+N_2+T$. However, as we will see in Section 1.5.2, if the underlying data is incomplete, Within transformations in general can not be represented with nice sample means, instead, some matrix operations have to be introduced. The same is not necessarily true for the *partial* transformations: (1.17) for example still defines a group mean over jt, no matter what form of incomplete data we have. The computational convenience is then obvious: if the Within transformation is hard to implement in incomplete data, and at the same time some dimensions are "short", there always exists a *partial* transformation, potentially robust to incomplete data,[4] wiping out the "long" effects only, letting the "short" effects to be incorporated to the model directly. Getting back to the example of model (1.2), while the Within transformation (1.11) involves matrix operations in case of incomplete data, estimating the transformed model (1.18) (based on the *partial* transformation (1.17)) becomes computationally less demanding.

Some caution is needed, however, when using *partial* transformations instead of *complete* ones. As partial transformations only eliminate some fixed effects parameters, non-eliminated ones still have to be incorporated to the model. Failure to do so results possibly biased and inconsistent estimators. If transformation (1.17) is employed on model (1.2), but $\tilde{\gamma}_j$ and $\tilde{\lambda}_t$, the non-eliminated fixed effects parameters are ignored during estimation, that is

$$\tilde{y}_{ijt} = \tilde{x}'_{ijt}\beta + \tilde{u}_{ijt}$$

is estimated with the $\tilde{u}_{ijt}$ disturbance where

$$\tilde{u}_{ijt} = \tilde{\gamma}_j + \tilde{\lambda}_t + \tilde{\varepsilon}_{ijt}\,,$$

a classical case of omitted variable problem emerges. Unless $\tilde{\gamma}_j = \tilde{\lambda}_t \equiv 0$, or the correlation is nil between the regressors and the omitted dummy variables, the estimator for β will be biased and inconsistent.

1.5.2 Incomplete Panels with the Within Estimator

We have covered briefly incompleteness in Section 1.4 already, but the Within estimators, and the underlying transformations, open a new way to deal with it.

No Self-flow Data

Let us start with the no self-flow data, and for a short time, assume, that the index sets i and j are the same, and so $N_1 = N_2 = N$.

In terms of the models from Section 1.2, the scalar transformations introduced

[4] Robustness here means that the partial transformation to be employed can be represented with group means even if the data is incomplete.

there can no longer be applied. Fortunately, the pattern of the missing observations is highly structured, allowing for the derivation of optimal transformations that are still quite simple and maintain the BLUE properties of the Within estimators based on them. Following the derivations of Balazsi et al. (2015), the transformation for the models are the following:

$$\begin{aligned} \tilde{y}_{ijt} &= y_{ijt} - \tfrac{N-1}{N(N-2)T}(y_{i++} + y_{+j+}) - \tfrac{1}{N(N-2)T}(y_{j++} + y_{+i+}) \\ &\quad - \tfrac{1}{N(N-1)} y_{++t} + \tfrac{2}{N(N-2)T} y_{+++} \end{aligned} \tag{1.19}$$

for model (1.2), and

$$\tilde{y}_{ijt} = y_{ijt} - \frac{1}{T} y_{ij+} \tag{1.20}$$

for model (1.3). For models (1.4), and (1.5) the no self-flow transformations are

$$\tilde{y}_{ijt} = y_{ijt} - \frac{1}{T} y_{ij+} - \frac{1}{N(N-1)} y_{++t} + \frac{1}{TN(N-1)} y_{+++}\,, \tag{1.21}$$

and

$$\tilde{y}_{ijt} = y_{ijt} - \frac{1}{N-1} y_{+jt}\,, \tag{1.22}$$

while for models (1.6), and (1.7), they are

$$\begin{aligned} \tilde{y}_{ijt} &= y_{ijt} - \tfrac{N-1}{N(N-2)}(y_{i+t} + y_{+jt}) - \tfrac{1}{N(N-2)}(y_{+it} + y_{j+t}) \\ &\quad + \tfrac{1}{(N-1)(N-2)} y_{++t}\,, \end{aligned} \tag{1.23}$$

and

$$\begin{aligned} \tilde{y}_{ijt} = \quad & y_{ijt} - \tfrac{N-3}{N(N-2)}(y_{i+t} + y_{+jt}) + \tfrac{N-3}{N(N-2)T}(y_{i++} + y_{+j+}) - \tfrac{1}{T} y_{ij+} \\ & + \tfrac{1}{N(N-2)}(y_{+it} + y_{j+t}) - \tfrac{1}{N(N-2)T}(y_{+i+} + y_{j++}) \\ & + \tfrac{N^2-6N+4}{N^2(N-1)(N-2)}(y_{++t} - y_{+++}) \end{aligned} \tag{1.24}$$

respectively. For the proof of the no self-flow transformations, and for an insightful detour on the no self-flow issue of purely cross-sectional panels, see Appendix B. So overall, the self-flow data problem can be overcome by using an appropriate Within transformation. Optimality of the estimators is preserved, as the transformations are all derived from the Frisch-Waugh-Lovell theorem.

General Incompleteness

Next, let us go along these lines, and work out suitable Within transformations for any general form of incompleteness. Now, we are back in the case when i and j are different index sets. As the expressions below are all derived from the Frisch-Waugh-Lovell theorem, the transformations are optimal, and the estimators are BLUE. Remember, that now $t \in T_{ij}$, and let $R = \sum_{ij} |T_{ij}|$ denote the total number of observations, where $|T_{ij}|$ is the cardinality of the set T_{ij} (the number of observations in the given set).

For models (1.3) and (1.5), the unbalanced nature of the data does not cause any problem (since they are in fact can be represented as 2D models of one fixed effect), the Within transformations can be used, and they have exactly the same properties, as in the balanced case. However, for models (1.2), (1.4), (1.6), and (1.7), we face some problems. As the Within transformations fail to fully eliminate the fixed effects for these models (somewhat similarly to the no self-flow case), the resulting Within estimators suffer from (potentially severe) biases. However, the Wansbeek and Kapteyn (1989) approach, can be extended to these four cases (in a slightly different manner than in Davis (2002)).

Let us start with model (1.2). The dummy variable matrix D has to be modified to reflect the unbalanced nature of the data. Let the U_t and V_t $(t = 1 \dots T)$ be the sequence of $(I_{N_1} \otimes \iota_{N_2})$ and $(\iota_{N_1} \otimes I_{N_2})$ matrices, respectively, in which the following adjustments are made: for each (ij) observation, we leave the row (representing (ij)) in U_t and V_t matrices untouched where $t \in T_{ij}$, but delete it from the remaining $T - |T_{ij}|$ matrices. In this way we end up with the following dummy variable setup

$$\begin{aligned} D_1^a &= (U_1', U_2', \dots, U_T')' \quad \text{of size} \quad (R \times N_1), \\ D_2^a &= (V_1', V_2', \dots, V_T')' \quad \text{of size} \quad (R \times N_2), \text{ and} \\ D_3^a &= \operatorname{diag}\{V_1 \cdot \iota_{N_1}, V_2 \cdot, \iota_{N_1} \dots, V_T \cdot \iota_{N_1}\} \quad \text{of size} \quad (R \times T). \end{aligned}$$

So the complete dummy variable structure is now $D_a = (D_1^a, D_2^a, D_3^a)$. In this case, let us note here that, just as in Wansbeek and Kapteyn (1989), index t goes 'slowly' and ij goes 'fast'. Using this modified dummy variable structure, the optimal projection removing the fixed effects can be obtained in three steps:

$$\begin{aligned} M_{D_a}^{(1)} &= I_R - D_1^a (D_1^{a'} D_1^a)^{-1} D_1^{a'}, \\ M_{D_a}^{(2)} &= M_{D_a}^{(1)} - M_{D_a}^{(1)} D_2^a (D_2^{a'} M_{D_a}^{(1)} D_2^a)^{-} D_2^{a'} M_{D_a}^{(1)}, \end{aligned}$$

and finally

$$M_{D_a} = M_{D_a}^{(3)} = M_{D_a}^{(2)} - M_{D_a}^{(2)} D_3^a (D_3^{a'} M_{D_a}^{(2)} D_3^a)^{-} D_3^{a'} M_{D_a}^{(2)}. \tag{1.25}$$

It is easy to see that in fact $M_{D_a} D_a = 0$ projects out all three dummy matrices. Note that in the balanced case $(D_1^{a'} D_1^a)^{-1} = I_{N_1}/(N_2 T)$, but now

$$(D_1^{a'} D_1^a)^{-1} = \operatorname{diag}\left\{ \frac{1}{\sum_j |T_{1j}|}, \frac{1}{\sum_j |T_{2j}|}, \dots, \frac{1}{\sum_j |T_{N_1 j}|} \right\} \quad \text{of size} \quad (N_1 \times N_1).$$

With this in hand, we only have to calculate two inverses, $(D_2^{a'} M_{D_a}^{(1)} D_2^a)^{-}$, and $(D_3^{a'} M_{D_a}^{(2)} D_3^a)^{-}$ with respective sizes $(N_2 \times N_2)$ and $(T \times T)$ instead of three. This is feasible for reasonable sample sizes.

For model (1.4), the job is essentially the same. Let the W_t $(t = 1 \dots T)$ be the sequence of $(I_{N_1N_2} \otimes I_{N_1N_2})$ matrices, where again for each (ij), we remove the rows corresponding to observation (ij) in those W_t, where $t \notin T_{ij}$. In this way,

$$\begin{aligned} D_1^b &= (W_1', W_2', \dots, W_T')' \quad \text{of size} \quad (R \times N_1N_2)\,, \\ D_2^b &= D_3^a \quad \text{of size} \quad (R \times T)\,. \end{aligned}$$

The first step in the projection is now

$$M_{D_b}^{(1)} = I_R - D_1^b (D_1^{b'} D_1^b)^{-1} D_1^{b'}\,,$$

so the optimal projection orthogonal to $D_b = (D_1^b, D_2^b)$ is simply

$$M_{D_b} = M_{D_b}^{(2)} = M_{D_b}^{(1)} - M_{D_b}^{(1)} D_2^b (D_2^{b'} M_{D_b}^{(1)} D_2^b)^{-} D_2^{b'} M_{D_b}^{(1)}\,. \tag{1.26}$$

Note that as

$$(D_1^{b'} D_1^b)^{-1} = \operatorname{diag}\left\{ \frac{1}{|T_{11}|}, \frac{1}{|T_{12}|}, \dots, \frac{1}{|T_{N_1N_2}|} \right\} \quad \text{of size} \quad (N_1N_2 \times N_1N_2)\,,$$

we only have to calculate the inverse of a $(T \times T)$ matrix – $D_2^{b'} M_{D_b}^{(1)} D_2^b$ – which is easily doable. Further, as discussed above, given that model (1.2) is nested in (1.4), transformation (1.26) is in fact also valid for model (1.2).

Let us move on to model (1.6). Now, after the same adjustments as before,

$$\begin{aligned} D_1^c &= \operatorname{diag}\{U_1, U_2, \dots, U_T\} \quad \text{of size} \quad (R \times N_1T) \quad \text{and} \\ D_2^c &= \operatorname{diag}\{V_1, V_2, \dots, V_T\} \quad \text{of size} \quad (R \times N_2T)\,, \end{aligned}$$

so the stepwise projection, removing $D_c = (D_1^c, D_2^c)$, is

$$M_{D_c}^{(1)} = I_R - D_1^c (D_1^{c'} D_1^c)^{-1} D_1^{c'}\,,$$

leading to

$$M_{D_c} = M_{D_c}^{(2)} = M_{D_c}^{(1)} - M_{D_c}^{(1)} D_2^c (D_2^{c'} M_{D_c}^{(1)} D_2^c)^{-} D_2^{c'} M_{D_c}^{(1)}\,. \tag{1.27}$$

Note that for M_{D_c}, we have to invert $(N_2T \times N_2T)$ matrices, which can be computationally difficult.

The last model to deal with is model (1.7). Let $D_d = (D_1^d, D_2^d, D_3^d)$, where the adjusted dummy matrices are all defined above:

$$\begin{aligned} D_1^d &= D_1^b \quad \text{of size} \quad (R \times N_1N_2)\,, \\ D_2^d &= D_1^c \quad \text{of size} \quad (R \times N_1T)\,, \\ D_3^d &= D_2^c \quad \text{of size} \quad (R \times N_2T)\,. \end{aligned}$$

Defining the partial projector matrices $M^{(1)}_{D_d}$ and $M^{(2)}_{D_d}$ as

$$\begin{aligned} M^{(1)}_{D_d} &= I_R - D^d_1(D^{d'}_1 D^d_1)^{-1} D^{d'}_1 \text{ and} \\ M^{(2)}_{D_d} &= M^{(1)}_{D_d} - M^{(1)}_{D_d} D^{d'}_2 (D^{d'}_2 M^{(1)}_{D_d} D^d_2)^{-} D^{d'}_2 M^{(1)}_{D_d} , \end{aligned}$$

the appropriate transformation for model (1.7) is now

$$M_{D_d} = M^{(3)}_{D_d} = M^{(2)}_{D_d} - M^{(2)}_{D_d} D^{d'}_3 (D^{d'}_3 M^{(2)}_{D_d} D^d_3)^{-} D^{d'}_3 M^{(2)}_{D_d} . \tag{1.28}$$

It can be easily verified that M_{D_d} is idempotent and $M_{D_d} D_d = 0$, so all the fixed effects are indeed eliminated.[5] As model (1.6) is covered by model (1.7), projection (1.28) eliminates the fixed effects from that model as well. Moreover, as suggested above, all three-way fixed effects models are in fact nested into model (1.7). It is therefore intuitive that transformation (1.28) clears the fixed effects in all model formulations. Using (1.7) is not always advantageous though, as the transformation involves the inversion of potentially large matrices (of order N_1T, and N_2T). In the case of most models studied, we can find suitable unbalanced transformations at the cost of only inverting $(T \times T)$ matrices; or in some cases, we can even derive scalar transformations. It is good to know, however, that there is a general projection that is universally applicable to all three-way models in the presence of all kinds of data issues.

It is worth noting that transformations (1.25), (1.26), (1.27), and (1.28) are all dealing in a natural way with the no self-flow problem, as only the rows corresponding to the $i = j$ observations need to be deleted from the corresponding dummy variable matrices.

All transformations detailed above can also be rewritten in a semi-scalar form. Let us show here how this idea works on transformation (1.28), as all subsequent transformations can be dealt with in the same way. Let

$$\phi = C^{-} \bar{D}' y \quad \text{and} \quad \omega = \tilde{C}^{-} (M^{(2)}_{D_d} D^d_3)' y \quad \xi = C^{-} \bar{D}' D^d_3 \omega ,$$

where

$$C = \left(D^d_2\right)' \bar{D} , \quad \bar{D} = \left(I_R - D^d_1(D^{d'}_1 D^d_1)^{-1} D^{d'}_1\right) D^d_2 , \text{ and} \quad \tilde{C} = D^{d'}_3 M^{(2)}_{D_d} D^d_3 .$$

Now the scalar representation of transformation (1.28) is

$$\begin{aligned} [M_{D_d} y]_{ijt} = \;& y_{ijt} - \tfrac{1}{|T_{ij}|} \textstyle\sum_{t \in T_{ij}} y_{ijt} + \tfrac{1}{|T_{ij}|} a'_{ij} \phi - \phi_{it} \\ & - \omega_{jt} + \tfrac{1}{|T_{ij}|} \tilde{a}'_{ij} \omega + \xi_{it} - \tfrac{1}{|T_{ij}|} \left(a^b_{ij}\right)' \xi , \end{aligned}$$

where a_{ij} and $\tilde{a}_{ij}$ are the column vectors corresponding to observations (ij) from

[5] A STATA program code for transformation (1.28) with a user-friendly detailed explanation is available at *www.personal.ceu.hu/staff/repec/pdf/stata-program_document-dofile.pdf.* Estimation of model (1.7) is then easily done for any kind of incompleteness.

matrices $A = D_2^{d'} D_1^d$ and $\tilde{A} = D_3^{d'} D_1^d$, respectively; ϕ_{it} is the (it)-th element of the $(N_1 T \times 1)$ column vector ϕ; ω_{jt} is the (jt)-th element of the $(N_2 T \times 1)$ column vector ω; and finally, ξ_{it} is the element corresponding to the (it)-th observation from the $(N_1 T \times 1)$ column vector, ξ.[6]

1.6 Dynamic Models

In the case of dynamic autoregressive models, the use of which is unavoidable if the data generating process has partial adjustment or some kind of memory, the Within estimators in a usual panel data framework are biased. In this section we generalize these well known results to this higher dimensional setup. We first derive a general semi-asymptotic bias formulae, then we make it specific for each of the models introduced in Section 1.2, lastly we propose consistent estimators for the problematic models.

1.6.1 Nickell Biases

The models of Section 1.2 can all be written in the general dynamic form

$$y = \rho y_{-1} + D\pi + \varepsilon\,, \tag{1.29}$$

where D and π correspond to any of the specific D and π discussed in Section 1.2. With M_D the projection matrix orthogonal to D,

$$\hat{\rho} = \frac{y'_{-1} M_D y}{y'_{-1} M_D y_{-1}} = \rho + \frac{\mathrm{tr}(M_D \varepsilon y'_{-1})}{\mathrm{tr}(M_D y_{-1} y'_{-1})}\,, \tag{1.30}$$

[6] From a computational point of view, the calculation of matrix M_{D_d} is by far the most resource requiring as we have to invert $(N_1 T \times N_1 T)$, and $(N_2 T \times N_2 T)$ sized matrices. Simplifications related to this can dramatically reduce CPU and storage requirements. This topic, however, is well beyond the scope of this chapter.

where y and y_{-1} are the column vectors of dependent and lagged dependent variables respectively of size N_1N_2T. Let

$$
\begin{aligned}
B_0 &= \begin{pmatrix} 0 & 0 \\ I_{T-1} & 0 \end{pmatrix} \quad \text{of size} \quad (T \times T)\,, \\
\Gamma_0 &= (I_T - \rho B_0)^{-1} = \begin{pmatrix} 1 & \dots & \dots & 0 \\ \rho & \ddots & & \vdots \\ & \ddots & \ddots & \vdots \\ \rho^{T-1} & & \rho & 1 \end{pmatrix} \quad \text{of size} \quad (T \times T)\,, \\
\Psi_0 &= \begin{pmatrix} 1 & \rho & & \rho^{T-1} \\ \rho & \ddots & \ddots & \\ & \ddots & \ddots & \rho \\ \rho^{T-1} & & \rho & 1 \end{pmatrix} = I_T + \rho(\Gamma_0 B_0 + (\Gamma_0 B_0)')
\end{aligned}
$$

of size $(T \times T)$, and let $B = I_{N_1N_2} \otimes B_0, \Gamma = I_{N_1N_2} \otimes \Gamma_0, \Psi = I_{N_1N_2} \otimes \Psi_0$ define matrices necessary for the general bias formulae. With e_1 the first unit vector of size $(T \times 1)$ and y_0 having N_1N_2 elements (the initial values of the y_{ijt} for all (ij) pair),

$$By = y_{-1} - y_0 \otimes e_1\,,$$

and so the model (1.29) can be rewritten as

$$y = \rho By + \rho y_0 \otimes e_1 + D\pi + \varepsilon\,, \quad \text{or} \quad (I_{N_1N_2T} - \rho B)y = \rho y_0 \otimes e_1 + D\pi + \varepsilon\,,$$

which ultimately leads to

$$y = \rho\Gamma(y_0 \otimes e_1) + \Gamma D\pi + \Gamma\varepsilon\,.$$

Let ε_+ be ε advanced by one time period. Then, under the stationarity of ε_{ijt},

$$\mathrm{E}(y_{-1}\varepsilon') = \mathrm{E}(y\varepsilon'_+) = \Gamma\,\mathrm{E}(\varepsilon\varepsilon'_+) = \sigma^2_\varepsilon \Gamma B\,.$$

So for the expectation of the numerator in (1.30) we obtain

$$\mathrm{E}(\mathrm{tr}(M_D\varepsilon y'_{-1})) = \sigma^2_\varepsilon \,\mathrm{tr}(M_D\Gamma B) = \frac{\sigma^2_\varepsilon}{2\rho}(\mathrm{tr}(M_D\Psi) - \mathrm{tr}(M_D))\,,$$

with $\Psi = (I_{N_1N_2T} + \rho(\Gamma B + (\Gamma B)'))$. For the denominator in (1.30),

$$
\begin{aligned}
\mathrm{E}(\mathrm{tr}(M_D y_{-1}y'_{-1})) &= \mathrm{E}(\mathrm{tr}(M_D yy')) \\
&= \rho^2\,\mathrm{E}(\mathrm{tr}(M_D y_{-1}y'_{-1})) + \sigma^2_\varepsilon\,\mathrm{tr}(M_D) + 2\,\mathrm{E}(\mathrm{tr}(M_D\varepsilon y'_{-1}))\,,
\end{aligned}
$$

so, as $\mathrm{E}(\mathrm{tr}(M_D\varepsilon y'_{-1})) = \sigma^2_\varepsilon\,\mathrm{tr}(M_D\Gamma B)$,

$$\mathrm{E}(\mathrm{tr}(M_D y_{-1}y'_{-1})) = \frac{\sigma^2_\varepsilon}{1-\rho^2}(\mathrm{tr}(M_D) + 2\,\mathrm{tr}(M_D\Gamma B)) = \frac{\sigma^2_\varepsilon}{1-\rho^2}\,\mathrm{tr}(M_D\Psi)\,.$$

Table 1.3 *Trace calculations in (1.30) for different models*

Model	$\mathrm{tr}(M_D)$	$\mathrm{tr}(M_D\Psi)$
(1.2)	$(N_1N_2-1)T-(N_1+N_2-2)$	$(N_1N_2-1)T-(N_1+N_2-2)\Theta$
(1.3)	$N_1N_2(T-1)$	$N_1N_2(T-\Theta)$
(1.4)	$(N_1N_2-1)(T-1)$	$(N_1N_2-1)(T-\Theta)$
(1.5)	$N_2T(N_1-1)$	$N_2T(N_1-1)$
(1.6)	$(N_1-1)(N_2-1)T$	$(N_1-1)(N_2-1)T$
(1.7)	$(N_1-1)(N_2-1)(T-1)$	$(N_1-1)(N_2-1)(T-\Theta)$

Table 1.4 *Semi-asymptotic bias of each model formulation*

Model	$\mathrm{plim}_{N_1,N_2\to\infty}(\hat{\rho}-\rho)$
(1.2)	$\frac{1-\rho^2}{2\rho}\left(1-\mathrm{plim}_{N_1,N_2\to\infty}\frac{(N_1N_2-1)T-(N_1+N_2-2)}{(N_1N_2-1)T-(N_1+N_2-2)\Theta}\right)=0$
(1.3), (1.4), (1.7)	$\frac{1-\rho^2}{2\rho}\left(1-\frac{T-1}{T-\Theta}\right)$
(1.5), (1.6)	0

Combining the expressions for the numerator and denominator we get

$$\mathrm{plim}_{N_1,N_2\to\infty}\hat{\rho}=\rho+\frac{1-\rho^2}{2\rho}\left(1-\mathrm{plim}_{N_1,N_2\to\infty}\frac{\mathrm{tr}(M_D)}{\mathrm{tr}(M_D\Psi)}\right). \tag{1.31}$$

As for the specific models, as $\mathrm{tr}(I_T\Psi_0)=\mathrm{tr}(I_T)$, we get $\mathrm{tr}(M_D)$ and $\mathrm{tr}(M_D\Psi)$ as in Table 1.3, with

$$\Theta=\mathrm{tr}(\bar{J}_T\Psi_0)=1+2\frac{\rho}{1-\rho}\left(1-\frac{1}{T}\frac{1-\rho^T}{1-\rho}\right).$$

The individual asymptotic biases, following (1.31), are collected in Table 1.4.[7]

1.6.2 Arellano–Bond Estimation

As seen above, we have problems with the N inconsistency of models (1.3), (1.4) and (1.7) in the dynamic case. Luckily, many of the well known instrumental variables (IV) estimators developed to deal with dynamic panel data models can be generalized to these higher dimensions as well, as the number of available orthogonality conditions increases together with the dimensions. Let us take the example

[7] A natural extension would be to compute these semi-asymptotic biases, when only $N_1\to\infty$ or $N_2\to\infty$. These two cases, however, give the same biases for all models as Table 1.4 suggests, except for model (1.2). In its case, the bias is not completely wiped out when only one individual dimension grows.

of one of the most frequently used, the Arellano and Bond IV estimator (see Arellano and Bond (1991) and (Harris et al., 2008, p. 260)) for the estimation of model (1.3).

The model is written up in first differences, such as

$$(y_{ijt} - y_{ijt-1}) = \rho\,(y_{ijt-1} - y_{ijt-2}) + (\varepsilon_{ijt} - \varepsilon_{ijt-1})\,, \quad t = 3,\ldots,T$$

or

$$\Delta y_{ijt} = \rho \Delta y_{ijt-1} + \Delta\varepsilon_{ijt}, \quad t = 3,\ldots,T.$$

The y_{ijt-k}, $(k = 2,\ldots,t-1)$ are valid instruments for Δy_{ijt-1}, as Δy_{ijt-1} is N asymptotically correlated with y_{ijt-k}, but y_{ijt-k} are not with $\Delta\varepsilon_{ijt}$. As a result, the full instrument set for a given cross sectional pair, (ij) is

$$z_{ij} = \begin{pmatrix} y_{ij1} & 0 & \cdots & \cdots & 0 & \cdots & 0 \\ 0 & y_{ij1} & y_{ij2} & 0 & 0 & \cdots & 0 \\ \vdots & \cdots & & \vdots & \vdots & \cdots & \vdots \\ 0 & \cdots & 0 & 0 & y_{ij1} & \cdots & y_{ijT-2} \end{pmatrix}$$

of size $((T-2)\times(T-1)(T-2)/2)$. The resulting IV estimator of ρ is

$$\hat{\rho}_{AB} = \left[\Delta y'_{-1} Z_{AB}\left(Z'_{AB}\Omega Z_{AB}\right)^{-1} Z'_{AB}\Delta y_{-1}\right]^{-1} \Delta y'_{-1} Z_{AB}\left(Z'_{AB}\Omega Z_{AB}\right)^{-1} Z'_{AB}\Delta y,$$

where Δy and Δy_{-1} are the panel first differences, $Z_{AB} = (z'_{11}, z'_{12}, \ldots, z'_{NN})'$ and $\Omega = (I_{N_1N_2} \otimes \Sigma)$ is the covariance matrix, with known form

$$\Sigma = \begin{pmatrix} 2 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & -1 & 2 & -1 \\ 0 & \cdots & 0 & -1 & 2 \end{pmatrix} \quad \text{of size } \left((T-2)\times(T-2)\right).$$

The generalized Arellano-Bond estimator behaves exactly in the same way as the ‘original’ two dimensional one, regardless the dimensionality of the model.

In the case of models (1.4) and (1.7), to derive an Arellano-Bond type estimator, we need to insert one further step. After taking the first differences, we implement a simple transformation in order to get to a model with only (ij) pairwise interaction effects, exactly as in model (1.3). We then proceed as above, as the Z_{AB} instruments are valid for these transformed models as well. Let us start with model (1.4) and take the first differences to get

$$\Delta y_{ijt} = \rho \Delta y_{ijt-1} + \Delta\lambda_t + \Delta\varepsilon_{ijt}\,.$$

Now, instead of estimating this equation directly with IV, we carry out the following cross-sectional transformation

$$\Delta\tilde{y}_{ijt} = \left(\Delta y_{ijt} - \frac{1}{N_1}\sum_i \Delta y_{ijt}\right) ,$$

or introducing the notation $\Delta\bar{y}_{.jt} = \frac{1}{N_1}\sum_i \Delta y_{ijt}$ and, also, noticing that the λ-s had been eliminated from the model

$$(\Delta y_{ijt} - \Delta\bar{y}_{.jt}) = \rho\,(\Delta y_{ijt-1} - \Delta\bar{y}_{.jt-1}) + (\Delta\varepsilon_{ijt} - \Delta\bar{\varepsilon}_{.jt})\ .$$

The Z_{AB} instruments proposed above are valid with $(\Delta y_{ijt-1} - \Delta\bar{y}_{.jt-1})$ now, as they are uncorrelated with $(\Delta\varepsilon_{ijt} - \Delta\bar{\varepsilon}_{.jt})$, but correlated with the former. The IV estimator of ρ, $\hat{\rho}_{AB}$ has again the form

$$\hat{\rho}_{AB} = \left[\Delta\tilde{y}'_{-1} Z_{AB}(Z'_{AB}\Omega Z_{AB})^{-1} Z'_{AB}\Delta\tilde{y}_{-1}\right]^{-1}\Delta\tilde{y}'_{-1} Z_{AB}(Z'_{AB}\Omega Z_{AB})^{-1} Z'_{AB}\Delta\tilde{y},$$

with $\Delta\tilde{y}$ and $\Delta\tilde{y}_{-1}$ being the transformed panel first differences of the dependent variable.

Continuing now with model (1.7), the transformation needed in this case is

$$\left(\Delta y_{ijt} - \frac{1}{N_1}\sum_i \Delta y_{ijt} - \frac{1}{N_2}\sum_j \Delta y_{ijt} + \frac{1}{N_1 N_2}\sum_{i,j}\Delta y_{ijt}\right) .$$

Picking up the previously introduced notation and using the fact that the fixed effects are cleared again, we get

$$\begin{aligned}(\Delta y_{ijt} - \Delta\bar{y}_{.jt} - \Delta\bar{y}_{i.t} + \Delta\bar{y}_{..t}) = \quad & \rho(\Delta y_{ijt-1} - \Delta\bar{y}_{.jt-1} - \Delta\bar{y}_{i.t-1} + \Delta\bar{y}_{..t-1}) \\ & +(\Delta\varepsilon_{ijt} - \Delta\bar{\varepsilon}_{.jt} - \Delta\bar{\varepsilon}_{i.t} + \Delta\bar{\varepsilon}_{..t})\end{aligned}$$

The Z_{AB} instruments can be used again, on this transformed model, to get a consistent estimator for ρ.

1.7 Heteroscedasticity and Cross-correlation

We have assumed so far throughout the chapter that the idiosyncratic disturbance terms in ε are in fact well-behaved white noises, that is, all heterogeneity is introduced into the model through the fixed effects. Conditioning on the individual dummy variables is however not always enough to address the dependence between individual units. In the presence of such remaining dependences the white noise assumption of the disturbances results biased estimators and spurious inferences. To handle this, we introduce a simple form of cross-correlation and heteroscedasticity among the disturbance terms and see how this influences the estimation methods introduced earlier. So far the approach has been to perform directly LSDV on the models, or alternatively, to transform the models in such a way that the fixed effects drop out, and then estimate the transformed models with OLS. Now, however, in order to use all available information in an optimal way, the structure of the disturbances has to be taken into account for the estimation, promoting Feasible GLS (FGLS) instead of OLS on the fixed effects model. From the joint FGLS estimator of the parameters we can express $\hat{\beta}$ by partialling out the fixed effects parameters as a second step.

1.7.1 The New Covariance Matrices and the GLS Estimator

The initial assumptions about the disturbance terms are now replaced by

$$\mathrm{E}(\varepsilon_{ijt}\varepsilon_{kls}) = \begin{cases} \sigma_{ij}^2 & \text{if } i=k, j=l, t=s \\ \rho_1 & \text{if } i=k, j\neq l, \forall t,s \\ \rho_2 & \text{if } i\neq k, j=l, \forall t,s \\ 0 & \text{otherwise} \end{cases}$$

which allows for a general form of cross-dependence and heteroscedasticity. Then the variance-covariance matrix of all models introduced in Section 1.3 takes the form

$$\mathrm{E}(\varepsilon\varepsilon') = \Omega = (\Upsilon \otimes I_T) + \rho_1(I_{N_1} \otimes J_{N_2T}) + \rho_2(J_{N_1} \otimes I_{N_2} \otimes J_T), \tag{1.32}$$

where

$$\Upsilon = \begin{pmatrix} \sigma_{11}^2 - \rho_1 - \rho_2 & 0 & \cdots & 0 \\ 0 & \sigma_{12}^2 - \rho_1 - \rho_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{N_1N_2}^2 - \rho_1 - \rho_2 \end{pmatrix}$$

is an $(N_1N_2 \times N_1N_2)$ diagonal matrix. Invoking the form of the general fixed effects model (1.1), and decomposing X and D to Z and β and π to δ, gives

$$y = Z\delta + \varepsilon\,.$$

The GLS estimator is then reads as

$$\hat{\delta} = \left(Z'\Omega^{-1}Z\right)^{-1} Z'\Omega^{-1}y\,. \tag{1.33}$$

As much as (1.33) is simple theoretically, it is as forbidding in practice: to carry the estimation out, we have to compute Ω^{-1} first, to get $\hat{\delta}$, then $(D'\Omega^{-1}D)^{-1}$, to express $\hat{\beta}$ from the joint estimator. With a decomposition of Ω (exact derivations are omitted), the largest matrix to work with is of order $\min\{N_1,N_2\}$ when computing Ω^{-1}, however there is no clear way to reduce the computation of $(D'\Omega^{-1}D)^{-1}$.

The situation is fundamentally different if, along with cross-correlations, homoscedasticity is assumed. In that case, Ω is simplified to

$$\Omega = (\sigma_\varepsilon^2 - \rho_1 - \rho_2) I_{N_1 N_2 T} + \rho_1 (I_{N_1} \otimes J_{N_2 T}) + \rho_2 (J_{N_1} \otimes I_{N_2} \otimes J_T) \ ,$$

with only three variance components, and its inverse is easily obtained with a decomposition similar to Wansbeek and Kapteyn (1982),

$$\Omega^{-1} = I_{N_1 N_2 T} + \theta_1 (I_{N_1} \otimes \bar{J}_{N_2 T}) + \theta_2 (\bar{J}_{N_1} \otimes I_{N_2} \bar{J}_T) + \theta_3 (I_{N_1} \otimes \bar{J}_{N_2 T})$$

with

$$\begin{aligned} \theta_1 &= -\tfrac{N_2 T \rho_1}{(N_2 T - 1)\rho_1 - \rho_2 + \sigma_\varepsilon^2} \ , \quad \theta_2 = -\tfrac{N_1 T \rho_2}{(N_1 T - 1)\rho_2 - \rho_1 + \sigma_\varepsilon^2} \quad \text{and} \\ \theta_3 &= \left(\tfrac{N_2 T \rho_1}{(N_2 T - 1)\rho_1 - \rho_2 + \sigma_\varepsilon^2} + \tfrac{N_1 T \rho_2}{(N_1 T - 1)\rho_2 - \rho_1 + \sigma_\varepsilon^2} - \tfrac{N_1 T \rho_2 + N_2 T \rho_1}{(N_1 T - 1)\rho_2 + (N_2 T - 1)\rho_1 + \sigma_\varepsilon^2} \right) \ . \end{aligned}$$

As now we have the exact form of Ω^{-1}, estimation (1.33) can be performed, and the (BLUE) $\hat{\delta}$ GLS estimators collected. Note, that this GLS estimation is equivalent to a two-step procedure, where we first transform y, X and D according to

$$\begin{aligned} \tilde{y}_{ijt} &= y_{ijt} - \left(1 - \sqrt{\theta_1 + 1}\right) \bar{y}_{i..} - \left(1 - \sqrt{\theta_2 + 1}\right) \bar{y}_{.j.} \\ &\quad + \left(1 - \sqrt{\theta_1 + 1} - \sqrt{\theta_2 + 1} + \sqrt{\theta_1 + \theta_2 + \theta_3 + 1}\right) \bar{y}_{...} \ , \end{aligned}$$

which is proportional to the scalar representation of $\Omega^{-\frac{1}{2}} y$, then perform an OLS on the transformed model. To express $\hat{\beta}$ out from the composite estimator $\hat{\delta}$, invoking the Frisch–Waugh–Lovell theorem, the transformed variables should be further pre-multiplied with the projector

$$M_{\Omega^{-1/2} D} = I - \Omega^{-1/2} D \left(D' \Omega^{-1} D\right)^{-} D' \Omega^{-1/2} \ ,$$

and an OLS should be performed on the twice-transformed variables y and X. As it turns out, the two consecutive transformations, $\Omega^{-\frac{1}{2}}$ and $M_{\Omega^{-\frac{1}{2}} D}$, together are identical to the Within transformation for all models except for (1.5), with α_{jt} fixed effects. In other words, the GLS equals the OLS so long the effects are symmetrical in i and j, as, quite intuitively, the Within transformation for those models eliminate the cross-correlations from the disturbance terms along with the fixed effects.

1.7.2 Estimation of the Variance Components and the Cross Correlations

What now remains to be done is to estimate the variance components in order to make the GLS feasible. In principle, the job is to find a set of identifying equations from which the variance components can be expressed. Remember, that during the estimation we have transformed the models and performed an OLS on that. However this, in the case of some models, highly limits the number of identifying equations available for the variance components. For some models, this even means that the variance components are non-estimable without further restrictions on the structure of the disturbances (for example, $\rho_1 = \rho_2$, or an even stronger one, $\rho_1 = \rho_2 = 0$). This would certainly impede our cause, so let us take another track. Along with the OLS residuals from the transformed models, we can produce an other type of residual: the one from the LSDV estimation. As we will see, we can estimate all the variance components from the LSDV residuals, and at the same time, we can obtain these residuals without directly estimating the possibly numerous fixed effects.

As Section 1.3 suggests, whenever the D dummy coefficient matrix has no full column rank, the composite fixed effects parameters, π can not be identified (and of course, estimated). However, this is not the case for $D\pi$, which is given by

$$D\hat{\pi} = D(D'D)^{-}D'(y - X\hat{\beta}) = (I - M_D)(y - X\hat{\beta})\ . \tag{1.34}$$

The LSDV residuals are

$$\hat{\varepsilon} = y - X\hat{\beta} - D\hat{\pi} = (I - (I - M_D))\,(y - X\hat{\beta}) = M_D(y - X\hat{\beta}) = \tilde{y} - \tilde{X}\hat{\beta} \tag{1.35}$$

where '$\sim$' denotes the appropriate Within transformation.

With the residuals in hand, the variance components can be expressed from the same identifying conditions regardless of the model specification. These are

$$\begin{aligned} \mathrm{E}\left(\varepsilon_{ijt}^2\right) &= \sigma_{ij}^2 \\ \mathrm{E}\left(\bar{\varepsilon}_{.jt}^2\right) &= \tfrac{1}{N_1^2}\left(\textstyle\sum_i \sigma_{ij}^2 + N_1(N_1-1)\rho_2\right) \\ \mathrm{E}\left(\bar{\varepsilon}_{i.t}^2\right) &= \tfrac{1}{N_2^2}\left(\textstyle\sum_j \sigma_{ij}^2 + N_2(N_2-1)\rho_1\right)\ . \end{aligned}$$

The last step is to "estimate" the identifying conditions by replacing expectations with sample means, and the disturbances with the residuals. That is,

$$\begin{aligned} \hat{\sigma}_{ij}^2 &= \tfrac{1}{T}\textstyle\sum_t \hat{\varepsilon}_{ijt}^2 \\ \hat{\rho}_2 &= \tfrac{1}{N_1(N_1-1)}\left(\tfrac{1}{N_2 T}\textstyle\sum_{jt}\left(\sum_i \hat{\varepsilon}_{ijt}\right)^2 - \sum_i \hat{\sigma}_{ij}^2\right) \\ \hat{\rho}_1 &= \tfrac{1}{N_2(N_2-1)}\left(\tfrac{1}{N_1 T}\textstyle\sum_{it}\left(\sum_j \hat{\varepsilon}_{ijt}\right)^2 - \sum_j \hat{\sigma}_{ij}^2\right)\ . \end{aligned} \tag{1.36}$$

Equation (1.36) gives consistent estimators of the variance components, as long as $T \to \infty$, as the number of heteroscedastic variances grows along with N_1 and N_2. Inserting these estimated variance components into (1.33) gives the FGLS estimator which handles the new and more flexible correlation structure.

When homoscedasticity is assumed along with the cross-correlations, the variance-components estimators become

$$
\begin{aligned}
\hat{\sigma}_\varepsilon^2 &= \tfrac{1}{N_1N_2T}\textstyle\sum_{ijt}\hat{\varepsilon}_{ijt}^2 \\
\hat{\rho}_2 &= \tfrac{1}{N_1-1}\left(\tfrac{1}{N_1N_2T}\textstyle\sum_{jt}\left(\sum_i\hat{\varepsilon}_{ijt}\right)^2-\hat{\sigma}_\varepsilon^2\right) \\
\hat{\rho}_1 &= \tfrac{1}{N_2-1}\left(\tfrac{1}{N_1N_2T}\textstyle\sum_{it}\left(\sum_j\hat{\varepsilon}_{ijt}\right)^2-\hat{\sigma}_\varepsilon^2\right)\,,
\end{aligned}
\tag{1.37}
$$

and T-asymptotics is not necessary any more ($N_1\to\infty$ or $N_2\to\infty$ is enough) to get the estimators consistent.

When the data is incomplete, the derived FGLS estimator for the model with homoscedasticity and cross-correlations is not appropriate as the decomposition of Ω can not be represented with Kronecker products any longer, and so the presented linear transformations to be employed on the data are incorrect. As the full analysis of such incomplete estimator would certainly be lengthy, we only provide some guidance on how to carry out the estimation. First, we leave out those rows from D (similarly as we did in Section 1.5.2), and rows and columns from Ω which correspond to missing observations. Then we proceed by performing a GLS with the adjusted covariance matrix, but to get its inverse, we now have to use partial inverse methods, to at least partially avoid the dimensionality issue. The last step is to estimate the variance components, for which we only have to adjust (1.36) (or (1.37)) to the incomplete sample sizes.

Remember, that the FGLS estimator in the presence of heteroscedasticity is consistent only for long panels (when $T\to\infty$). So how to proceed when the data is small in the time dimension? Let us consider the case of heteroscedasticity only, so set the cross correlations to null ($\rho_1=\rho_2=0$). This special case can be estimated in two ways. The optimal way is to first to transform the model according to the optimal Within transformation as before, then carry out an FGLS with the heteroscedastic covariance matrix

$$
\Omega_h=\operatorname{diag}\left\{\sigma_{11}^2 I_{|T_{11}|},\,\sigma_{12}^2 I_{|T_{12}|},\,\ldots,\,\sigma_{nm}^2 I_{|T_{N_1N_2}|}\right\}\,,
$$

which is diagonal regardless of the potential data issues. The variance components are then estimated from

$$
\hat{\sigma}_{ij}^2=\frac{1}{|T_{ij}|}\sum_t\hat{\varepsilon}_{ijt}^2,
$$

like before, with the $\hat{\varepsilon}_{ijt}$ being the LSDV residuals. However, this FGLS, as before, is still only T consistent. When the data is short in time, it is better to estimate the transformed model with OLS, which is still an unbiased and consistent estimator of β in all the asymptotic cases studied before (though not optimal any more), and use heteroscedasticity robust White covariance matrix to estimate $\mathrm{Var}(\beta)$. Then we get

$$\begin{aligned}\text{Var}(\hat{\beta}) &= (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\hat{\Omega}_h\tilde{X}(\tilde{X}'\tilde{X})^{-1} \\ &= \Big(\textstyle\sum_{ijt}\tilde{x}_{ijt}\tilde{x}'_{ijt}\Big)^{-1}\Big(\textstyle\sum_{ijt}\tilde{x}_{ijt}\tilde{x}'_{ijt}\frac{1}{|T_{ij}|}\sum_t\hat{\varepsilon}^2_{ijt}\Big)\Big(\textstyle\sum_{ijt}\tilde{x}_{ijt}\tilde{x}'_{ijt}\Big)^{-1},\end{aligned}$$

where '$\sim$' indicates that the variables are transformed. Notice again, that only the data X has to be transformed, but not Ω_h, due to the idempotent nature of the projection matrix.

1.8 Extensions to Higher Dimensions

In four and higher dimensions the number of specific effects, and therefore models, available is staggering. As a consequence, we have to restrict somehow the model formulations taken into account. The restriction used in this chapter is to allow for pairwise interaction effects only. Without attempting to be comprehensive, the most relevant four dimensional models are introduced in this section. Then, on a kind of benchmark model, we show intuitively how to estimate them for complete data, and also, in the case of the same data problems brought up in Sections 1.4 and 1.5. This is carried in a way that gives indications on how to proceed beyond four dimensions.

1.8.1 Different Forms of Heterogeneity

The dependent variable is now observed along four indices, such as $ijst$. The generalization of model (1.4) (and also, that of the 2D fixed effects model with both individual and time effects) is

$$y_{ijst} = x'_{ijst}\beta + \gamma_{ijs} + \lambda_t + \varepsilon_{ijst}\,,$$

or alternatively, a more restrictive formulation is

$$y_{ijst} = x'_{ijst}\beta + \alpha_i + \alpha^*_j + \gamma_s + \lambda_t + \varepsilon_{ijst}\,.$$

As in the case of 3D models, we can benefit from the multi-dimensional nature of the data, and let the fixed effects to be time dependent

$$y_{ijst} = x'_{ijst}\beta + \alpha_{it} + \gamma_{jt} + \delta_{st} + \varepsilon_{ijst}$$

that is we can allow all individual heterogeneity to vary over time as well. Finally, let us take the four-dimensional extension of the all-encompassing model (1.7), with pair-wise interaction effects:

$$y_{ijst} = x'_{ijst}\beta + \gamma^0_{ijs} + \gamma^1_{ijt} + \gamma^2_{jst} + \gamma^3_{ist} + \varepsilon_{ijst}\,, \tag{1.38}$$

with $i = 1 \dots N_1$, $j = 1 \dots N_2$, $s = 1 \dots N_3$, and $t = 1 \dots T$. This is what we consider from now on as a benchmark model, and show step-by-step how to estimate it.

1.8.2 Least Squares and the Within Estimators

If we keep maintaining the standard OLS assumptions lined up in Section 1.2, the LSDV estimator of model (1.38), following (1.8), is BLUE. Also, if we define the Within projector M_D, to get $\hat{\beta}$, the maximum matrix size to be worked with is still $(K \times K)$. For model (1.38), the composite dummy matrix D is

$$D = ((I_{N_1N_2N_3} \otimes \iota_T),\ (I_{N_1N_2} \otimes \iota_{N_3} \otimes I_T),\ (\iota_{N_1} \otimes I_{N_2N_3T}),\ (I_{N_1} \otimes \iota_{N_2} \otimes I_{N_3T}))$$

with size

$$(N_1N_2N_3T \times (N_1N_2N_3 + N_1N_2T + N_2N_3T + N_1N_3T))\,,$$

and column rank

$$(N_1N_2N_3T - (N_1-1)(N_2-1)(N_3-1)(T-1))\,,$$

leading to

$$\begin{aligned} M_D \;=\;& I_{N_1N_2N_3T} - (\bar{J}_{N_1} \otimes I_{N_2N_3T}) - (I_{N_1} \otimes \bar{J}_{N_2} \otimes I_{N_3T}) \\ & - (I_{N_1N_2} \otimes \bar{J}_{N_3} \otimes I_T) - (I_{N_1N_2N_3} \otimes \bar{J}_T) + (\bar{J}_{N_1N_2} \otimes I_{N_3T}) \\ & + (\bar{J}_{N_1} \otimes I_{N_2} \otimes \bar{J}_{N_3} \otimes I_T) + (\bar{J}_{N_1} \otimes I_{N_2N_3} \otimes \bar{J}_T) \\ & + (I_{N_1} \otimes \bar{J}_{N_2N_3} \otimes I_T) + (I_{N_1} \otimes \bar{J}_{N_2} \otimes I_{N_3} \otimes \bar{J}_T) + (I_{N_1N_2} \otimes \bar{J}_{N_3T}) \\ & - (\bar{J}_{N_1N_2N_3} \otimes I_T) - (\bar{J}_{N_1N_2} \otimes I_{N_3} \otimes \bar{J}_T) - (\bar{J}_{N_1} \otimes I_{N_2} \otimes \bar{J}_{N_3T}) \\ & - (I_{N_1} \otimes \bar{J}_{N_2N_3T}) + \bar{J}_{N_1N_2N_3T}\,. \end{aligned}$$

Just as before, M_D defines the optimal Within transformation to be performed on the data, so we can avoid matrix manipulations. That is, the LSDV estimator of β is analogous to the optimal Within estimator, which is obtained by first transforming the data according to

$$\begin{aligned} \tilde{y}_{ijst} \;=\;& y_{ijst} - \bar{y}_{.jst} - \bar{y}_{i.st} - \bar{y}_{ij.t} - \bar{y}_{ijs.} + \bar{y}_{..st} + \bar{y}_{.j.t} + \bar{y}_{.js.} \\ & + \bar{y}_{i..t} + \bar{y}_{i.s.} + \bar{y}_{ij..} - \bar{y}_{...t} - \bar{y}_{..s.} - \bar{y}_{.j..} - \bar{y}_{i...} + \bar{y}_{....} \end{aligned} \tag{1.39}$$

(which eliminates $(\gamma^0_{ijs},\ \gamma^1_{ijt},\ \gamma^2_{jst},\ \gamma^3_{ist})$), then running an OLS on the transformed variables $\tilde{y}_{ijst}, \tilde{x}'_{ijst}$.

The properties of these estimators are identical to those of the three-way models, with the only modification, that now even more asymptotic cases could be considered. In general, the estimator of a fixed effects parameter is consistent, if an index with which the effect is fixed, goes to infinity. The resulting variances of any of the estimators should be normalized with the sample sizes which grow, and further, the degrees of freedom should be corrected to reflect the column rank deficiency in D.[8]

1.8.3 Some Data Problems

In theory, the missing data problem is corrected for by leaving out those rows from D which correspond to missing observations. LSDV estimation should then be done with the modified $\tilde{D}$, or alternatively, with $M_{\tilde{D}} = I - \tilde{D}(\tilde{D}'\tilde{D})^{-}\tilde{D}'$. Unfortunately, as now M_D has no clear structure, the resulting LSDV estimator can not be reached at an acceptable costs when the data is large. However, the optimal Within estimator offers a better way to tackle the problem at only moderate costs. Just like in Section 1.5, we have to come up with adjusted transformations, that clear out the fixed effects in the case of missing data. The no self-flow and unbalanced transformations in Section 1.5 can easily be generalized to any higher dimensions. For model (1.38), the no self-flow robust transformation reads as

$$\begin{aligned}\tilde{y}_{ijst} = \; & y_{ijst} - \tfrac{1}{N-1}y_{+jst} - \tfrac{1}{N-1}y_{i+st} - \tfrac{1}{N_3}y_{ij+t} - \tfrac{1}{T}y_{ijs+} + \tfrac{1}{(N-1)^2}y_{++st} \\ & + \tfrac{1}{(N-1)N_3}y_{+j+t} + \tfrac{1}{(N-1)T}y_{+js+} + \tfrac{1}{(N-1)N_3}y_{i++t} + \tfrac{1}{(N-1)T}y_{i+s+} \\ & + \tfrac{1}{N_3T}y_{ij++} - \tfrac{1}{(N-1)^2N_3}y_{+++t} - \tfrac{1}{(N-1)^2T}y_{++s+} - \tfrac{1}{(N-1)N_3T}y_{+j++} \\ & - \tfrac{1}{(N-1)N_3T}y_{i+++} + \tfrac{1}{(N-1)^2N_3T}y_{++++} - \tfrac{1}{(N-1)N_3T}y_{ji++} \\ & + \tfrac{1}{(N-1)T}y_{jis+} + \tfrac{1}{(N-1)N_3}y_{ji+t} - \tfrac{1}{N-1}y_{jist} \, .\end{aligned} \tag{1.40}$$

Note, that in the no self-flow case $N_1 = N_2 = N$ had to be assumed.

Incomplete data can also be handled quite flexibly in case of four-dimensional models. Remember, that the key (iterative) unbalanced-robust transformation in Section 1.5 was (1.28), which can be generalized simply into a four dimensional setup. Let the dummy variables matrices for the four fixed effects in (1.38) be denoted by $D_e = (D_1^e, D_2^e, D_3^e, D_4^e)$ and let $M_{D_e}^{(k)}$ be the transformation that clears the first k fixed effects out; namely, $M_{D_e}^{(k)} \cdot \left(D_1^e, \ldots, D_k^e\right) = (0, \ldots, 0)$ for $k = 1 \ldots 4$. The appropriate Within transformation to clear out the first k fixed effects is then

$$M_{D_e}^{(k)} = M_{D_e}^{(k-1)} - \left(M_{D_e}^{(k-1)}D_k^e\right)\left[\left(M_{D_e}^{(k-1)}D_k^e\right)'\left(M_{D_e}^{(k-1)}D_k^e\right)\right]^{-}\left(M_{D_e}^{(k-1)}D_k^e\right)', \tag{1.41}$$

where the first step in the iteration is

$$M_{D_e}^{(1)} = I - D_1^e\left((D_1^e)'D_1^e\right)^{-1}(D_1^e)',$$

and the iteration should be processed until $k = 4$. Note that none of this hinges on

[8] For example for model (1.38), the correct degrees of freedom (coming from the rank of M_D) is $(N_1 - 1)(N_2 - 1)(N_3 - 1)(T - 1) - K$.

the model specification and can be done to any other multi-dimensional fixed effects model. The drawback, which can not really be addressed at this point, is again the increasing size of the matrices involved in the calculations. If this is the case, direct inverse calculations are feasible only up to some point, and further tricks (parallel computations, iterative inverting methods) should be used. However, this is beyond the scope of this chapter.

1.9 Conclusion

In the case of three and higher dimensional fixed effects panel data models, due to the many interaction effects, the number of dummy variables in the model increases dramatically. This circumvents the direct estimation of all model parameters. In order to estimate the slope parameters at least, we either have to partial out the intercept parameters from the LSDV estimation, or alternatively (producing numerically equivalent estimates), we can use the appropriate Within estimators which do not require the explicit incorporation of the fixed effects into the model. Although these Within estimators are more complex than for the usual two dimensional panel data models, and usually not unique, they turn out to be quite useful in these higher dimensional setups. Along with the estimators, finite and asymptotic properties are considered, as well as some insights on testing for parameter values. Both estimators (the LSDV and the Within), however, are biased and inconsistent in the case of some relevant data problems, like the lack of self-flows, or general incompleteness in the data. The chapter offers two ways to correct for this inconsistency: one iterative, following Carneiro et al. (2008) and Guimaraes and Portugal (2009), and another, which is derived from the Frisch-Waugh theorem. The chapter also extends the results of 2D dynamic autoregressive models, and generalizes the so-called Nickell-bias to show that the estimators of 3D fixed effects models are biased in general. Interestingly, the Within estimator is inconsistent only for some of the considered three-dimensional models, which inconsistency in turn is easily tackled by Arellano-Bond-type estimators. Next, we have allowed heteroscedasticity, and a simple form of cross-correlation to the disturbance terms, and derived (F)GLS estimators for such augmented models. Lastly, we have shown, through the lenses of a four- and higher-dimensional extension, that generalization of any result of the chapter is straightforward, and thus can be done without trouble. The derived estimators and properties of this chapter should be taken into account by all researchers relying on these methods.

Appendix A – Background Calculations – Obtaining M_D from D

Let us consider model (1.2) in details, the rest is analogous. Remember, that now the dummy matrix is

$$D = ((I_{N_1} \otimes \iota_{N_2T}),\ (\iota_{N_1} \otimes I_{N_2} \otimes \iota_T),\ (\iota_{N_1N_2} \otimes I_T))\ .$$

Wansbeek (1991) has shown, that the column space of D does not change by replacing I_{N_1} (or similarly I_{N_2} or I_T) with any $(G_{N_1}, \bar{\iota}_{N_1})$ orthonormal matrix of order $(N_1 \times (N_1 - 1))$, where G_{N_1} has to satisfy the following conditions:

$$G'_{N_1}\iota_{N_1} = 0, \quad \text{and} \quad G'_{N_1}G_{N_1} = I_{N_1-1}, \quad \text{with} \quad \bar{\iota}_{N_1} \equiv \iota_{N_1}/\sqrt{N_1}\,.$$

As matrix D spans the same vector space as the following orthonormal matrix

$$\hat{D} \equiv ((G_{N_1} \otimes \bar{\iota}_{N_2T}),\ (\bar{\iota}_{N_1} \otimes G_{N_2} \otimes \bar{\iota}_T),\ (\bar{\iota}_{N_1N_2} \otimes G_T),\ \bar{\iota}_{N_1N_2T})\,,$$

which has in fact full column rank $(N_1 + N_2 + T - 2)$, the projection matrix of size $(N_1N_2T \times N_1N_2T)$ to get rid of D is simply

$$\begin{aligned} M_D &\equiv I_{N_1N_2T} - \hat{D}\hat{D}' \\ &= I_{N_1N_2T} - (Q_{N_1} \otimes \bar{J}_{N_2T}) - (\bar{J}_{N_1} \otimes Q_{N_2} \otimes \bar{J}_T) - (\bar{J}_{N_1N_2} \otimes Q_T) - \bar{J}_{N_1N_2T} \\ &= I_{N_1N_2T} - (I_{N_1} \otimes \bar{J}_{N_2T}) - (\bar{J}_{N_1} \otimes I_{N_2} \otimes \bar{J}_T) - (\bar{J}_{N_1N_2} \otimes I_T) + 2\bar{J}_{N_1N_2T}\,, \end{aligned}$$

with $Q_{N_1} \equiv G_{N_1}G'_{N_1} = I_{N_1} - \bar{J}_{N_1}$.

Appendix B – Background Calculations – Derivations of the No Self-flow Transformations

B.1 No Self-flow Derivation for the Pure Cross-sectional Panel Model

To get an insight for the exact derivations, let us consider first the case of $T = 1$. This corresponds to pure cross-sectional panels, where the only reasonable model formulation (ignoring the trivial case of one fixed effect) is, with $i, j = 1, \ldots, N$

$$y_{ij} = \beta' x_{ij} + \alpha_i + \gamma_j + \varepsilon_{ij}\,, \tag{B.42}$$

or in matrix form,

$$\begin{aligned} y &= X\beta + (I_N \otimes \iota_N)\alpha + (\iota_N \otimes I_N)\gamma + \varepsilon = X\beta + D_\alpha\alpha + D_\gamma\gamma + \varepsilon \\ &= X\beta + D(\alpha', \gamma')' + \varepsilon\,. \end{aligned}$$

As there are no data with $i = j$, we eliminate these from the model by using the selection matrix L of order $N^2 \times N(N-1)$ to get

$$L'y = L'X\beta + L'D(\alpha', \gamma')' + L'\varepsilon.$$

So the optimal effects-eliminating projection matrix is

$$M_{L'D} = I_{N(N-1)} - L'DW^{+}D'L,$$

with $W = D'LL'D$ and "$^{+}$" denoting the Moore-Penrose generalized inverse. We want to have a simple expression for the elements of $M_{L'D}L'y$, indicated by a tilde. When in the data $i = j$ are observed, this expression is

$$\tilde{y}_{ij} = y_{ij} - y_{i.} - y_{.j} + y_{..} \, .$$

Now, it is more complicated. For $i \neq j$, $(e_i \otimes e_j)'D = (e_i \otimes e_j)'LL'D = (e_i', e_j')$, so

$$\tilde{y}_{ij} = (e_i \otimes e_j)'LM_D L'y = y_{ij} - (e_i', e_j')W^{+}D'LL'y \tag{B.43}$$

with e_i being the ith unit vector of size N. We have to further elaborate on W. Since

$$D_\alpha' LL' D_\alpha = D_\gamma' LL' D_\gamma = (N-1)I_N \quad \text{and} \quad D_\alpha' LL' D_\gamma = J_N - I_N$$

and, as before, $\bar{J}_N = J_N/N$ and $Q_N = I_N - \bar{J}_N$, we obtain

$$\begin{aligned} W &= \begin{pmatrix} N-1 & -1 \\ -1 & N-1 \end{pmatrix} \otimes I_N + \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \otimes J_N \\ &= \begin{pmatrix} N-1 & -1 \\ -1 & N-1 \end{pmatrix} \otimes Q_N + (N-1) \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \otimes \bar{J}_N \, . \end{aligned}$$

Since Q_N and $\bar{J}_N$ are idempotent and mutually orthogonal, the Moore-Penrose inverse W^{+} of W is

$$\begin{aligned} W^{+} &= \tfrac{1}{N(N-2)} \begin{pmatrix} N-1 & 1 \\ 1 & N-1 \end{pmatrix} \otimes Q_N + \tfrac{1}{4(N-1)} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \otimes \bar{J}_N \\ &= \tfrac{1}{N(N-2)} \begin{pmatrix} N-1 & 1 \\ 1 & N-1 \end{pmatrix} \otimes I_N + \tfrac{1}{N} \begin{pmatrix} p & q \\ q & p \end{pmatrix} \otimes J_N \, , \end{aligned}$$

with

$$p = \frac{1}{4(N-1)} - \frac{N-1}{N(N-2)} \quad \text{and} \quad q = \frac{1}{4(N-1)} - \frac{1}{N(N-2)} \, .$$

Now, with this updated form of W,

$$\begin{aligned} (e_i', e_j')W^{+} &= \tfrac{1}{N(N-2)} \left((N-1)e_i' + e_j', e_i' + (N-1)e_j' \right) \\ &\quad - \tfrac{1}{2(N-1)(N-2)} \left(\iota_N', \iota_N' \right) . \end{aligned}$$

Moreover, with Y being the $(N \times N)$ data matrix containing the y_{ij} observations, with zeros filled in the empty diagonal elements,

$$D'LL'y = \begin{pmatrix} Y'\iota_N \\ Y\iota_N \end{pmatrix} .$$

So, after multiplying $(e_i', e_i')W^+$ and $D'LL'y$, and as $y_{++} = \iota_N' Y \iota_N = \iota_N' Y' \iota_N$, we get

$$\tilde{y}_{ij} = y_{ij} - \frac{N-1}{N(N-2)}\left(y_{i+} + y_{+j}\right) - \frac{1}{N(N-2)}\left(y_{+i} + y_{j+}\right) + \frac{1}{(N-1)(N-2)} y_{++}\,, \tag{B.44}$$

with

$$y_{+i} = \sum_i y_{ij,(j=i)}, \quad y_{j+} = \sum_j y_{ij,(i=j)}\,.$$

When N grows larger, the effect of the missing diagonal elements becomes smaller, which is reflected in the above expression by the third term at the right-hand side of formula (B.44) being of lower order than N. The introduction of the "+" notation is needed to avoid confusions with the indexing. For example, now we take sums with respect to the first index, fixing the second j, but we also take the same kind of sum, only fixing i now as the second index. This cannot be properly represented with our previous $\bar{y}_{\cdot j}$-type notations. Let us illustrate these new sums with a small example. Let $N = 3$ and let $y_{ij} = y_{12}$. Then $y_{+i} = y_{+1} = y_{11} + y_{21} + y_{31}$, and similarly, $y_{j+} = y_{2+} = y_{21} + y_{22} + y_{23}$.

B.2 No Self-flow Transformation for Model (1.2)

We start from the matrix form of (1.2),

$$\begin{aligned} y &= X\beta + (I_N \otimes \iota_N \otimes \iota_T)\alpha + (\iota_N \otimes I_N \otimes \iota_T)\gamma + (\iota_N \otimes \iota_N \otimes I_T)\lambda + \varepsilon \\ &= X\beta + D_{\alpha_*}\alpha + D_{\gamma_*}\gamma + D_\lambda \lambda + \varepsilon\,. \end{aligned}$$

To adjust the model for the no self-flow type data, we pre-multiply it with the selection matrix $\tilde{L}$ of order $(N^2T \times (N(N-1)T))$, to get

$$\begin{aligned} \tilde{L}'y &= \tilde{L}'X\beta + \tilde{L}'D_{\alpha_*}\alpha + \tilde{L}'D_{\gamma_*}\gamma + \tilde{L}'D_\lambda \lambda + \tilde{L}'\varepsilon \\ &= \tilde{L}'X\beta + \tilde{L}'D(\alpha', \gamma', \lambda')' + \tilde{L}'\varepsilon\,. \end{aligned}$$

Note, that $\tilde{L}\tilde{L}' = I_{N^2T} - \sum_i (e_i e_i' \otimes e_i e_i') \otimes I_T$. The optimal effects-eliminating projection matrix, orthogonal to $\tilde{L}'D$, is a direct application of the Frisch-Waugh theorem,

$$M_{\tilde{L}'D} = I_{N(N-1)T} - \tilde{L}'DW_1^+D'\tilde{L}$$

with $W_1 = D'\tilde{L}\tilde{L}'D$. From now on, we have three things to do. First, get the $\tilde{L}'D$ for a particular (ijt) element. Second, elaborate on W_1^+ (this is the hardest part), and third, get $D'\tilde{L}\tilde{L}'y$. Notice, that $(e_i \otimes e_j \otimes e_t)\tilde{L}\tilde{L}'D = (e_i', e_j', e_t')$, so

$$\tilde{y}_{ijt} = y_{ijt} - (e_i', e_j', e_t')W_1^+D'\tilde{L}\tilde{L}'y\,.$$

Now

$$
\begin{aligned}
D'_{\alpha_*}\tilde{L}\tilde{L}'D_{\alpha_*} &= (I_N \otimes \iota_N \otimes \iota_T)'\{I_{N^2T} - \textstyle\sum_i (e_ie_i' \otimes e_ie_i') \otimes I_T\}(I_N \otimes \iota_N \otimes \iota_T)\\
&= (I_N \otimes \iota_N'\iota_N \otimes \iota_T'\iota_T) - (\iota_N'e_i)^2\iota_T'\iota_T \textstyle\sum_i e_ie_i'\\
&= (N-1)TI_N
\end{aligned}
$$

$$
\begin{aligned}
D'_{\gamma_*}\tilde{L}\tilde{L}'D_{\gamma_*} &= (N-1)TI_N\\
D'_{\lambda}\tilde{L}\tilde{L}'D_{\lambda} &= (\iota_N \otimes \iota_N \otimes I_T)'\{I_{N^2T} - \textstyle\sum_i (e_ie_i' \otimes e_ie_i') \otimes I_T\}(\iota_N \otimes \iota_N \otimes I_T)\\
&= (\iota_N'\iota_N \otimes \iota_N'\iota_N \otimes I_T) - \textstyle\sum_i (\iota_N'e_ie_i'\iota_N)^2 \otimes I_T\\
&= N(N-1)I_T
\end{aligned}
$$

$$
\begin{aligned}
D'_{\alpha_*}\tilde{L}\tilde{L}'D_{\gamma_*} &= (I_N \otimes \iota_N \otimes \iota_T)'\{I_{N^2T} - \textstyle\sum_i (e_ie_i' \otimes e_ie_i') \otimes I_T\}(\iota_N \otimes I_N \otimes \iota_T)\\
&= (J_N \otimes \iota_T'\iota_T) - (\iota_N'e_i)^2\iota_T'\iota_T \textstyle\sum_i e_ie_i'\\
&= T(J_N - I_N)
\end{aligned}
$$

$$
\begin{aligned}
D'_{\gamma_*}\tilde{L}\tilde{L}'D_{\alpha_*} &= T(J_N - I_N)\\
D'_{\alpha_*}\tilde{L}\tilde{L}'D_{\lambda} &= (I_N \otimes \iota_N \otimes \iota_T)'\{I_{N^2T} - \textstyle\sum_i (e_ie_i' \otimes e_ie_i') \otimes I_T\}(\iota_N \otimes \iota_N \otimes I_T)\\
&= (\iota_N \otimes \iota_N'\iota_N \otimes \iota_T') - (\iota_N'e_i)^2 \textstyle\sum_i e_i \otimes \iota_T'\\
&= (N-1)\iota_N\iota_T'\\
D'_{\gamma_*}\tilde{L}\tilde{L}'D_{\lambda} &= (N-1)\iota_N\iota_T'\\
D'_{\lambda}\tilde{L}\tilde{L}'D_{\alpha_*} &= (N-1)\iota_T\iota_N' \quad \text{and}\\
D'_{\lambda}\tilde{L}\tilde{L}'D_{\gamma_*} &= (N-1)\iota_T\iota_N',
\end{aligned}
$$

From this, W_1 can be constructed as

$$
W_1 = \left(\begin{array}{cc|c}
(N-1)TI_N & T(J_N - I_N) & (N-1)\iota_N\iota_T'\\
T(J_N - I_N) & (N-1)TI_N & (N-1)\iota_N\iota_T'\\
\hline
(N-1)\iota_T\iota_N' & (N-1)\iota_T\iota_N' & N(N-1)I_T\,,
\end{array}\right)
$$

so its inverse is simply

$$
W_1^+ = \left(\begin{array}{cc}
\frac{1}{TN(N-2)}\begin{pmatrix} N-1 & 1\\ 1 & N-1\end{pmatrix} \otimes I_N + \begin{pmatrix} r & s\\ s & r\end{pmatrix} \otimes J_N & -\frac{1}{TN(N-1)}\iota_{2N}\iota_T'\\
-\frac{1}{TN(N-1)}\iota_T\iota_{2N}' & \frac{1}{N(N-1)}(I_T + \bar{J}_T)
\end{array}\right)
$$

with

$$
r = -\frac{-N^2 + 2N - 4}{4N^2(N-1)(N-2)T} \quad \text{and} \quad s = \frac{3N^2 - 10N + 4}{4N^2(N-1)(N-2)T}\,.
$$

As

$$\begin{aligned}(e_i', e_j', e_t')W_1^+ &= \tfrac{1}{TN(N-2)}\Big((N-1)(e_i', e_j', 0) + (e_j', e_i', 0)\Big) \\ &\quad + \Big(r+s-\tfrac{1}{TN(N-1)}\Big)(\iota_N', \iota_N', 0) + \tfrac{1}{TN(N-1)}(0, 0, Te_t' - \iota_T') \\ &= \tfrac{1}{TN(N-2)}\Big((N-1)(e_i', e_j', 0) + (e_j', e_i', 0)\Big) \\ &\quad - \tfrac{1}{2(N-1)(N-2)T}(\iota_N', \iota_N', 0) + \tfrac{1}{TN(N-1)}(0, 0, Te_t' - \iota_T'),\end{aligned}$$

the final form of the projection is obtained when the above expression is multiplied by $D'\tilde{L}\tilde{L}'y$, which in turn gives

$$\begin{aligned}\tilde{y}_{ijt} &= y_{ijt} - \tfrac{N-1}{N(N-2)T}(y_{i++} + y_{+j+}) - \tfrac{1}{N(N-2)T}(y_{j++} + y_{+i+}) \\ &\quad - \tfrac{1}{N(N-1)}y_{++t} + \tfrac{2}{N(N-2)T}y_{+++}.\end{aligned}$$

B.3 No Self-flow Transformation for Model (1.6)

The model in matrix form reads as

$$\begin{aligned}y &= X\beta + (I_N \otimes \iota_N \otimes I_T)\alpha + (\iota_N \otimes I_N \otimes I_T)\alpha^* + \varepsilon \\ &= X\beta + D(\alpha', \alpha^{*'})' + \varepsilon.\end{aligned}$$

With the same selection matrix $\tilde{L}$, we adjust the model to reflect no self-flow data,

$$\tilde{L}'y = \tilde{L}'X\beta + \tilde{L}'D(\alpha', \alpha^{*'})' + \tilde{L}'\varepsilon.$$

The optimal effects-eliminating projection matrix is

$$M_{\tilde{L}'D} = I_{N(N-1)T} - \tilde{L}'DW_2^+D'\tilde{L},$$

with $W_2 = D'\tilde{L}\tilde{L}'D$. As $(e_i \otimes e_j \otimes e_t)'\tilde{L}\tilde{L}'D = (e_i', e_j', e_t')$ holds again,

$$\begin{aligned}\tilde{y}_{ijt} &= (e_i \otimes e_j \otimes e_t)'\tilde{L}M_{\tilde{L}'D}\tilde{L}'y \\ &= y_{ijt} - (e_i', e_j', e_t')W_2^+D'\tilde{L}\tilde{L}'y.\end{aligned}$$

Further, as

$$D'\tilde{L}\tilde{L}'D = \begin{pmatrix} (N-1)(I_N \otimes I_T) & ((J_N - I_N) \otimes I_T) \\ ((J_N - I_N) \otimes I_T) & (N-1)(I_N \otimes I_T) \end{pmatrix},$$

W_2 is simply

$$\begin{aligned}W_2 &= \begin{pmatrix} N-1 & -1 \\ -1 & N-1 \end{pmatrix} \otimes I_{NT} + \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \otimes J_N \otimes I_T \\ &= \begin{pmatrix} N-1 & -1 \\ -1 & N-1 \end{pmatrix} \otimes Q_N \otimes I_T + (N-1)\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \otimes \bar{J}_N \otimes I_T,\end{aligned}$$

Since Q_N and $\bar{J}_N$ are idempotent and mutually orthogonal, the Moore-Penrose inverse W_2^+ of W_2 is

$$\begin{aligned} W_2^+ &= \tfrac{1}{N(N-2)} \begin{pmatrix} N-1 & 1 \\ 1 & N-1 \end{pmatrix} \otimes Q_N \otimes I_T + \tfrac{1}{4(N-1)} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \otimes \bar{J}_N \otimes I_T \\ &= \tfrac{1}{N(N-2)} \begin{pmatrix} N-1 & 1 \\ 1 & N-1 \end{pmatrix} \otimes I_{NT} + \tfrac{1}{N} \begin{pmatrix} p & q \\ q & p \end{pmatrix} \otimes J_N \otimes I_T \,, \end{aligned}$$

with again,

$$p = \frac{1}{4(N-1)} - \frac{N-1}{N(N-2)} \quad \text{and} \quad q = \frac{1}{4(N-1)} - \frac{1}{N(N-2)} \,.$$

Now, with this new form of W_2^+,

$$\begin{aligned} (e_i', e_j', e_t') W_2^+ &= \tfrac{1}{N(N-2)} \Big((N-1)e_i' + e_j', e_i' + (N-1)e_j', e_t' \Big) \\ &\quad - \tfrac{1}{2(N-1)(N-2)} \left(\iota_N', \iota_N', e_t' \right) . \end{aligned}$$

So, multiplying $(e_i', e_i', e_t') W_2^+$ by $D'\tilde{L}\tilde{L}'y$ gives

$$\begin{aligned} \tilde{y}_{ijt} &= y_{ijt} - \tfrac{N-1}{N(N-2)} \left(y_{i+t} + y_{+jt} \right) - \tfrac{1}{N(N-2)} \left(y_{+it} + y_{j+t} \right) \\ &\quad + \tfrac{1}{(N-1)(N-2)} y_{++t} \,. \end{aligned}$$

B.4 No Self-flow Transformation for Model (1.7)

As the method developed so far leads to messy matrix formulations in case of model (1.7), we propose an other way to find the optimal scalar no self-flow Within transformation. With the same selection matrix $\tilde{L}$ as before, $\tilde{L}'y$ is the shortened y vector without the elements $i = j$. The main job is again to get an expression for $M_{\tilde{L}'D}$, or for $\tilde{L}M_{\tilde{L}'D}\tilde{L}'$ when we fill out arbitrary values for the missing elements with $i = j$, to preserve a simple data format. Remember, that D corresponds to the specific model's dummy variable structure. Let $H = \sum_i (e_i e_i' \otimes e_i)$, so

$$H'H = I_N \quad HH' = \sum_i (e_i e_i' \otimes e_i e_i') \quad \tilde{L}\tilde{L}' = I_{N^2T} - (HH' \otimes I_T) \quad \tilde{L}'(H \otimes I_T) = 0.$$

With $G = \hat{D}'(H \otimes I_T)$ and $V = (I - G'G)^-$, we obtain

$$\begin{aligned} M_{\tilde{L}'D} &= M_{\tilde{L}'\hat{D}} = I - \tilde{L}'\hat{D}(\hat{D}'\tilde{L}\tilde{L}'\hat{D})^- \hat{D}'\tilde{L} \\ &= I - \tilde{L}'\hat{D} \left[I - \hat{D}'(H \otimes I_T)(H \otimes I_T)'\hat{D} \right]^- \hat{D}'\tilde{L} \\ &= I - \tilde{L}'\hat{D}(I - GG')^- \hat{D}'\tilde{L} \\ &= I - \tilde{L}'\hat{D} \left[I + G(I - G'G)^- G' \right] \hat{D}'\tilde{L} \\ &= \tilde{L}'(I - \hat{D}\hat{D}')\tilde{L} - \tilde{L}'\hat{D}GVG'\hat{D}'\tilde{L} \\ &= \tilde{L}'M_D\tilde{L} - \tilde{L}'\hat{D}\hat{D}'(H \otimes I_T)V(H \otimes I_T)'\hat{D}\hat{D}'\tilde{L} . \end{aligned}$$

Notice, that first term leads to the usual optimal transformation M_D. The second term corrects for the missing $i = j$ observations. This is general and holds for any no self-flow model, however we want to apply it for model (1.7), so we use the specific dummy matrixes

$$\begin{aligned}
D &= ((I_N \otimes I_N \otimes \iota_T),\ (I_N \otimes \iota_N \otimes I_T),\ (\iota_N \otimes I_N \otimes I_T)) \\
\hat{D} &= ((G_N \otimes G_N \otimes \bar{\iota}_T),\ (G_N \otimes \bar{\iota}_N \otimes G_T),\ (\bar{\iota}_N \otimes G_N \otimes G_T),\ (G_N \otimes \bar{\iota}_N \otimes \bar{\iota}_T), \\
&\quad ,(\bar{\iota}_N \otimes G_N \otimes \bar{\iota}_T),\ (\bar{\iota}_N \otimes \bar{\iota}_N \otimes G_T),\ (\bar{\iota}_N \otimes \bar{\iota}_N \otimes \bar{\iota}_T))
\end{aligned}$$

and get

$$\begin{aligned}
\hat{D}\hat{D}' &= I - (Q_N \otimes Q_N \otimes Q_T), \quad \text{so} \\
\tilde{L}'\hat{D}\hat{D}'(H \otimes I_T) &= -\tilde{L}'(Q_N \otimes Q_N \otimes Q_T)(H \otimes I_T)\,.
\end{aligned}$$

To elaborate on, we need some auxiliary results.

$$\begin{aligned}
H'(Q_N \otimes Q_N)H &= Q_N \cdot Q_N = (I_N - \tfrac{1}{N}J_N) \cdot (I_N - \tfrac{1}{N}J_N) \\
&= (1 - \tfrac{2}{N})I_N + \tfrac{1}{N^2}J_N = (1 - \tfrac{2}{N})(Q_N + \bar{J}_N) + \tfrac{1}{N}\bar{J}_N \\
&= \tfrac{N-2}{N}Q_N + \tfrac{N-1}{N}\bar{J}_N\,.
\end{aligned}$$

Next,

$$\begin{aligned}
I - G'G &= I - (H \otimes I_T)'\hat{D}\hat{D}'(H \otimes I_T) \\
&= I - (H \otimes I_T)'(I - Q_N \otimes Q_N \otimes Q_T)(H \otimes I_T) \\
&= (H \otimes I_T)'(Q_N \otimes Q_N \otimes Q_T)(H \otimes I_T) \\
&= (Q_N \cdot Q_N) \otimes Q_T\,,
\end{aligned}$$

so

$$\begin{aligned}
V &= (I - G'G)^- = \left(\tfrac{N}{N-2}Q_N + \tfrac{N}{N-1}\bar{J}_N\right) \otimes Q_T \\
&= \left(\tfrac{N}{N-2}I_N - \tfrac{1}{(N-1)(N-2)}J_N\right) \otimes Q_T\,.
\end{aligned}$$

Further, since $e_i'Q_N = e_i' - \frac{1}{N}\iota_N'$, there holds for $i \neq j$

$$\begin{aligned}
(e_i \otimes e_j \otimes e_t)'(Q_N \otimes Q_N \otimes Q_T)(H \otimes I_T) &= [(e_i'Q_N \otimes e_j'Q_N)H] \otimes e_t'Q_T \\
&= (Q_N e_i \cdot Q_N e_j)' \otimes e_t'Q_T \\
&= \tfrac{1}{N^2}[\iota_N - N(e_i + e_j)]' \otimes e_t'Q_T\,.
\end{aligned}$$

So multiplying this elaborated form of $(e_i \otimes e_j \otimes e_t)'(Q_N \otimes Q_N \otimes Q_T)(H \otimes I_T)$ by V, we get

$$\begin{aligned}
&\left(\tfrac{1}{N^2}(\iota_N' - N(e_i' + e_j')) \otimes (e_t' - \tfrac{1}{T}\iota_T')\right) \cdot \left(\tfrac{N}{N-2}I_N - \tfrac{1}{(N-1)(N-2)}J_N\right) \otimes (I_T - \bar{J}_T) \\
&\quad = \left(\tfrac{1}{(N-1)(N-2)}\iota_N' - \tfrac{1}{N-2}(e_i' + e_j')\right) \otimes (e_t' - \tfrac{1}{T}\iota_T')\,.
\end{aligned}$$

Let Y now be of order $(N^2 \times T)$ such that $\mathrm{vec}(Y') = y$. Then

$$\begin{aligned}
(H \otimes I_T)'\hat{D}\hat{D}'\tilde{L}\tilde{L}'y &= (H \otimes I_T)'(I - Q_N \otimes Q_N \otimes Q_T)\mathrm{vec}(Y') \\
&= \mathrm{vec}(Y'H) - \mathrm{vec}(Q_T Y'(Q_N \otimes Q_N)H)\,.
\end{aligned}$$

After elaborating on this, we get $\tilde{y}$, a vector of size NT with typical elements

$$\tilde{y}_{it} = \frac{1}{N}y_{i+t} + \frac{1}{N}y_{+it} - \frac{1}{N^2}y_{++t} - \frac{1}{NT}y_{i++} - \frac{1}{NT}y_{+i+} + \frac{1}{N^2T}y_{+++}$$

So putting everything together, the adjustment we have to make on the complete data optimal Within transformation is

$$\begin{aligned}&-\tfrac{1}{N(N-2)}(y_{i+t}+y_{+it}+y_{j+t}+y_{+jt})+\tfrac{1}{N(N-2)T}(y_{i++}+y_{+i+}+y_{j++}+y_{+j+})\\&+\tfrac{3N-2}{N^2(N-1)(N-2)}(y_{++t}-y_{+++}),\end{aligned}$$

leading to the optimal no self-flow Within transformation

$$\begin{aligned}y_{ijt} \quad &-\tfrac{N-3}{N(N-2)}(y_{i+t}+y_{+jt})+\tfrac{N-3}{N(N-2)T}(y_{i++}+y_{+j+})+\tfrac{1}{N(N-2)}(y_{+it}+y_{j+t})\\&-\tfrac{1}{N(N-2)T}(y_{+i+}+y_{j++})-\tfrac{1}{T}y_{ij+}+\tfrac{N^2-6N+4}{N^2(N-1)(N-2)}(y_{++t}-y_{+++}).\end{aligned}$$

Note, that this method is also perfectly legitimate for all fixed effects models listed in Section 1.2. Let's consider for example model (1.3). We know, that its optimal Within transformation in case of the lack of self-flow is, for $i \neq j$,

$$\tilde{y}_{ijt} = y_{ijt} - \frac{1}{T}\bar{y}_{ij+}.$$

In principal,

$$M_{\tilde{L}'D} = \tilde{L}'M_D\tilde{L} - \tilde{L}'\hat{D}\hat{D}'(H\otimes I_T)V(H\otimes I_T)'\hat{D}\hat{D}'\tilde{L}$$

should also give the same result. Notice, that as

$$\hat{D} = (I_N \otimes I_N \otimes \bar{\iota}_T),$$

now $\hat{D}\hat{D}' = (I_{N^2}\otimes Q_T)$, and $M_D = I_{N^2T} - (I_{N^2}\otimes \bar{J}_T)$. But as, for $i \neq j$,

$$\begin{aligned}(e_i\otimes e_j\otimes e_t)'\tilde{L}'\hat{D}\hat{D}'(H\otimes I_T) &= (e_i\otimes e_j\otimes e_t)'(I_{N^2}\otimes Q_T)\cdot(H\otimes I_T)\\&= (e_i\otimes e_j\otimes e_t)'\cdot(H\otimes Q_T)\\&= (e_i\otimes e_j\otimes e_t)'\cdot(\textstyle\sum_i(e_ie_i'\otimes e_i)\otimes Q_T)\\&= 0,\end{aligned}$$

$M_{\tilde{L}'D}$ reduces to $\tilde{L}'M_D\tilde{L}$, which in turn gives the scalar form

$$\tilde{y}_{ijt} = (e_i\otimes e_j\otimes e_t)'\tilde{L}'M_D\tilde{L}\tilde{L}'y = y_{ijt} - \frac{1}{T}\bar{y}_{ij+}.$$

10.14754/CEU.2017.05

2

Modelling Multi-dimensional Panel Data: A Random Effects Approach

Sections 2.2–2.4 are joint works with Badi H. Baltagi, Laszlo Matyas and Daria Pus, Sections 2.5 and 2.6.2–2.6.3 are joint works with Mark N. Harris, Felix Chan and Maurice Bun. Sections 2.7 and the rest of 2.6.1 are solely my own.

2.1 Introduction

The disturbances of an econometric model in principle include all factors influencing the behaviour of the dependent variable, which cannot be explicitly specified. In a statistical sense this means all terms about which we do not have enough information. In this chapter we deal with the cases when the individual and/or time specific factors, and the possible interaction effects between them are considered as unobserved heterogeneity, and as such are represented by random variables, and are part of the composite disturbance terms. From a more practical point of view, unlike the fixed effects approach, as seen in Chapter 1, this random effects approach has the advantage that the number of parameters to take into account does not increase with the sample size. It also makes possible the identification of parameters associated with some time and/or individual invariant variables (see *e.g.* p.60 in Baltagi et al., 2008 or Hornok, 2011).

Historically, multi-dimensional random effects (or error components) models can be traced back to the variance component analysis literature (see Rao and Kleffe, 1980, or the seminal results of Laird and Ware, 1982 or Leeuw and Kreft, 1986), and are related to the multi-level models, well known in statistics (see, for example, Scott et al., 2013, Luke, 2004, Goldstein, 1995, and Bryk and Raudenbush, 1992). We, however, assume fixed slope parameters for the regressors (rather than a composition of fixed and random elements), and zero means for the random components.

This chapter follows in spirit the analysis of the two-way panels by Baltagi et al.

(2008), that is, in Section 2.2 we introduce the most frequently used models in a three-dimensional (3D) panel data setup, Section 2.3 deals with the Feasible GLS estimation of these models, while Section 2.4 analyses the behaviour of this estimator for incomplete/unbalanced data. Section 2.5 investigates the case when regressors are potentially correlated with the random effects and proposes instrumental variable (IV) estimators to obtain consistent estimators for all model parameters. Section 2.6 lists various tests concerning random effects panel models: a test for model selection/specification, a test for identifying the persistence and the sources of endogeneity, and a test for instrument validity are proposed. Section 2.7 generalizes the presented models to four and higher dimensional data sets, and extends the random effects approach toward a mixed effects framework, and finally, Section 2.8 concludes.

2.2 Different Model Specifications

In this section, we present the most relevant three-dimensional model formulations, paying special attention to the different interaction effects. The models we encounter have empirical relevance, and correspond to some fixed effects model formulations known from the literature (see, for example, Baltagi et al., 2003; Egger and Pfaffermayr, 2003; Baldwin and Taglioni, 2006; Baier and Bergstrand, 2007).

The general form of these random effects (or error components) models can be casted as

$$y = X\beta + u, \tag{2.1}$$

where y and X are respectively the vector and matrix of observations of the dependent and explanatory variables, β is the vector of unknown (slope) parameters, and we want to exploit the structure embedded in the random disturbance terms u. As it is well known from the Gauss-Markov theorem, the General Least Squares (GLS) estimator is BLUE for β. To make it operational, in principle, we have to perform three steps. First, using the specific structure of u, we have to derive the variance-covariance matrix of model (2.1), $E(uu') = \Omega$, then, preferably using spectral decomposition, we have to derive its inverse. This is important, as multi-dimensional data often tend to be very large, leading to some Ω-s of extreme order. And finally, we need to estimate the unknown variance components of Ω to arrive to the well known Feasible GLS (FGLS) formulation.

2.2.1 Various Heterogeneity Formulations

The most general model formulation in a three-dimensional setup encompassing all pairwise random effects is

$$y_{ijt} = x'_{ijt}\beta + \mu_{ij} + \upsilon_{it} + \zeta_{jt} + \varepsilon_{ijt}\,, \tag{2.2}$$

where $i = 1\ldots N_1$, $j = 1\ldots N_2$, and $t = 1\ldots T$. Note, that y_{ijt}, x'_{ijt}, and $u_{ijt} = \mu_{ij} + v_{it} + \zeta_{jt} + \varepsilon_{ijt}$ are an element of the $(N_1N_2T \times 1)$, $(N_1N_2T \times K)$, and $(N_1N_2T \times 1)$ sized vectors and matrix y, X, and u respectively, of the general formulation (2.1), and β is the $(K \times 1)$ vector of parameters. We assume the random effects to be well-behaved ones, that is, to be pairwise uncorrelated, $E(\mu_{ij}) = 0$, $E(\upsilon_{it}) = 0$, $E(\zeta_{jt}) = 0$, and further,

$$\begin{aligned}
E(\mu_{ij}\mu_{i'j'}) &= \begin{cases} \sigma^2_\mu & i = i' \text{ and } j = j' \\ 0 & \text{otherwise} \end{cases} \\
E(\upsilon_{it}\upsilon_{i't'}) &= \begin{cases} \sigma^2_\upsilon & i = i' \text{ and } t = t' \\ 0 & \text{otherwise} \end{cases} \\
E(\zeta_{jt}\zeta_{j't'}) &= \begin{cases} \sigma^2_\zeta & j = j' \text{ and } t = t' \\ 0 & \text{otherwise} \end{cases}
\end{aligned}$$

Most importantly, the regressors are assumed to be uncorrelated with the random effects (strict exogeneity), which underpins the use of GLS-type estimators. This assumption is later relaxed in Section (2.5). The covariance matrix of such error components structure is simply

$$\Omega = E(uu') = \sigma^2_\mu(I_{N_1N_2} \otimes J_T) + \sigma^2_\upsilon(I_{N_1} \otimes J_{N_2} \otimes I_T) + \sigma^2_\zeta(J_{N_1} \otimes I_{N_2T}) + \sigma^2_\varepsilon I_{N_1N_2T}\,, \tag{2.3}$$

where we keep sticking to standard ANOVA notation, so I_{N_1} and J_{N_1} are the identity, and the square matrix of ones respectively, with the size in the index.

All other relevant model specifications are obtained by applying some restrictions on the random effects structure, that is all covariance structures are nested into that of model (2.2). The model which only uses individual-time-varying effects reads as

$$y_{ijt} = x'_{ijt}\beta + \upsilon_{it} + \zeta_{jt} + \varepsilon_{ijt}\,, \tag{2.4}$$

together with the appropriate assumptions listed for model (2.2). Now

$$\Omega = \sigma^2_\upsilon(I_{N_1} \otimes J_{N_2} \otimes I_T) + \sigma^2_\zeta(J_{N_1} \otimes I_{N_2} \otimes I_T) + \sigma^2_\varepsilon I_{N_1N_2T}\,. \tag{2.5}$$

A further restriction on the above model is

$$y_{ijt} = x'_{ijt}\beta + \zeta_{jt} + \varepsilon_{ijt}\,, \tag{2.6}$$

which in fact is a generalization of the approach used in multi-level modeling, see for example, Ebbes et al. (2004) or Hubler (2006).[1] The covariance matrix now is

$$\Omega = \sigma_\zeta^2 (J_{N_1} \otimes I_{N_2 T}) + \sigma_\varepsilon^2 I_{N_1 N_2 T} \,. \tag{2.7}$$

Another restrictions of model (2.2) is to leave in the pair-wise random effects, and restrict the individual-time-varying terms. Specifically, model

$$y_{ijt} = x'_{ijt}\beta + \mu_{ij} + \lambda_t + \varepsilon_{ijt} \tag{2.8}$$

incorporates both time and individual-pair random effects. We assume, as before, that $E(\lambda_t) = 0$, and that

$$E(\lambda_t \lambda'_t) = \begin{cases} \sigma_\lambda^2 & t = t' \\ 0 & \text{otherwise} \end{cases}$$

Now

$$\Omega = \sigma_\mu^2 (I_{N_1 N_2} \otimes J_T) + \sigma_\lambda^2 (J_{N_1 N_2} \otimes I_T) + \sigma_\varepsilon^2 I_{N_1 N_2 T} \,. \tag{2.9}$$

A restriction of the above model, when we assume, that $\mu_{ij} = \upsilon_i + \zeta_j$ is[2]

$$y_{ijt} = x'_{ijt}\beta + \upsilon_i + \zeta_j + \lambda_t + \varepsilon_{ijt} \tag{2.10}$$

with the usual assumptions $E(\upsilon_i) = E(\zeta_j) = E(\lambda_t) = 0$, and

$$\begin{aligned} E(\upsilon_i \upsilon_{i'}) &= \begin{cases} \sigma_\upsilon^2 & i = i' \\ 0 & \text{otherwise} \end{cases} \\ E(\zeta_j \zeta_{j'}) &= \begin{cases} \sigma_\zeta^2 & j = j' \\ 0 & \text{otherwise} \end{cases} \\ E(\lambda_t \lambda_{t'}) &= \begin{cases} \sigma_\lambda^2 & t = t' \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

Its covariance structure is

$$\Omega = \sigma_\upsilon^2 (I_{N_1} \otimes J_{N_2 T}) + \sigma_\zeta^2 (J_{N_1} \otimes I_{N_2} \otimes J_T) + \sigma_\lambda^2 (J_{N_1 N_2} \otimes I_T) + \sigma_\varepsilon^2 I_{N_1 N_2 T} \,. \tag{2.11}$$

Lastly, the simplest model is

$$y_{ijt} = x'_{ijt}\beta + \mu_{ij} + \varepsilon_{ijt} \tag{2.12}$$

with

$$\Omega = \sigma_\mu^2 (I_{N_1 N_2} \otimes J_T) + \sigma_\varepsilon^2 I_{N_1 N_2 T} \,. \tag{2.13}$$

[1] The symmetric counterpart of model (2.6), with υ_{it} random effects, could also be listed here, however, as it has the exact same properties as model (2.6), we take the two models together.

[2] This model has in fact been introduced in Matyas (1998), and before that, in Ghosh (1976).

Note, that model (2.12) and (2.8) can be considered in fact as straight panel data models, where the individuals are now the (ij) pairs (so essentially it does not take into account the three-dimensional nature of the data). In this sense, models (2.2), (2.4) and (2.10) are more advantageous and are flexible in tackling problems, which can not be handled with standard 2D panels.

2.2.2 Spectral Decomposition of the Covariance Matrices

To estimate the above models, the inverse of Ω is needed, a matrix of size $(N_1N_2T \times N_1N_2T)$. For even moderately large samples, this can be unfeasible in practice without further elaboration. The common practice is to use the spectral decomposition of Ω, which in turn gives the inverse as a function of fairly standard matrices (see Wansbeek and Kapteyn, 1982). We derive the algebra for model (2.2), Ω^{-1} for all other models can de derived likewise, so we only present the final results. First, consider a simple rewriting of the identity matrix

$$I_{N_1} = Q_{N_1} + \bar{J}_{N_1}\,, \quad \text{where, remember} \quad Q_{N_1} = I_{N_1} - \bar{J}_{N_1}\,,$$

with $\bar{J}_{N_1} = 1/N_1 J_{N_1}$. Now Ω becomes

$$\begin{aligned}\Omega = \;& T\sigma_\mu^2((Q_{N_1} + \bar{J}_{N_1}) \otimes (Q_{N_2} + \bar{J}_{N_2}) \otimes \bar{J}_T) \\ & + N_2\sigma_\upsilon^2((Q_{N_1} + \bar{J}_{N_1}) \otimes \bar{J}_{N_2} \otimes (Q_T + \bar{J}_T)) \\ & + N_1\sigma_\zeta^2(\bar{J}_{N_1} \otimes (Q_{N_2} + \bar{J}_{N_2}) \otimes Q_T) \\ & + \sigma_\varepsilon^2((Q_{N_1} + \bar{J}_{N_1}) \otimes (Q_{N_2} + \bar{J}_{N_2}) \otimes (Q_T + \bar{J}_T))\,.\end{aligned}$$

If we unfold the brackets, the terms we get are in fact the between-group variations of each possible groups in three-dimensional data. For example, the building block

$$B_{ij.} = (Q_{N_1} \otimes Q_{N_2} \otimes \bar{J}_T)$$

captures the variation between i and j. All other B matrices are defined in a similar manner: the indices in the subscript indicate the variation with respect to which it is captured. The two extremes, B_{ijt} and $B_{...}$ are thus

$$B_{ijt} = (Q_{N_1} \otimes Q_{N_2} \otimes Q_T) \quad \text{and} \quad B_{...} = (\bar{J}_{N_1} \otimes \bar{J}_{N_2} \otimes \bar{J}_T)\,.$$

Notice, that the covariance matrix of all three-way error components model can be represented by these B building blocks. For model (2.2), this means

$$\begin{aligned}\Omega = \;& \sigma_\varepsilon^2 B_{ijt} + (\sigma_\varepsilon^2 + T\sigma_\mu^2)B_{ij.} + (\sigma_\varepsilon^2 + N_2\sigma_\upsilon^2)B_{i.t} + (\sigma_\varepsilon^2 + N_1\sigma_\zeta^2)B_{.jt} \\ & + (\sigma_\varepsilon^2 + T\sigma_\mu^2 + N_2\sigma_\upsilon^2)B_{i..} + (\sigma_\varepsilon^2 + T\sigma_\mu^2 + N_1\sigma_\zeta^2)B_{.j.} \\ & + (\sigma_\varepsilon^2 + N_2\sigma_\upsilon^2 + N_1\sigma_\zeta^2)B_{..t} + (\sigma_\varepsilon^2 + T\sigma_\mu^2 + N_2\sigma_\upsilon^2 + N_1\sigma_\zeta^2)B_{...}\,.\end{aligned} \tag{2.14}$$

Also notice, that all B matrices are idempotent and mutually orthogonal by construction (as $Q_{N_1}\bar{J}_{N_1} = 0$, likewise with N_2 and T), so

$$\begin{aligned}\Omega^{-1} = \; & \tfrac{1}{\sigma_\varepsilon^2}B_{ijt} + \tfrac{1}{\sigma_\varepsilon^2+T\sigma_\mu^2}B_{ij.} + \tfrac{1}{\sigma_\varepsilon^2+N_2\sigma_\upsilon^2}B_{i.t} + \tfrac{1}{\sigma_\varepsilon^2+N_1\sigma_\zeta^2}B_{.jt} \\ & + \tfrac{1}{\sigma_\varepsilon^2+T\sigma_\mu^2+N_2\sigma_\upsilon^2}B_{i..} + \tfrac{1}{\sigma_\varepsilon^2+T\sigma_\mu^2+N_1\sigma_\zeta^2}B_{.j.} \\ & + \tfrac{1}{\sigma_\varepsilon^2+N_2\sigma_\upsilon^2+N_1\sigma_\zeta^2}B_{..t} + \tfrac{1}{\sigma_\varepsilon^2+T\sigma_\mu^2+N_2\sigma_\upsilon^2+N_1\sigma_\zeta^2}B_{...}\,.\end{aligned}$$

This means that we can get the inverse of a covariance matrix at virtually no computational cost, as a function of some standard B matrices. After some simplification, we get

$$\begin{aligned}\sigma_\varepsilon^2\Omega^{-1} = \; & I_{N_1N_2T} - (1-\theta_1)(\bar{J}_{N_1}\otimes I_{N_2T}) - (1-\theta_2)(I_{N_1}\otimes\bar{J}_{N_2}\otimes I_T) \\ & -(1-\theta_3)(I_{N_1N_2}\otimes\bar{J}_T) + (1-\theta_1-\theta_2+\theta_4)(\bar{J}_{N_1N_2}\otimes I_T) \\ & +(1-\theta_1-\theta_3+\theta_5)(\bar{J}_{N_1}\otimes I_{N_2}\otimes\bar{J}_T) \\ & +(1-\theta_2-\theta_3+\theta_6)(I_{N_1}\otimes\bar{J}_{N_2T}) \\ & -(1-\theta_1-\theta_2-\theta_3+\theta_4+\theta_5+\theta_6-\theta_7)\bar{J}_{N_1N_2T}\,,\end{aligned} \tag{2.15}$$

with

$$\begin{aligned}\theta_1 & = \tfrac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+N_1\sigma_\zeta^2}, \quad \theta_2 = \tfrac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+N_2\sigma_\upsilon^2}, \quad \theta_3 = \tfrac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+T\sigma_\mu^2} \\ \theta_4 & = \tfrac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+N_2\sigma_\upsilon^2+N_1\sigma_\zeta^2}, \quad \theta_5 = \tfrac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+T\sigma_\mu^2+N_1\sigma_\zeta^2}, \\ \theta_6 & = \tfrac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+T\sigma_\mu^2+N_2\sigma_\upsilon^2}, \quad \text{and} \quad \theta_7 = \tfrac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+T\sigma_\mu^2+N_2\sigma_\upsilon^2+N_1\sigma_\zeta^2}\,.\end{aligned}$$

The good thing is that we can fully get rid of the matrix notations, following Fuller and Battese (1973), as $\sigma_\varepsilon^2\Omega^{-1/2}y$ can be written up in scalar form as well. This transformation can be represented with its typical element

$$\begin{aligned}\tilde{y}_{ijt} = \; & y_{ijt} - (1-\sqrt{\theta_1})\bar{y}_{.jt} - (1-\sqrt{\theta_2})\bar{y}_{i.t} - (1-\sqrt{\theta_3})\bar{y}_{ij.} \\ & +(1-\sqrt{\theta_1}-\sqrt{\theta_2}+\sqrt{\theta_4})\bar{y}_{..t} \\ & +(1-\sqrt{\theta_1}-\sqrt{\theta_3}+\sqrt{\theta_5})\bar{y}_{.j.} + (1-\sqrt{\theta_2}-\sqrt{\theta_3}+\sqrt{\theta_6})\bar{y}_{i..} \\ & -(1-\sqrt{\theta_1}-\sqrt{\theta_2}-\sqrt{\theta_3}+\sqrt{\theta_4}+\sqrt{\theta_5}+\sqrt{\theta_6}-\sqrt{\theta_7})\bar{y}_{...}\,.\end{aligned}$$

By using the OLS on these transformed variables, we get back the GLS estimator. For other models, the job is essentially the same. For model (2.4),

$$\begin{aligned}\sigma_\varepsilon^2\Omega^{-1} = \; & I_{N_1N_2T} - (I_{N_1}\otimes\bar{J}_{N_2}\otimes I_T) - (\bar{J}_{N_1}\otimes I_{N_2T}) + (\bar{J}_{N_1N_2}\otimes I_T) \\ & + \tfrac{\sigma_\varepsilon^2}{N_1\sigma_\zeta^2+\sigma_\varepsilon^2}((\bar{J}_{N_1}\otimes I_{N_2T}) - (\bar{J}_{N_1N_2}\otimes I_T)) \\ & + \tfrac{\sigma_\varepsilon^2}{N_2\sigma_\upsilon^2+\sigma_\varepsilon^2}((I_{N_1}\otimes\bar{J}_{N_2}\otimes I_T)) - (\bar{J}_{N_1N_2}\otimes I_T)) \\ & + \tfrac{\sigma_\varepsilon^2}{N_2\sigma_\upsilon^2+N_1\sigma_\zeta^2+\sigma_\varepsilon^2}(\bar{J}_{N_1N_2}\otimes I_T)\,,\end{aligned}$$

and so $\sigma_\varepsilon^2\Omega^{-1/2}y$ in scalar form reads as, with a typical $\tilde{y}_{ijt}$ element,

$$\tilde{y}_{ijt} = y_{ijt} - (1-\sqrt{\theta}_8)\bar{y}_{i.t} - (1-\sqrt{\theta}_9)\bar{y}_{.jt} + (1-\sqrt{\theta}_8-\sqrt{\theta}_9+\sqrt{\theta}_{10})\bar{y}_{..t}\,,$$

with

$$\theta_8 = \frac{\sigma_\varepsilon^2}{N_2\sigma_\upsilon^2 + \sigma_\varepsilon^2}\,, \quad \theta_9 = \frac{\sigma_\varepsilon^2}{N_1\sigma_\zeta^2 + \sigma_\varepsilon^2}\,, \quad \theta_{10} = \frac{\sigma_\varepsilon^2}{N_2\sigma_\upsilon^2 + N_1\sigma_\zeta^2 + \sigma_\varepsilon^2}\,.$$

For model (2.6), the inverse of the covariance matrix is even simpler,

$$\sigma_\varepsilon^2\Omega^{-1} = I_{N_1N_2T} - (\bar{J}_{N_1}\otimes I_{N_2T}) + \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + N_1\sigma_\zeta^2}(\bar{J}_{N_1}\otimes I_{N_2T})\,,$$

so $\sigma_\varepsilon^2\Omega^{-1/2}y$ defines the scalar transformation

$$\tilde{y}_{ijt} = y_{ijt} - (1-\sqrt{\theta}_{11})\bar{y}_{.jt}\,, \quad \text{with} \quad \theta_{11} = \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + N_1\sigma_\zeta^2}\,.$$

For Model (2.8), it is

$$\begin{aligned}\sigma_\varepsilon^2\Omega^{-1} = \;& I_{N_1N_2T} - (I_{N_1N_2}\otimes\bar{J}_T) - (\bar{J}_{N_1N_2}\otimes I_T) + \bar{J}_{N_1N_2T}\\ &+\tfrac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+T\sigma_\mu^2}((I_{N_1N_2}\otimes\bar{J}_T) - \bar{J}_{N_1N_2T})\\ &+\tfrac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+N_1N_2\sigma_\lambda^2}((\bar{J}_{N_1N_2}\otimes I_T) - \bar{J}_{N_1N_2T}) + \tfrac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+T\sigma_\mu^2+N_1N_2\sigma_\lambda^2}\bar{J}_{N_1N_2T}\,,\end{aligned}$$

so $\sigma_\varepsilon^2\Omega^{-1/2}y$ in scalar form is

$$\tilde{y}_{ijt} = y_{ijt} - (1-\sqrt{\theta}_{12})\bar{y}_{ij.} - (1-\sqrt{\theta}_{13})\bar{y}_{..t} + (1-\sqrt{\theta}_{12}-\sqrt{\theta}_{13}+\sqrt{\theta}_{14})\bar{y}_{...}\,,$$

with

$$\theta_{12} = \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T\sigma_\mu^2}\,, \quad \theta_{13} = \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + N_1N_2\sigma_\lambda^2}\,, \quad \theta_{14} = \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T\sigma_\mu^2 + N_1N_2\sigma_\lambda^2}\,.$$

The spectral decomposition of model (2.10), which was in fact proposed by Baltagi (1987), is

$$\begin{aligned}\sigma_\varepsilon^2\Omega^{-1} = \;& I_{N_1N_2T} - (\bar{J}_{N_1N_2}\otimes I_T) - (\bar{J}_{N_1}\otimes I_{N_2}\otimes\bar{J}_T) - (I_{N_1}\otimes\bar{J}_{N_2T})\\ &+2\bar{J}_{N_1N_2T} + \tfrac{\sigma_\varepsilon^2}{N_2T\sigma_\upsilon^2+\sigma_\varepsilon^2}((I_{N_1}\otimes\bar{J}_{N_2T}) - \bar{J}_{N_1N_2T})\\ &+\tfrac{\sigma_\varepsilon^2}{N_1T\sigma_\zeta^2+\sigma_\varepsilon^2}((\bar{J}_{N_1}\otimes I_{N_2}\otimes\bar{J}_T) - \bar{J}_{N_1N_2T})\\ &+\tfrac{\sigma_\varepsilon^2}{N_1N_2\sigma_\lambda^2+\sigma_\varepsilon^2}((\bar{J}_{N_1N_2}\otimes I_T) - \bar{J}_{N_1N_2T})\\ &+\tfrac{\sigma_\varepsilon^2}{N_2T\sigma_\upsilon^2+N_1T\sigma_\zeta^2+N_1N_2\sigma_\lambda^2+\sigma_\varepsilon^2}\bar{J}_{N_1N_2T}\,.\end{aligned}$$

With the covariance matrix in hand, $\sigma_\varepsilon^2\Omega^{-1/2}y$ translates into

$$\begin{aligned}\tilde{y}_{ijt} = \;& y_{ijt} - (1-\sqrt{\theta}_{15})\bar{y}_{i..} - (1-\sqrt{\theta}_{16})\bar{y}_{.j.} - (1-\sqrt{\theta}_{17})\bar{y}_{..t}\\ &+(2-\sqrt{\theta}_{15}-\sqrt{\theta}_{16}-\sqrt{\theta}_{17}+\sqrt{\theta}_{18})\bar{y}_{...}\,,\end{aligned}$$

10.14754/CEU.2017.05

Table 2.1 *Structure of the Ω^{-1} matrices*

Model	(2.2)	(2.4)	(2.6)	(2.8)	(2.10)	(2.12)
I_{N^2T}	$+$[a]	$+$	$+$	$+$	$+$	$+$
$(I_{N^2} \otimes \bar{J}_T)$	$+$			$+$		$+$
$(I_N \otimes \bar{J}_N \otimes I_T)$	$+$	$+$				
$(\bar{J}_N \otimes I_{NT})$	$+$	$+$	$+$			
$(I_N \otimes \bar{J}_{NT})$	$+$				$+$	
$(\bar{J}_N \otimes I_N \otimes \bar{J}_T)$	$+$				$+$	
$(\bar{J}_{N^2} \otimes I_T)$	$+$	$+$		$+$	$+$	
$\bar{J}_{N^2T}$	$+$			$+$	$+$	

[a] Notes: a "+" sign in a column says which building element is part of the given model's Ω^{-1}. If the "+"-s in the column a given model A cover that of another model B's means that model B is nested into model A. It can be seen, for example, that all models are in fact nested into (2.2), or that model (2.12) is nested into model (2.8).

where

$$\theta_{15} = \frac{\sigma_\varepsilon^2}{N_2 T \sigma_\upsilon^2 + \sigma_\varepsilon^2}\,, \qquad \theta_{16} = \frac{\sigma_\varepsilon^2}{N_1 T \sigma_\zeta^2 + \sigma_\varepsilon^2}\,, \quad \theta_{17} = \frac{\sigma_\varepsilon^2}{N_1 N_2 \sigma_\lambda^2 + \sigma_\varepsilon^2}\,, \quad \text{and}$$
$$\theta_{18} = \frac{\sigma_\varepsilon^2}{N_2 T \sigma_\upsilon^2 + N_1 T \sigma_\zeta^2 + N_1 N_2 \sigma_\lambda^2 + \sigma_\varepsilon^2}\,.$$

For model (2.12), the inversion gives

$$\sigma_\varepsilon^2 \Omega^{-1} = I_{N_1N_2T} - (I_{N_1N_2} \otimes \bar{J}_T) + \frac{\sigma_\varepsilon^2}{T\sigma_\mu^2 + \sigma_\varepsilon^2}(I_{N_1N_2} \otimes \bar{J}_T)\,,$$

and so $\sigma_\varepsilon^2 \Omega^{-1/2} y$ can be written up in scalar form, represented by a typical element

$$\tilde{y}_{ijt} = y_{ijt} - (1 - \sqrt{\theta}_{19})\bar{y}_{ij.}\,, \quad \text{with} \quad \theta_{19} = \frac{\sigma_\varepsilon^2}{T\sigma_\mu^2 + \sigma_\varepsilon^2}\,.$$

Table 2.1 summarizes the key elements in each models' inverse covariance matrix in the finite case.

When the number of observations grow in one or more dimensions, it can be interesting to find the limits of the θ_k weights. It is easy to see, that if all N_1, N_2, and $T \to \infty$, all θ_k, $(k = 1, \dots, 19)$ are in fact going to zero. That is, if the data grows in all directions, the GLS estimator (and in turn the FGLS) is identical to the Within Estimator. Hence, for example for model (2.2), in the limit, $\sigma_\varepsilon^2 \Omega^{-1}$ is simply given by

$$\begin{aligned}\lim_{N_1,N_2,T\to\infty} \sigma_\varepsilon^2 \Omega^{-1} = \;& I_{N_1N_2T} - (\bar{J}_{N_1} \otimes I_{N_2T}) - (I_{N_1} \otimes \bar{J}_{N_2} \otimes I_T) \\ & -(I_{N_1N_2} \otimes \bar{J}_T) + (\bar{J}_{N_1N_2} \otimes I_T) + (\bar{J}_{N_1} \otimes I_{N_2} \otimes \bar{J}_T) \\ & +(I_{N_1} \otimes \bar{J}_{N_2T}) - \bar{J}_{N_1N_2T}\,,\end{aligned}$$

Table 2.2 *Asymptotic conditions when the models' FGLS converges to a Within estimator*

Model	Condition
(2.2)	$N_1 \to \infty, N_2 \to \infty, T \to \infty$
(2.4)	$N_1 \to \infty, N_2 \to \infty$
(2.6)	$N_1 \to \infty$
(2.8)	$(N_1 \to \infty,\ T \to \infty)$ or $(N_2 \to \infty,\ T \to \infty)$
(2.10)	$(N_1 \to \infty,\ N_2 \to \infty)$ or $(N_1 \to \infty,\ T \to \infty)$ or $(N_2 \to \infty,\ T \to \infty)$
(2.12)	$T \to \infty$

which is the covariance matrix of the Within estimator. Table 2.2 collects the asymptotic conditions, when the models' (F)GLS estimator is converging to a Within estimator.

2.3 FGLS Estimation

To make the FGLS estimator operational, we need estimators for the variance components. Let us start again with model (2.2), for the other models, the job is essentially the same. Using the assumptions that the error components are pairwise uncorrelated,

$$\begin{aligned} \mathrm{E}(u_{ijt}^2) &= \mathrm{E}((\mu_{ij} + \upsilon_{it} + \zeta_{jt} + \varepsilon_{ijt})^2) \\ &= \mathrm{E}(\mu_{ij}^2) + \mathrm{E}(\upsilon_{it}^2) + \mathrm{E}(\zeta_{jt}^2) + \mathrm{E}(\varepsilon_{ijt}^2) = \sigma_\mu^2 + \sigma_\upsilon^2 + \sigma_\zeta^2 + \sigma_\varepsilon^2 \,. \end{aligned}$$

By introducing different Within transformations and so projecting the error components into different subspaces of the original three-dimensional space, we can derive further identifying equations. The appropriate Within transformation for model (2.2) (see for details Balazsi et al., 2015) is

$$\tilde{u}_{ijt} = u_{ijt} - \bar{u}_{.jt} - \bar{u}_{i.t} - \bar{u}_{ij.} + \bar{u}_{..t} + \bar{u}_{.j.} + \bar{u}_{i..} - \bar{u}_{...} \,. \tag{2.16}$$

Note, that this transformation corresponds to the projection matrix

$$\begin{aligned} M = \;& I_{N_1N_2T} - (I_{N_1N_2} \otimes \bar{J}_T) - (I_{N_1} \otimes \bar{J}_{N_2} \otimes I_T) - (\bar{J}_{N_1} \otimes I_{N_2T}) \\ & + (I_{N_1} \otimes \bar{J}_{N_2T}) + (\bar{J}_{N_1} \otimes I_{N_2} \otimes \bar{J}_T) + (\bar{J}_{N_1N_2} \otimes I_T) - \bar{J}_{N_1N_2T} \,, \end{aligned}$$

and u has to be pre-multiplied with it. Transforming u_{ijt} according to this wipes out μ_{ij}, υ_{it}, ζ_{jt}, and gives, with $i = 1 \dots N_1$, and $j = 1 \dots N_2$,

$$\begin{aligned} \mathrm{E}(\tilde{u}_{ijt}^2) = \mathrm{E}(\tilde{\varepsilon}_{ijt}^2) &= \mathrm{E}((\varepsilon_{ijt} - \bar{\varepsilon}_{.jt} - \bar{\varepsilon}_{i.t} - \bar{\varepsilon}_{ij.} + \bar{\varepsilon}_{..t} + \bar{\varepsilon}_{.j.} + \bar{\varepsilon}_{i..} - \bar{\varepsilon}_{...})^2) \\ &= \tfrac{(N_1-1)(N_2-1)(T-1)}{N_1N_2T} \sigma_\varepsilon^2 \,, \end{aligned}$$

where $\frac{(N_1-1)(N_2-1)(T-1)}{N_1N_2T}$ is the rank/order ratio of M, likewise for all other subsequent transformations. Further, transforming u_{ijt} according to

$$\begin{aligned}\tilde{u}^a_{ijt} &= u_{ijt} - \bar{u}_{.jt} - \bar{u}_{i.t} + \bar{u}_{..t}\text{ ,or with the underlying matrix}\\ M^a &= I_{N_1N_2T} - (\bar{J}_{N_1} \otimes I_{N_2T}) - (I_{N_1} \otimes \bar{J}_{N_2} \otimes I_T) + (\bar{J}_{N_1N_2} \otimes I_T)\end{aligned}$$

eliminates $\upsilon_{it} + \zeta_{jt}$, and gives

$$\begin{aligned}\mathrm{E}((\tilde{u}^a_{ijt})^2) = \mathrm{E}((\tilde{\mu}^a_{ij} + \tilde{\varepsilon}^a_{ijt})^2) &= \mathrm{E}((\tilde{\mu}^a_{ij})^2) + \mathrm{E}((\tilde{\varepsilon}^a_{ijt})^2)\\ &= \tfrac{(N_1-1)(N_2-1)}{N_1N_2}(\sigma^2_\mu + \sigma^2_\varepsilon)\,.\end{aligned}$$

Transforming according to

$$\begin{aligned}\tilde{u}^b_{ijt} &= u_{ijt} - \bar{u}_{ij.} - \bar{u}_{.jt} + \bar{u}_{.j.}\,, \quad \text{or}\\ M^b &= I_{N_1N_2T} - (I_{N_1N_2} \otimes \bar{J}_T) - (\bar{J}_{N_1} \otimes I_{N_2T}) + (\bar{J}_{N_1} \otimes I_{N_2} \otimes \bar{J}_T)\end{aligned}$$

eliminates $\mu_{ij} + \zeta_{jt}$, and gives

$$\mathrm{E}((\tilde{u}^b_{ijt})^2) = \mathrm{E}((\tilde{\upsilon}^b_{it} + \tilde{\varepsilon}^b_{ijt})^2) = \mathrm{E}((\tilde{\upsilon}^b_{it})^2) + \mathrm{E}((\tilde{\varepsilon}^b_{ijt})^2) = \frac{(N_1-1)(T-1)}{N_1T}(\sigma^2_\upsilon + \sigma^2_\varepsilon)\,.$$

Finally, using

$$\begin{aligned}\tilde{u}^c_{ijt} &= u_{ijt} - \bar{u}_{ij.} - \bar{u}_{i.t} + \bar{u}_{i..}\,, \quad \text{or}\\ M^c &= I_{N_1N_2T} - (I_{N_1N_2} \otimes \bar{J}_T) - (I_{N_1} \otimes \bar{J}_{N_2} \otimes I_T) + (I_{N_1} \otimes \bar{J}_{N_2T})\end{aligned}$$

wipes μ_{ij} and υ_{it} out, and gives

$$\begin{aligned}\mathrm{E}((\tilde{u}^c_{ijt})^2) = \mathrm{E}((\tilde{\zeta}^c_{jt} + \tilde{\varepsilon}^c_{ijt})^2) &= \mathrm{E}((\tilde{\zeta}^c_{jt})^2) + \mathrm{E}((\tilde{\varepsilon}^c_{ijt})^2)\\ &= \tfrac{(N_2-1)(T-1)}{N_2T}(\sigma^2_\zeta + \sigma^2_\varepsilon)\,.\end{aligned}$$

Putting the four identifying equations together gives a solvable system of four equations. Let $\hat{u}_{ijt}$ be the residual from the OLS estimation of $y = X\beta + u$. With this notation, the estimators for the variance components are

$$\begin{aligned}\hat{\sigma}^2_\varepsilon &= \tfrac{1}{(N_1-1)(N_2-1)(T-1)}\textstyle\sum_{ijt}\tilde{\hat{u}}^2_{ijt}\\ \hat{\sigma}^2_\mu &= \tfrac{1}{(N_1-1)(N_2-1)T}\textstyle\sum_{ijt}(\tilde{\hat{u}}^a_{ijt})^2 - \hat{\sigma}^2_\varepsilon\\ \hat{\sigma}^2_\upsilon &= \tfrac{1}{(N_1-1)N_2(T-1)}\textstyle\sum_{ijt}(\tilde{\hat{u}}^b_{ijt})^2 - \hat{\sigma}^2_\varepsilon\\ \hat{\sigma}^2_\zeta &= \tfrac{1}{N_1(N_2-1)(T-1)}\textstyle\sum_{ijt}(\tilde{\hat{u}}^c_{ijt})^2 - \hat{\sigma}^2_\varepsilon\,.\end{aligned}$$

where, obviously, $\tilde{\hat{u}}_{ijt}$, $\tilde{\hat{u}}^a_{ijt}$, $\tilde{\hat{u}}^b_{ijt}$, and $\tilde{\hat{u}}^c_{ijt}$ are the transformed residuals according to M, M^a, M^b, and M^c respectively.

Note, however, that the FGLS estimator of model (2.2) is only consistent if the data grows in at least two dimensions, that is, any two of $N_1 \to \infty$, $N_2 \to \infty$, and $T \to \infty$ has to hold. This is, because σ^2_μ (the variance of μ_{ij}) cannot be estimated consistently, when only $T \to \infty$, σ^2_υ, or when only $N_1 \to \infty$, and so on. For the consistency of the FGLS we need all variance components to be estimated consistently, something which holds only if the data grows in at least two dimensions. Table 2.3 collects the conditions needed for consistency for all models considered. So what if, for example, the data is such that N_1 is large, but N_2 and T are small (like in case, for example, of an employee-firm data with an extensive number of workers, but with few hiring firms observed annually)? This would mean, that σ^2_μ and σ^2_υ is estimated consistently, unlike σ^2_ζ. In such cases, it makes more sense to assume ζ_{jt} to be fixed instead of random (while still assuming the randomness of μ_{ij} and υ_{it}), arriving to the so-called "mixed effects models", something explored in Section 2.7.

We can estimate the variance components of the other models in a similar way. As the algebra is essentially the same, we only present here the main results. For model (2.4),

$$\begin{aligned} \mathrm{E}(\tilde{u}^2_{ijt}) &= \tfrac{(N_1-1)(N_2-1)}{N_1N_2}\sigma^2_\varepsilon\,, \quad \mathrm{E}((\tilde{u}^a_{ijt})^2) = \tfrac{N_1-1}{N_1}(\sigma^2_\upsilon+\sigma^2_\varepsilon) \quad \text{and} \\ \mathrm{E}((\tilde{u}^b_{ijt})^2) &= \tfrac{N_2-1}{N_2}(\sigma^2_\zeta+\sigma^2_\varepsilon)\,, \end{aligned}$$

now with $\tilde{u}_{ijt} = u_{ijt} - \bar{u}_{.jt} - \bar{u}_{i.t} + \bar{u}_{..t}$, and $\tilde{u}^a_{ijt} = u_{ijt} - \bar{u}_{.jt}$, and $\tilde{u}^b_{ijt} = u_{ijt} - \bar{u}_{i.t}$, which correspond to the projection matrices

$$\begin{aligned} M &= I_{N_1N_2T} - (\bar{J}_{N_1}\otimes I_{N_2T}) - (I_{N_1}\otimes\bar{J}_{N_2}\otimes I_T) + (\bar{J}_{N_1N_2}\otimes I_T) \\ M^a &= I_{N_1N_2T} - (\bar{J}_{N_1}\otimes I_{N_2T}) \\ M^b &= I_{N_1N_2T} - (I_{N_1}\otimes\bar{J}_{N_2}\otimes I_T) \end{aligned}$$

respectively. The estimators for the variance components then are

$$\begin{aligned} &\hat{\sigma}^2_\varepsilon = \tfrac{1}{(N_1-1)(N_2-1)T}\textstyle\sum_{ijt}\tilde{\hat{u}}^2_{ijt}\,, \quad \hat{\sigma}^2_\upsilon = \tfrac{1}{(N_1-1)N_2T}\textstyle\sum_{ijt}(\tilde{\hat{u}}^a_{ijt})^2 - \hat{\sigma}^2_\varepsilon\,, \quad \text{and} \\ &\hat{\sigma}^2_\zeta = \tfrac{1}{N_1(N_2-1)T}\textstyle\sum_{ijt}(\tilde{\hat{u}}^b_{ijt})^2 - \hat{\sigma}^2_\varepsilon\,, \end{aligned}$$

where again, $\tilde{\hat{u}}_{ijt}$, $\tilde{\hat{u}}^a_{ijt}$ and $\tilde{\hat{u}}^b_{ijt}$ are obtained by transforming the residual $\hat{u}_{ijt}$ according to M, M^a, and M^b respectively. For model (2.6), as

$$\mathrm{E}(u^2_{ijt}) = \sigma^2_\zeta + \sigma^2_\varepsilon\,, \quad \text{and} \quad \mathrm{E}(\tilde{u}^2_{ijt}) = \frac{N_i-1}{N_1}\sigma^2_\varepsilon\,,$$

with now $\tilde{u}_{ijt} = u_{ijt} - \bar{u}_{.jt}$ (or with $M = I_{N_1N_2T} - (\bar{J}_{N_1}\otimes I_{N_2T})$), the appropriate estimators are simply

$$\hat{\sigma}^2_\varepsilon = \frac{1}{(N_1-1)N_2T}\sum_{ijt}\tilde{\hat{u}}^2_{ijt}\,, \quad \text{and} \quad \hat{\sigma}^2_\zeta = \frac{1}{N_1N_2T}\sum_{ijt}\hat{u}^2_{ijt} - \hat{\sigma}^2_\varepsilon\,.$$

For model (2.8),

$$\begin{aligned} \mathrm{E}(\tilde{u}^2_{ijt}) &= \tfrac{(N_1N_2-1)(T-1)}{N_1N_2T}\sigma^2_\varepsilon\,, \quad \mathrm{E}((\tilde{u}^a_{ijt})^2) = \tfrac{N_1N_2-1}{N_1N_2}(\sigma^2_\mu+\sigma^2_\varepsilon)\,, \quad \text{and} \\ \mathrm{E}((\tilde{u}^b_{ijt})^2) &= \tfrac{T-1}{T}(\sigma^2_\lambda+\sigma^2_\varepsilon)\,, \end{aligned}$$

with $\tilde{u}_{ijt} = u_{ijt} - \bar{u}_{..t} - \bar{u}_{ij.} + \bar{u}_{...}$, and $\tilde{u}^a_{ijt} = u_{ijt} - \bar{u}_{..t}$, and $\tilde{u}^b_{ijt} = u_{ijt} - \bar{u}_{ij.}$ which correspond to

$$\begin{aligned} M &= I_{N_1N_2T} - (\bar{J}_{N_1N_2} \otimes I_T) - (I_{N_1N_2} \otimes \bar{J}_T) + \bar{J}_{N_1N_2T} \\ M^a &= I_{N_1N_2T} - (\bar{J}_{N_1N_2} \otimes I_T) \\ M^b &= I_{N_1N_2T} - (I_{N_1N_2} \otimes \bar{J}_T) \end{aligned}$$

respectively. The estimators for the variance components are

$$\begin{aligned} \hat{\sigma}^2_\varepsilon &= \tfrac{1}{(N_1N_2-1)(T-1)} \textstyle\sum_{ijt} \tilde{\hat{u}}^2_{ijt}\,, \quad \hat{\sigma}^2_\mu = \tfrac{1}{(N_1N_2-1)T} \textstyle\sum_{ijt} (\tilde{\hat{u}}^a_{ijt})^2 - \hat{\sigma}^2_\varepsilon\,, \quad \text{and} \\ \hat{\sigma}^2_\lambda &= \tfrac{1}{N_1N_2(T-1)} \textstyle\sum_{ijt} (\tilde{\hat{u}}^b_{ijt})^2 - \hat{\sigma}^2_\varepsilon\,. \end{aligned}$$

For model (2.10), as

$$\begin{aligned} \mathrm{E}(\tilde{u}^2_{ijt}) &= \tfrac{(N_1N_2-1)T-(N_1-1)-(N_2-1)}{N_1N_2T}\sigma^2_\varepsilon \\ \mathrm{E}((\tilde{u}^a_{ijt})^2) &= \tfrac{(N_1N_2-1)T-(N_2-1)}{N_1N_2T}(\sigma^2_\upsilon + \sigma^2_\varepsilon) \\ \mathrm{E}((\tilde{u}^b_{ijt})^2) &= \tfrac{(N_1N_2-1)T-(N_1-1)}{N_1N_2T}(\sigma^2_\zeta + \sigma^2_\varepsilon) \\ \mathrm{E}((\tilde{u}^c_{ijt})^2) &= \tfrac{N_1N_2T-N_1-N_2+1}{N_1N_2T}(\sigma^2_\mu + \sigma^2_\varepsilon) \end{aligned}$$

with $\tilde{u}_{ijt} = u_{ijt} - \bar{u}_{..t} - \bar{u}_{.j.} - \bar{u}_{i..} + 2\bar{u}_{...}$, $\tilde{u}^a_{ijt} = u_{ijt} - \bar{u}_{..t} - \bar{u}_{.j.} + \bar{u}_{...}$, $\tilde{u}^b_{ijt} = u_{ijt} - \bar{u}_{..t} - \bar{u}_{i..} + \bar{u}_{...}$, and $\tilde{u}^c_{ijt} = u_{ijt} - \bar{u}_{i..} - \bar{u}_{.j.} + \bar{u}_{...}$ which all correspond to the projection matrices

$$\begin{aligned} M &= I_{N_1N_2T} - (\bar{J}_{N_1N_2} \otimes I_T) - (\bar{J}_{N_1} \otimes I_{N_2} \otimes \bar{J}_T) - (I_{N_1} \otimes \bar{J}_{N_2T}) + 2\bar{J}_{N_1N_2T} \\ M^a &= I_{N_1N_2T} - (\bar{J}_{N_1N_2} \otimes I_T) - (\bar{J}_{N_1} \otimes I_{N_2} \otimes \bar{J}_T) + \bar{J}_{N_1N_2T} \\ M^b &= I_{N_1N_2T} - (\bar{J}_{N_1N_2} \otimes I_T) - (I_{N_1} \otimes \bar{J}_{N_2T}) + \bar{J}_{N_1N_2T} \\ M^c &= I_{N_1N_2T} - (\bar{J}_{N_1} \otimes I_{N_2} \otimes \bar{J}_T) - (I_{N_1} \otimes \bar{J}_{N_2T}) + \bar{J}_{N_1N_2T} \end{aligned}$$

respectively. The estimators for the variance components are

$$\begin{aligned} \hat{\sigma}^2_\varepsilon &= \tfrac{1}{(N_1N_2-1)T-(N_1-1)-(N_2-1)} \textstyle\sum_{ijt} \tilde{\hat{u}}^2_{ijt} \\ \hat{\sigma}^2_\upsilon &= \tfrac{1}{(N_1N_2-1)T-(N_2-1)} \textstyle\sum_{ijt} (\tilde{\hat{u}}^a_{ijt})^2 - \hat{\sigma}^2_\varepsilon \\ \hat{\sigma}^2_\zeta &= \tfrac{1}{(N_1N_2-1)T-(N_1-1)} \textstyle\sum_{ijt} (\tilde{\hat{u}}^b_{ijt})^2 - \hat{\sigma}^2_\varepsilon \\ \hat{\sigma}^2_\lambda &= \tfrac{1}{N_1N_2T-N_1-N_2+1} \textstyle\sum_{ijt} (\tilde{\hat{u}}^c_{ijt})^2 - \hat{\sigma}^2_\varepsilon\,. \end{aligned}$$

Lastly, for model (2.12) we get

$$\mathrm{E}(u^2_{ijt}) = \sigma^2_\mu + \sigma^2_\varepsilon\,, \quad \text{and} \quad \mathrm{E}(\tilde{u}^2_{ijt}) = \frac{T-1}{T}\sigma^2_\varepsilon\,,$$

with $\tilde{u}_{ijt} = u_{ijt} - \bar{u}_{ij.}$ (which is the same as a general element of Mu with $M = I_{N_1N_2T} - (I_{N_1N_2} \otimes \bar{J}_T)$). With this, the estimators are

$$\hat{\sigma}^2_\varepsilon = \frac{1}{N_1N_2(T-1)} \sum_{ijt} \tilde{\hat{u}}^2_{ijt}\,, \quad \text{and} \quad \hat{\sigma}^2_\mu = \frac{1}{N_1N_2T} \sum_{ijt} \hat{u}^2_{ijt} - \hat{\sigma}^2_\varepsilon\,.$$

Standard errors are computed accordingly, using $\mathrm{Var}(\hat{\beta}_{FGLS}) = (X'\hat{\Omega}^{-1}X)^{-1}$. In the

Table 2.3 *Sample conditions for the consistency of the FGLS Estimator*

Model	Consistency requirements
(2.2)	$(N_1 \to \infty,\, N_2 \to \infty)$ or $(N_1 \to \infty,\, T \to \infty)$ or $(N_2 \to \infty,\, T \to \infty)$
(2.4)	$(T \to \infty)$ or $(N_1 \to \infty,\, N_2 \to \infty)$
(2.6)	$(N_2 \to \infty)$ or $(T \to \infty)$
(2.8)	$(N_1 \to \infty,\, T \to \infty)$ or $(N_2 \to \infty,\, T \to \infty)$
(2.10)	$(N_1 \to \infty,\, N_2 \to \infty,\, T \to \infty)$
(2.12)	$(N_1 \to \infty)$ or $(N_2 \to \infty)$

limiting cases, the usual normalization factors are needed to obtain finite variances. If, for example N_1 and T are growing, $\sqrt{N_1 T}(\hat{\beta}_{FGLS} - \beta)$ has a normal distribution with zero mean, and $Q^{-1}_{X\Omega X}$ variance, where $Q^{-1}_{X\Omega X} = \text{plim}_{N_1,T\to\infty} \frac{X'\hat{\Omega}^{-1}X}{N_1 T}$ is assumed to be a finite, positive definite matrix. This holds model-wide.

We have no such luck, however, with the OLS estimator. The issue is best illustrated with model (2.12). It can be shown, just as with the usual 2D panel models, $\text{Var}(\hat{\beta}_{OLS}) = (X'X)^{-1}X'\hat{\Omega}X(X'X)^{-1}$ (with $\hat{\Omega}$ being model-specific, but let us assume for now, that it corresponds to (2.13)).

In the asymptotic case, when $N_1, N_2 \to \infty$, $\sqrt{N_1 N_2}(\hat{\beta}_{OLS} - \beta)$ has a normal distribution with finite variance, but this variance grows without bound (at rate $O(T)$) once $T \to \infty$. That is, an extra $1/\sqrt{T}$ normalization factor has to be added to regain a normal distribution with bounded variance. Table 2.4 collects normalization factors needed for a finite $\text{Var}(\hat{\beta}_{OLS})$ for the different models considered. As it is uncommon to normalize with 1, or with expression like $\frac{\sqrt{N_1 N_2}}{\sqrt{A}}$, some insights into the normalizations are given in Appendix A.

Another interesting aspect is revealed by comparing Tables 2.2 and 2.3, that is the consistency requirements for the estimation of the variance components (Table 2.2) and the asymptotic results, when the FGLS converges to the Within estimator (Table 2.3).

As can be seen from Table 2.5, for all models the FGLS is consistent if all N_1, N_2, T go to infinity, but in these cases the (F)GLS estimator converges to the Within one. This is problematic, as some parameters, previously estimable, become suddenly unidentified. In such cases, we have to rely on the OLS estimates, rather than the FGLS. This is generally the case whenever a "+" sign is found in Table 2.5, most significant for models (2.8) and (2.10). For them, the FGLS is only consistent, when it is in fact the Within Estimator, leading to likely severe identification issues. The best case scenarios are indicated with a "−" sign, where the respective asymptotics are already enough for the consistency of the FGLS, but do

Table 2.4 *Normalization factors for the finiteness of* $\hat{\beta}_{OLS}$

Model	(2.2)	(2.4)	(2.6)	(2.8)	(2.10)	(2.12)
$N_1 \to \infty$	1	1	1	1	1	$\sqrt{N_1}$
$N_2 \to \infty$	1	1	$\sqrt{N_2}$	1	1	$\sqrt{N_2}$
$T \to \infty$	1	$\sqrt{T}$	$\sqrt{T}$	1	1	1
$N_1, N_2 \to \infty$	$\frac{\sqrt{N_1 N_2}}{\sqrt{A}}$[a]	$\frac{\sqrt{N_1 N_2}}{\sqrt{A}}$	$\sqrt{N_2}$	1	1	$\sqrt{N_1 N_2}$
$N_1, T \to \infty$	$\frac{\sqrt{N_1 T}}{\sqrt{A}}$	$\sqrt{T}$	$\sqrt{T}$	$\frac{\sqrt{N_1 T}}{\sqrt{A}}$	1	$\sqrt{N_1}$
$N_2, T \to \infty$	$\frac{\sqrt{N_2 T}}{\sqrt{A}}$	$\sqrt{T}$	$\sqrt{N_2 T}$	$\frac{\sqrt{N_2 T}}{\sqrt{A}}$	1	$\sqrt{N_2}$
$N_1, N_2, T \to \infty$	$\frac{\sqrt{N_1 N_2 T}}{\sqrt{A}}$	$\frac{\sqrt{N_1 N_2}}{\sqrt{A}}\sqrt{T}$	$\sqrt{N_2 T}$	$\frac{\sqrt{N_1 N_2 T}}{\sqrt{A}}$	$\frac{\sqrt{N_1 N_2 T}}{\sqrt{A_1 A_2}}$[b]	$\sqrt{N_1 N_2}$

[a] A is the sample size which grows with the highest rate, (N_1, N_2, or T)
[b] A_1, A_2 are the two sample sizes which grow with the highest rates.

Table 2.5 *Asymptotic results when the OLS should be used*

Model	(2.2)	(2.4)	(2.6)	(2.8)	(2.10)	(2.12)
$N_1 \to \infty$			$+$[a]			$-$
$N_2 \to \infty$			$-$			$-$
$T \to \infty$		$-$	$-$			$+$
$N_1, N_2 \to \infty$	$-$	$+$	$+$		$+$	$-$
$N_1, T \to \infty$	$-$	$-$	$+$	$+$	$+$	$+$
$N_2, T \to \infty$	$-$	$-$	$-$	$+$	$+$	$+$
$N_1, N_2, T \to \infty$	$+$	$+$	$+$	$+$	$+$	$+$

[a] Notes: a "$-$" sign indicates that the model is estimated consistently with FGLS, a "$+$" sign indicates that OLS should be used as some parameters are not identified, and a box is left blank if the model can not estimated consistently (under the respective asymptotics).

not yet cause identification problems. Lastly, blank spaces are left in the table if, under the given asymptotic, the FGLS is not consistent. In such cases we can again rely on the consistency of the OLS, but its standard errors are inconsistent, just as with the FGLS.

2.4 Incomplete Data

2.4.1 Structure of the Covariance Matrices

Our analysis has concentrated so far on balanced panels. We know, however, that real life data sets usually have some kind of incompleteness embedded. This can be more visible in the case of higher dimensional panels, where the number of missing observations can be substantial. As known from the analysis of the standard two-way error components models, in this case the estimators of the variance components, and in turn, those of the slope parameters are inconsistent, and further, the spectral decomposition of Ω is inapplicable. Next, we present the covariance matrices of the different models in an incomplete data framework, we show a feasible way to invert them, and then propose a method to estimate the variance components in this general setup.

In our modelling framework, just like in Chapter 1, incompleteness means, that for any (ij) pair of individuals, $t \in T_{ij}$, where T_{ij} index-set is a subset of the general $\{1,\dots,T\}$ index-set of the time periods spanned by the data. Further, let $|T_{ij}|$ denote the cardinality of T_{ij}, i.e., the number of its elements. Note, that for complete (balanced) data, $T_{ij} = \{1,\dots,T\}$, and $|T_{ij}| = T$ for all (ij). We also assume, that for each t there is at least one (ij) pair, for each i, there is at least one (jt) pair, and for each j, there is at least one (it) pair observed. This assumption is almost natural, as it simply requires individuals or time periods with no underlying observation to be dropped from the data set. As the structure of the data now is quite complex, we need to introduce a few new notation and definitions along the way. Formally, let us call n_{it}, n_{jt}, n_i, n_j, and n_t the total number of observations for a given (it), (jt) pair, and for given individuals i, j, and time t, respectively. Further, let us call $\tilde{n}_{ij}$, $\tilde{n}_{it}$, $\tilde{n}_{jt}$ the total number of (ij), (it), and (jt) pairs present in the data. Remember, that in the balanced case, $\tilde{n}_{ij} = N_1N_2$, $\tilde{n}_{it} = N_1T$, and $\tilde{n}_{jt} = N_2T$. It would make sense to define similarly $\tilde{n}_i$, $\tilde{n}_j$, and $\tilde{n}_t$, however, we can assume, without the loss of generality, that there are still N_1 i, N_2 j, individuals, and T total time periods in the data (of course, there are holes in it).

For the all-encompassing model (2.2), u_{ijt} can be stacked into vector u. Remember, that in the complete case it is

$$\begin{aligned} u &= (I_{N_1} \otimes I_{N_2} \otimes \iota_T)\mu + (I_{N_1} \otimes \iota_{N_2} \otimes I_T)\upsilon + (\iota_{N_1} \otimes I_{N_2} \otimes I_T)\zeta + I_{N_1N_2T}\varepsilon \\ &= D_1\mu + D_2\upsilon + D_3\zeta + \varepsilon\,, \end{aligned}$$

with μ, υ, ζ, ε begin the stacked vectors of μ_{ij}, υ_{it}, ζ_{jt}, and ε_{ijt}, of respective lengths N_1N_2, N_1T, N_2T, N_1N_2T, and ι is the column of ones with size on the index. The covariance matrix can then be represented by

$$E(uu') = \Omega = D_1D_1'\sigma_\mu^2 + D_2D_2'\sigma_\upsilon^2 + D_3D_3'\sigma_\zeta^2 + I\sigma_\varepsilon^2\,,$$

which is identical to (2.3). However, in the case of missing data, we have to modify the underlying D_k dummy matrices to reflect the unbalanced nature of the data. For every (ij) pair, let V_{ij} denote the size $(|T_{ij}| \times T)$ matrix, which we obtain from the $(T \times T)$ identity matrix by deleting rows corresponding to missing observations.[3] With this, the incomplete D_k dummies are

$$\begin{aligned}
D_1 &= \mathrm{diag}\{V_{11}\iota_T, V_{12}\iota_T, \ldots, V_{N_1N_2}\iota_T\} \quad \text{of size} \quad \left(\textstyle\sum_{ij}|T_{ij}| \times \tilde{n}_{ij}\right), \\
D_2 &= \mathrm{diag}\left\{(V'_{11}, V'_{12}, \ldots, V'_{1N_2})', \ldots, (V'_{N_11}, V'_{N_12}, \ldots, V'_{N_1N_2})'\right\} \\
&\qquad\qquad\qquad\qquad\qquad\qquad \text{of size} \quad \left(\textstyle\sum_{ij}|T_{ij}| \times \tilde{n}_{it}\right) \\
D_3 &= \left(\mathrm{diag}\{V'_{11}, V'_{12}, \ldots, V'_{1N_2}\}', \ldots, \mathrm{diag}\{V'_{N_11}, V'_{N_12}, \ldots, V'_{N_1N_2}\}'\right)' \\
&\qquad\qquad\qquad\qquad\qquad\qquad \text{of size} \quad \left(\textstyle\sum_{ij}|T_{ij}| \times \tilde{n}_{jt}\right).
\end{aligned}$$

These then can be used to construct the covariance matrix as

$$\Omega = E(uu') = I_{\sum_{ij}|T_{ij}|}\sigma^2_\varepsilon + D_1D'_1\sigma^2_\mu + D_2D'_2\sigma^2_\upsilon + D_3D'_3\sigma^2_\zeta$$

of size $\left(\sum_{ij}|T_{ij}| \times \sum_{ij}|T_{ij}|\right)$. If the data is complete, the above covariance structure in fact gives back (2.3). The job is the same for other models. For models (2.4) and (2.6),

$$u = D_2\upsilon + D_3\zeta + \varepsilon$$

and

$$u = D_3\zeta + \varepsilon$$

respectively, with the incompleteness adjusted D_2 and D_3 defined above, giving in turn

$$\Omega = I_{\sum_{ij}|T_{ij}|}\sigma^2_\varepsilon + D_2D'_2\sigma^2_\upsilon + D_3D'_3\sigma^2_\zeta$$

for model (2.4), and

$$\Omega = I_{\sum_{ij}|T_{ij}|}\sigma^2_\varepsilon + D_3D'_3\sigma^2_\zeta$$

for model (2.6). Again, if the panel were in fact complete, we would get back (2.5) and (2.7). The incomplete data covariance matrix of model (2.8) is

$$\Omega = I_{\sum_{ij}|T_{ij}|}\sigma^2_\varepsilon + D_1D'_1\sigma^2_\mu + D_4D'_4\sigma^2_\lambda,$$

with

$$D_4 = (V'_{11}, V'_{12}, \ldots, V'_{N_1N_2})' \quad \text{of size} \quad \left(\sum_{ij}|T_{ij}| \times T\right).$$

The covariance matrix for model (2.10) is

$$\Omega = I_{\sum_{ij}|T_{ij}|}\sigma^2_\varepsilon + D_5D'_5\sigma^2_\upsilon + D_6D'_6\sigma^2_\zeta + D_4D'_4\sigma^2_\lambda,$$

[3] If, for example, $t = 1, 4, 10$ are missing for some (ij), we delete rows 1, 4, and 10 from I_T to get V_{ij}.

where

$$
\begin{aligned}
D_5 &= \mathrm{diag}\left\{(V_{11}'\iota_T, V_{12}'\iota_T, \ldots, V_{1N_2}'\iota_T)', \ldots, (V_{N_1 1}'\iota_T, V_{N_1 2}'\iota_T, \ldots, V_{N_1 N_2}'\iota_T)'\right\} \\
D_6 &= \big(\mathrm{diag}\{V_{11}'\iota_T, V_{12}'\iota_T, \ldots, V_{1N_2}'\iota_T\}', \ldots \times \\
&\quad \times \ldots, \mathrm{diag}\{V_{N_1 1}'\iota_T, V_{N_1 2}'\iota_T, \ldots, V_{N_1 N_2}'\iota_T\}'\big)' .
\end{aligned}
$$

of sizes $(\sum_{ij}|T_{ij}| \times N_1)$, and $(\sum_{ij}|T_{ij}| \times N_2)$. Lastly, for model (2.12) we simply get

$$\Omega = I_{\sum_{ij}|T_{ij}|}\sigma_\varepsilon^2 + D_1 D_1' \sigma_\mu^2 .$$

An important practical difficulty is that the spectral decomposition of the covariance matrices introduced in Section 2.3 are no longer valid, so the inversion of Ω for very large data sets can be forbidding. To go around this problem, let us construct the *quasi-spectral decomposition* of the incomplete data covariance matrices, which is simply done by leaving out the missing rows from the appropriate B. Specifically, let us call B^* the incompleteness-adjusted versions of any B, which we get by removing the rows corresponding to the missing observations. For example, the spectral decomposition (2.14) for the all-encompassing model reads as

$$
\begin{aligned}
\Omega^* = \; & \sigma_\varepsilon^2 B_{ijt}^* + (\sigma_\varepsilon^2 + T\sigma_\mu^2) B_{ij.}^* + (\sigma_\varepsilon^2 + N_2\sigma_\upsilon^2) B_{i.t}^* + (\sigma_\varepsilon^2 + N_1\sigma_\zeta^2) B_{.jt}^* \\
& + (\sigma_\varepsilon^2 + T\sigma_\mu^2 + N_2\sigma_\upsilon^2) B_{i..}^* + (\sigma_\varepsilon^2 + T\sigma_\mu^2 + N_1\sigma_\zeta^2) B_{.j.}^* \\
& + (\sigma_\varepsilon^2 + N_2\sigma_\upsilon^2 + N_1\sigma_\zeta^2) B_{..t}^* + (\sigma_\varepsilon^2 + T\sigma_\mu^2 + N_2\sigma_\upsilon^2 + N_1\sigma_\zeta^2) B_{...}^* ,
\end{aligned}
$$

where now all B^* have number of rows equal to $\sum_{ij}|T_{ij}|$. Of course, this is not a correct spectral decomposition of Ω, but helps to define the following conjecture.[4] Namely, when the number of missing observations relative to the total number of observations is small, the inverse of Ω based on the quasi-spectral decomposition of it, Ω^{*-1}, approximate arbitrarily well Ω^{-1}. More precisely, if $[N_1N_2T - \sum_i\sum_j|T_{ij}|]/[N_1N_2T] \to 0$, then $(\Omega^{-1} - \Omega^{*-1}) \to 0$. This means that in large data sets, when the number of missing observation is small relative to the total number of observations, Ω^{*-1} can safely be used in the GLS estimator instead of Ω^{-1}. Let us give an example. Multi-dimensional panel data are often used to deal with trade (gravity) models. In these cases, however, when country i trade with country j, there are no (ii) (or (jj)) observations, there is no self-trade. Then the total number of observations is $N^2T - NT$ with NT being the number of missing observations due to no self-trade. Given that $[N^2T - (N^2T - NT)]/N^2T \to 0$ as the sample size increases, the quasi-spectral decomposition can be used in large data.

[4] This can be demonstrated by simulation.

2.4.2 The Inverse of the Covariance Matrices

The solution proposed above, however, suffers from two potential drawbacks. First, the inverse, though reached at very low cost, may not be accurate enough, and second, when the "holes" in the data are substantial this method cannot be used. These reasons spur us to derive the analytically correct inverse of the covariance matrices at the lowest possible cost. To do that, we have to reach back to the comprehensive incomplete data analysis carried out by Baltagi and Chang (1994), and later Baltagi et al. (2002) for one- and two-way error component models, Baltagi et al. (2001) for nested three-way models, and also, we have to generalize the results of Wansbeek and Kapteyn (1989) (in a slightly different manner though, than seen in Davis, 2002). This leads us, for model (2.2), to

$$\sigma_\varepsilon^2\Omega^{-1} = P^b - P^b D_3 (R^c)^{-1} D_3' P^b \tag{2.17}$$

where P^b and R^c are obtained in steps:

$$\begin{aligned} R^c &= D_3' P^b D_3 + \tfrac{\sigma_\varepsilon^2}{\sigma_\zeta^2} I, \quad P^b = P^a - P^a D_2 (R^b)^{-1} D_2' P^a, \\ R^b &= D_2' P^a D_2 + \tfrac{\sigma_\varepsilon^2}{\sigma_\upsilon^2} I, \quad P^a = I - D_1 (R^a)^{-1} D_1', \quad \text{and} \\ R^a &= D_1' D_1 + \tfrac{\sigma_\varepsilon^2}{\sigma_\mu^2} I, \end{aligned}$$

where D_1, D_2, D_3 are the incompleteness-adjusted dummy variable matrices, and are used to construct the P and R matrices sequentially: first, construct R^a to get P^a, then construct R^b to get P^b, and finally, construct R^c to get P^c. Proof of (2.17) can be found in Appendix B. Note, that to get the inverse, we have to invert $\min\{N_1T; N_2T; N_1N_2\}$ matrices. The quasi-scalar form of (2.17) (which corresponds to the incomplete data version of transformation (2.15)) is

$$y_{ijt} - \left(1 - \sqrt{\frac{\sigma_\varepsilon^2}{|T_{ij}|\sigma_\mu^2 + \sigma_\varepsilon^2}}\right) \frac{1}{|T_{ij}|} \sum_t y_{ijt} - \omega_{ijt}^a - \omega_{ijt}^b,$$

with

$$\omega_{ijt}^a = \chi_{ijt}^a \cdot \psi^a, \quad \text{and} \quad \omega_{ijt}^b = \chi_{ijt}^b \cdot \psi^b,$$

where χ_{ijt}^a is the row corresponding to observation (ijt) from $P^a D_2$, ψ^a is the column vector $(R^b)^{-1} D_2' P^a y$, ω_{ijt}^b is the row from matrix $P^b D_3$ corresponding to observation (ijt), and finally, ψ^b is the column vector $(R^c)^{-1} D_3' P^b y$.

For the other models, the job is essentially the same, only the number of steps in obtaining the inverse is smaller (as the number of different random effects decreases). For model (2.4), it is, with appropriately redefining P and R,

$$\sigma_\varepsilon^2\Omega^{-1} = P^a - P^a D_3 (R^b)^{-1} D_3' P^a, \tag{2.18}$$

where now

$$R^b = D_3'P^aD_3 + \frac{\sigma_\varepsilon^2}{\sigma_\zeta^2}I, \quad P^a = I - D_2(R^a)^{-1}D_2' \quad \text{and} \quad R^a = D_2'D_2 + \frac{\sigma_\varepsilon^2}{\sigma_\upsilon^2}I,$$

with the largest matrix to inverted now of size $\min\{N_1T; N_2T\}$. For model (2.6), it is even more simple,

$$\sigma_\varepsilon^2\Omega^{-1} = I - D_3(R^a)^{-1}D_3' \quad \text{with} \quad R^a = D_3'D_3 + \frac{\sigma_\varepsilon^2}{\sigma_\zeta^2}I, \tag{2.19}$$

defining the scalar transformation

$$\tilde{y}_{ijt} = y_{ijt} - \left(1 - \sqrt{\frac{\sigma_\varepsilon^2}{n_{jt}\sigma_\zeta^2 + \sigma_\varepsilon^2}}\right)\frac{1}{n_{jt}}\sum_i y_{ijt},$$

with n_{jt} being the number of observations for a given (jt) pair. For model (2.8), the inverse is

$$\sigma_\varepsilon^2\Omega^{-1} = P^a - P^aD_4(R^b)^{-1}D_4'P^a \tag{2.20}$$

where

$$R^b = D_4'P^aD_4 + \frac{\sigma_\varepsilon^2}{\sigma_\lambda^2}I, \quad P^a = I - D_1(R^a)^{-1}D_1' \quad \text{and} \quad R^a = D_1'D_1 + \frac{\sigma_\varepsilon^2}{\sigma_\mu^2}I.$$

and we have to invert a $\min\{N_1N_2; T\}$ sized matrix. For model (2.10), the inverse is again the result of a three-step procedure:

$$\sigma_\varepsilon^2\Omega^{-1} = P^b - P^bD_4(R^c)^{-1}D_4'P^b, \tag{2.21}$$

where

$$\begin{aligned} R^c &= D_4'P^bD_4 + \tfrac{\sigma_\varepsilon^2}{\sigma_\lambda^2}I,\ P^b = P^a - P^aD_6(R^b)^{-1}D_6'P^a, \\ R^b &= D_6'P^aD_6 + \tfrac{\sigma_\varepsilon^2}{\sigma_\zeta^2}I,\ P^a = I - D_5(R^a)^{-1}D_5', \text{ and } \quad R^a = D_5'D_5 + \tfrac{\sigma_\varepsilon^2}{\sigma_\upsilon^2}I, \end{aligned}$$

(with inverting a matrix of size $\min\{N_1; N_2; T\}$) and finally, the inverse of the simplest model is

$$\sigma_\varepsilon^2\Omega^{-1} = I - D_1(R^a)^{-1}D_1' \quad \text{with} \quad R^a = D_1'D_1 + \frac{\sigma_\varepsilon^2}{\sigma_\mu^2}I, \tag{2.22}$$

defining the scalar transformation

$$\tilde{y}_{ijt} = y_{ijt} - \left(1 - \sqrt{\frac{\sigma_\varepsilon^2}{|T_{ij}|\sigma_\mu^2 + \sigma_\varepsilon^2}}\right)\frac{1}{|T_{ij}|}\sum_t y_{ijt}$$

on a typical y_{ijt} variable.

2.4.3 Estimation of the Variance Components

Let us proceed to the estimation of the variance components. The estimators used for complete data are no longer applicable here, as for example, transformation (2.16) does not eliminate μ_{ij}, υ_{it}, and ζ_{jt} from the composite disturbance term $u_{ijt} = \mu_{ij} + \upsilon_{it} + \zeta_{jt} + \varepsilon_{ijt}$, when the data is incomplete. This problem can be tackled in two ways. We can derive incompleteness-robust alternative to (2.16), i.e., a transformation which clears the non-idiosyncratic random effects from u_{ijt}, in the case of incomplete data (see Balazsi et al., 2015). The problem is that most of these transformations involve the manipulation of large matrices resulting in heavy computational burden. To avoid this we propose simple linear transformations, which on the one hand, are robust to incomplete data, and on the other hand, identify the variance components. Let us see, how this works for model (2.2). As before

$$\mathrm{E}(u_{ijt}^2) = \sigma_\mu^2 + \sigma_\upsilon^2 + \sigma_\zeta^2 + \sigma_\varepsilon^2\,, \tag{2.23}$$

but now, let us define

$$\begin{aligned} \tilde{u}_{ijt}^a &= u_{ijt} - \tfrac{1}{|T_{ij}|}\textstyle\sum_t u_{ijt}\,, \quad \tilde{u}_{ijt}^b = u_{ijt} - \tfrac{1}{n_{it}}\textstyle\sum_j u_{ijt}\,, \quad \text{and} \\ \tilde{u}_{ijt}^c &= u_{ijt} - \tfrac{1}{n_{jt}}\textstyle\sum_i u_{ijt}\,. \end{aligned}$$

It can be seen that

$$\begin{aligned} \mathrm{E}((\tilde{u}_{ijt}^a)^2) &= \tfrac{|T_{ij}|-1}{|T_{ij}|}(\sigma_\upsilon^2 + \sigma_\zeta^2 + \sigma_\varepsilon^2)\,, \quad \mathrm{E}((\tilde{u}_{ijt}^b)^2) = \tfrac{n_{it}-1}{n_{it}}(\sigma_\mu^2 + \sigma_\zeta^2 + \sigma_\varepsilon^2)\,, \\ \text{and} \quad & \mathrm{E}((\tilde{u}_{ijt}^c)^2) = \tfrac{n_{jt}-1}{n_{jt}}(\sigma_\mu^2 + \sigma_\upsilon^2 + \sigma_\varepsilon^2)\,. \end{aligned} \tag{2.24}$$

Combining (2.23) with (2.24) identifies all four variance components. The appropriate estimators are then

$$\begin{aligned} \hat{\sigma}_\mu^2 &= \tfrac{1}{\sum_{ij}|T_{ij}|}\textstyle\sum_{ijt}\hat{u}_{ijt}^2 - \tfrac{1}{\tilde{n}_{ij}}\textstyle\sum_{ij}\tfrac{1}{|T_{ij}|-1}\textstyle\sum_t(\tilde{\hat{u}}_{ijt}^a)^2 \\ \hat{\sigma}_\upsilon^2 &= \tfrac{1}{\sum_{ij}|T_{ij}|}\textstyle\sum_{ijt}\hat{u}_{ijt}^2 - \tfrac{1}{\tilde{n}_{it}}\textstyle\sum_{it}\tfrac{1}{n_{it}-1}\textstyle\sum_j(\tilde{\hat{u}}_{ijt}^b)^2 \\ \hat{\sigma}_\zeta^2 &= \tfrac{1}{\sum_{ij}|T_{ij}|}\textstyle\sum_{ijt}\hat{u}_{ijt}^2 - \tfrac{1}{\tilde{n}_{jt}}\textstyle\sum_{jt}\tfrac{1}{n_{jt}-1}\textstyle\sum_i(\tilde{\hat{u}}_{ijt}^c)^2 \\ \hat{\sigma}_\varepsilon^2 &= \tfrac{1}{\sum_{ij}|T_{ij}|}\textstyle\sum_{ijt}\hat{u}_{ijt}^2 - \hat{\sigma}_\mu^2 - \hat{\sigma}_\upsilon^2 - \hat{\sigma}_\zeta^2\,, \end{aligned} \tag{2.25}$$

where $\hat{u}_{ijt}$ are the OLS residuals, and $\tilde{\hat{u}}_{ijt}^k$ are its transformations ($k = a,b,c$), where $\tilde{n}_{ij}$, $\tilde{n}_{it}$, and $\tilde{n}_{jt}$ denote the total number of observations for the (ij), (it), and (jt) pairs respectively in the data.

The estimation strategy of the variance components is exactly the same for all the other models. Let us keep for now the definitions of $\tilde{u}_{ijt}^b$, and $\tilde{u}_{ijt}^c$. For model

(2.4), with $u_{ijt} = \upsilon_{it} + \zeta_{jt} + \varepsilon_{ijt}$, the estimators read as

$$\begin{aligned} \hat{\sigma}^2_\upsilon &= \tfrac{1}{\sum_{ij}|T_{ij}|}\sum_{ijt}\hat{u}^2_{ijt} - \tfrac{1}{\tilde{n}_{it}}\sum_{it}\tfrac{1}{n_{it}-1}\sum_j(\tilde{\hat{u}}^b_{ijt})^2 \\ \hat{\sigma}^2_\zeta &= \tfrac{1}{\sum_{ij}|T_{ij}|}\sum_{ijt}\hat{u}^2_{ijt} - \tfrac{1}{\tilde{n}_{jt}}\sum_{jt}\tfrac{1}{n_{jt}-1}\sum_i(\tilde{\hat{u}}^c_{ijt})^2 \\ \hat{\sigma}^2_\varepsilon &= \tfrac{1}{\sum_{ij}|T_{ij}|}\sum_{ijt}\hat{u}^2_{ijt} - \hat{\sigma}^2_\upsilon - \hat{\sigma}^2_\zeta\,, \end{aligned} \tag{2.26}$$

whereas for model (2.6), with $u_{ijt} = \zeta_{jt} + \varepsilon_{ijt}$, they are

$$\begin{aligned} \hat{\sigma}^2_\zeta &= \tfrac{1}{\sum_{ij}|T_{ij}|}\sum_{ijt}\hat{u}^2_{ijt} - \tfrac{1}{\tilde{n}_{jt}}\sum_{jt}\tfrac{1}{n_{jt}-1}\sum_i(\tilde{\hat{u}}^c_{ijt})^2 \\ \hat{\sigma}^2_\varepsilon &= \tfrac{1}{\sum_{ij}|T_{ij}|}\sum_{ijt}\hat{u}^2_{ijt} - \hat{\sigma}^2_\zeta\,, \end{aligned} \tag{2.27}$$

Note, that these latter two estimators can be obtained from (2.25), by assuming $\hat{\sigma}^2_\mu = 0$ for model (2.4), and $\hat{\sigma}^2_\mu = \hat{\sigma}^2_\upsilon = 0$ for model (2.6).

For model (2.8), let us redefine the $\tilde{u}^k_{ijt}$-s, as

$$\tilde{u}^a_{ijt} = u_{ijt} - \frac{1}{|T_{ij}|}\sum_t u_{ijt}\,, \quad \text{and} \quad \tilde{u}^b_{ijt} = u_{ijt} - \frac{1}{n_t}\sum_{ij} u_{ijt}\,,$$

with n_t being the number of individual pairs at time t. With $u_{ijt} = \mu_{ij} + \lambda_t + \varepsilon_{ijt}$,

$$\begin{aligned} &\mathrm{E}((\tilde{u}^a_{ijt})^2) = \tfrac{|T_{ij}|-1}{|T_{ij}|}\quad(\sigma^2_\lambda + \sigma^2_\varepsilon)\,, \quad \mathrm{E}((\tilde{u}^b_{ijt})^2) = \tfrac{n_t-1}{n_t}(\sigma^2_\mu + \sigma^2_\varepsilon)\,, \\ &\qquad\qquad\qquad\quad \text{and} \quad \mathrm{E}(u^2_{ijt}) = \sigma^2_\mu + \sigma^2_\lambda + \sigma^2_\varepsilon\,. \end{aligned}$$

From this set of identifying equations, the estimators are simply

$$\begin{aligned} \hat{\sigma}^2_\mu &= \tfrac{1}{\sum_{ij}|T_{ij}|}\sum_{ijt}\hat{u}^2_{ijt} - \tfrac{1}{\tilde{n}_{ij}}\sum_{ij}\tfrac{1}{|T_{ij}|-1}\sum_t(\tilde{\hat{u}}^a_{ijt})^2 \\ \hat{\sigma}^2_\lambda &= \tfrac{1}{\sum_{ij}|T_{ij}|}\sum_{ijt}\hat{u}^2_{ijt} - \tfrac{1}{T}\sum_t\tfrac{1}{n_t-1}\sum_{ij}(\tilde{\hat{u}}^b_{ijt})^2 \\ \hat{\sigma}^2_\varepsilon &= \tfrac{1}{\sum_{ij}|T_{ij}|}\sum_{ijt}\hat{u}^2_{ijt} - \hat{\sigma}^2_\mu - \hat{\sigma}^2_\lambda\,. \end{aligned} \tag{2.28}$$

For model (2.12), with $u_{ijt} = \mu_{ij} + \varepsilon_{ijt}$, keeping the definition of $\tilde{u}^a_{ijt}$,

$$\begin{aligned} \hat{\sigma}^2_\mu &= \tfrac{1}{\sum_{ij}|T_{ij}|}\sum_{ijt}\hat{u}^2_{ijt} - \tfrac{1}{\tilde{n}_{ij}}\sum_{ij}\tfrac{1}{|T_{ij}|-1}\sum_t(\tilde{\hat{u}}^a_{ijt})^2 \\ \hat{\sigma}^2_\varepsilon &= \tfrac{1}{\sum_{ij}|T_{ij}|}\sum_{ijt}\hat{u}^2_{ijt} - \hat{\sigma}^2_\mu\,. \end{aligned} \tag{2.29}$$

Finally, for model (2.10), as now $u_{ijt} = \upsilon_i + \zeta_j + \lambda_t + \varepsilon_{ijt}$, using

$$\tilde{u}^a_{ijt} = u_{ijt} - \frac{1}{n_i}\sum_{jt} u_{ijt}\,, \quad \tilde{u}^b_{ijt} = u_{ijt} - \frac{1}{n_j}\sum_{it} u_{ijt}\,, \quad \tilde{u}^c_{ijt} = u_{ijt} - \frac{1}{n_t}\sum_{ij} u_{ijt}\,,$$

with n_i and n_j being the number of observation-pairs for individual i, and j, respectively, the identifying equations are

$$\begin{aligned} \mathrm{E}((\tilde{u}^a_{ijt})^2) &= \tfrac{n_i-1}{n_i}(\sigma^2_\zeta + \sigma^2_\lambda + \sigma^2_\varepsilon)\,, \quad \mathrm{E}((\tilde{u}^b_{ijt})^2) = \tfrac{n_j-1}{n_j}(\sigma^2_\upsilon + \sigma^2_\lambda + \sigma^2_\varepsilon)\,, \\ \mathrm{E}((\tilde{u}^c_{ijt})^2) &= \tfrac{n_t-1}{n_t}(\sigma^2_\upsilon + \sigma^2_\zeta + \sigma^2_\varepsilon)\,, \quad \text{and} \quad \mathrm{E}(u^2_{ijt}) = \sigma^2_\upsilon + \sigma^2_\zeta + \sigma^2_\lambda + \sigma^2_\varepsilon\,, \end{aligned}$$

in turn leading to

$$
\begin{aligned}
\hat{\sigma}_{\upsilon}^2 &= \tfrac{1}{\sum_{ij}|T_{ij}|}\textstyle\sum_{ijt}\hat{u}_{ijt}^2 - \tfrac{1}{\tilde{n}_i}\sum_{ij}\tfrac{1}{n_i-1}\sum_{jt}(\tilde{\hat{u}}_{ijt}^a)^2 \\
\hat{\sigma}_{\upsilon}^2 &= \tfrac{1}{\sum_{ij}|T_{ij}|}\textstyle\sum_{ijt}\hat{u}_{ijt}^2 - \tfrac{1}{\tilde{n}_j}\sum_{it}\tfrac{1}{n_j-1}\sum_{it}(\tilde{\hat{u}}_{ijt}^b)^2 \\
\hat{\sigma}_{\zeta}^2 &= \tfrac{1}{\sum_{ij}|T_{ij}|}\textstyle\sum_{ijt}\hat{u}_{ijt}^2 - \tfrac{1}{T}\sum_{jt}\tfrac{1}{n_t-1}\sum_{ij}(\tilde{\hat{u}}_{ijt}^c)^2 \\
\hat{\sigma}_{\varepsilon}^2 &= \tfrac{1}{\sum_{ij}|T_{ij}|}\textstyle\sum_{ijt}\hat{u}_{ijt}^2 - \hat{\sigma}_{\upsilon}^2 - \hat{\sigma}_{\zeta}^2 - \hat{\sigma}_{\lambda}^2 \,.
\end{aligned}
\tag{2.30}
$$

2.5 Endogenous Regressors

When the unobserved effects are correlated with the regressors, the (F)GLS estimators proposed in Sections 2.2-2.3 are biased and inconsistent. An alternative approach is to seek an appropriate transformation to eliminate the unobserved heterogeneity, like we did in Chapter 1. While these Within estimators are generally consistent under various assumptions, they have two major shortcomings. First, these estimators eliminate all time-invariant and individual-invariant variables from the model. This includes the unobserved heterogeneity as well as some of the explanatory variables in x_{ijt}. As a result, any parameters associated with the time-invariant or individual-invariant variables cannot be estimated using these estimators. This can be a potentially very important issue if these variables are key policy ones, for example. Second, these estimators eliminate unobserved heterogeneities by computing the deviation of each variable from different means, such as group means (averages over time) and overall means (average over time and individual). This approach often leads to information lost and this is reflected by the fact that these estimators, while consistent, are generally not efficient.

From a practical perspective, the efficiency issue is generally a lesser concern as multi-dimensional datasets have an overwhelmingly large number of observations over all (most) indexes. Thus, the ease of computation of these estimators often outweighs the efficiency benefit from the more computationally complicated, but more efficient, estimators. The identifiability of parameters associated with time-invariant and individual-invariant variables are often the more serious issue. For example, standard Gravity models of trade employ distance, the GDP of the export and the import countries as key regressors (see, for example, Harris et al., 2002; Bun and Klaassen, 2007). Distance is clearly be time-invariant; GDP of the exporting country is invariant with respect to all import countries; and likewise, the GDP of the importing country is invariant with respect to all export countries, thus they are both individual-invariant. Under the assumption that $u_{ijt} = \mu_{ij} + \upsilon_{it} + \zeta_{jt}$, the standard Within-type approach will eliminate all t-invariant, j-invariant and i-invariant explanatory variables and thus it is impossible to estimate their effects.

This section, based on Balazsi et al. (2017) thus proposes a consistent way to estimate all model parameters under the assumption that the random effects are possibly correlated with some of the regressors. To do this we use the 2D results of Hausman and Taylor (1981) (hereafter HT), and extend their approach to our multi-dimensional random effects panel data models.

2.5.1 The Hausman-Taylor-like Instrumental Variable Estimator

Let's consider model (2.10) first:

$$y_{ijt} = \underline{x}'_{ijt}\beta + \upsilon_i + \zeta_j + \lambda_t + \varepsilon_{ijt} = \underline{x}'_{ijt}\beta + u_{ijt} \tag{2.31}$$

and show in details how to extend the HT approach. To make the subsequent analysis more transparent, x'_{ijt} in (2.10) is re-written as $\underline{x}_{ijt}$, and x'_{ijt} is assigned with a slightly different meaning. This change in notation is only used in Section 2.5 and in Sections 2.6.2 and 2.6.3, where tests concerning endogenous regressors are detailed.

Sources of Endogeneity

In addition to the specification as stated in equation (2.31), we consider the random effects in u_{ijt} to be well-behaved, as outlined in Section 2.2, however, we do not impose any correlation restriction between the regressors and the unobserved heterogeneity. In order to accommodate the most general correlation structure between regressors and any of these unobserved heterogeneity terms. For ease of exposition, we divide the explanatory variables according to their index properties as follows:

$$\underline{x}'_{ijt} = (\bar{x}'_{ijt},\, x'_i,\, x'_j,\, x'_t).$$

Note that $\bar{x}'_{ijt} = (x'_{ijt},\, x'_{it},\, x'_{jt},\, x'_{ij})$, that is, it includes all regressors that vary over at least two indices. This particular partition highlights the fact that any parameters associated with variables that vary over more than one index can be identified and estimated from Within-type estimation, unlike the parameters of x'_i, x'_j, or x'_t, as these variables would be eliminated during the Within transformations.

Without loss in generality, we partition each group of variables as follows

$$\begin{aligned}
\bar{x}'_{ijt} &= (x'_{1ijt},\, x'_{2ijt},\, x'_{3ijt},\, x'_{4ijt},\, x'_{5ijt},\, x'_{6ijt},\, x'_{7ijt},\, x'_{8ijt}) \\
x'_i &= (x'_{1i},\, x'_{2i}) \\
x'_j &= (x'_{1j},\, x'_{2j}) \\
x'_t &= (x'_{1t},\, x'_{2t})
\end{aligned}$$

where each partition is assumed to have a different correlation structure with the unobserved heterogeneities. These are summarized in Table 2.6. The subsequent analysis does not explicitly impose any further assumptions on the correlations between the regressors; multicollinearity does not generally violate the feasibility and consistency of the estimators with the exception of perfect collinearity, which case will be excluded from the rank condition.

Table 2.6 *Sources of endogeneity on the level of partitions of the regressors for model (2.31)*

Correlated with	Partition			
None	x'_{1ijt}	x'_{1i}	x'_{1j}	x'_{1t}
υ_i	x'_{2ijt}	x'_{2i}		
ζ_j	x'_{3ijt}		x'_{2j}	
λ_t	x'_{4ijt}			x'_{2t}
υ_i, ζ_j	x'_{5ijt}			
υ_i, λ_t	x'_{6ijt}			
ζ_j, λ_t	x'_{7ijt}			
υ_i, ζ_j, λ_t	x'_{8ijt}			

Following HT and Wyhowski (1994), the basic idea is to construct a set of internal instruments by using the group means of variables in the partition $\bar{x}_{ijt}$. The approach can be outlined as follows. First, the parameters associated with $\bar{x}_{ijt}$ are estimated via standard Within estimator, which is consistent. Second, the group means of $\bar{x}_{ijt}$ are used to construct instruments for the endogenous partitions in x'_i, x'_j and x'_t.

The Hausman-Taylor Estimator

Using matrix notation for the partitions of the regressors, we define $X^{(1)}$ as the stacked matrix version of $\bar{x}'_{ijt}$, namely

$$X^{(1)} = \left(X_1^{(1)},\, X_2^{(1)},\, X_3^{(1)},\, X_4^{(1)},\, X_5^{(1)},\, X_6^{(1)},\, X_7^{(1)},\, X_8^{(1)}\right),$$

with respective columns $k_l^{(1)}$, $l = 1\ldots 8$. Similarly, we define $X^{(2)}, X^{(3)}, X^{(4)}$ for x'_i, x'_j and x'_t, respectively. The number of columns in each partition $X_l^{(m)}$ is $k_l^{(m)}$ with the associated parameter vector being $\beta_l^{(m)}$.

Consider the following transformation on $X^{(1)}$

$$\begin{aligned} H_1 = \;& \big(I - (\bar{J}_{N_1} \otimes \bar{J}_{N_2} \otimes \bar{J}_T) - (Q_{N_1} \otimes \bar{J}_{N_2} \otimes \bar{J}_T) \\ & - (\bar{J}_{N_1} \otimes Q_{N_2} \otimes \bar{J}_T) - (\bar{J}_{N_1} \otimes \bar{J}_{N_2} \otimes Q_T)\big) X^{(1)} \end{aligned}$$

which can be used as instrument for $X^{(1)}$ in order to obtain consistent estimate of $\beta_l^{(1)}$ for $l = 1, \ldots, 8$. It is clear that H_1 is correlated with $X^{(1)}$, but the transformation also removes the unobserved heterogeneities, namely it removes υ_i, ζ_j and λ_t when the transformation applies to (2.31). While this transformation looks complicated, it can be interpreted quite easily. Essentially, it is equivalent to $(x_{ijt} - \bar{x}_{i.} - \bar{x}_{.j.} - \bar{x}_{..t} + 2\bar{x}_{...})$.

Intuitively, to instrument x'_{2i}, the endogenous part of x'_i, we can use all regressors from x'_{ijt} which are uncorrelated with υ_i but this will work only if these instruments are also not correlated with ζ_j and λ_t. One way to ensure this is to remove j- and t- variations from the instruments. As a result, the instrument set for x'_{2i} is simply

$$H_2 = (Q_{N_1} \otimes \bar{J}_{N_2T}) \cdot \left(X_1^{(1)},\, X_3^{(1)},\, X_4^{(1)},\, X_7^{(1)},\, X_1^{(2)}\right)$$

where the exogenous $X_1^{(2)}$ does not require any instrument.

Also note that H_2 is a matrix containing the j,t-group means minus the overall sample mean of the included variables. That is, it only contains variation over i and therefore, this transform will remove ζ_j and λ_t. It is also important to note that H_2 only comprises of variables that are uncorrelated with υ_i, which underpins its usage as an instrument for $X_2^{(2)}$, the individual-specific regressors correlated with υ_i.

Following the similar arguments, the largest variable set uncorrelated with ζ_j is

$$H_3 = (\bar{J}_{N_1} \otimes Q_{N_2} \otimes \bar{J}_T) \cdot \left(X_1^{(1)},\, X_2^{(1)},\, X_4^{(1)},\, X_6^{(1)},\, X_1^{(3)}\right),$$

and finally, the largest set uncorrelated with λ_t is

$$H_4 = (\bar{J}_{N_1N_2} \otimes Q_T) \cdot \left(X_1^{(1)},\, X_2^{(1)},\, X_3^{(1)},\, X_4^{(1)},\, X_1^{(4)}\right).$$

In other words, we have, by construction

$$\begin{aligned}
\operatorname*{plim}_{N_1\to\infty} \tfrac{1}{N_1N_2T} H_2'(\upsilon + \varepsilon) &= 0 \\
\operatorname*{plim}_{N_2\to\infty} \tfrac{1}{N_1N_2T} H_3'(\zeta + \varepsilon) &= 0 \\
\operatorname*{plim}_{T\to\infty} \tfrac{1}{N_1N_2T} H_4'(\lambda + \varepsilon) &= 0\,.
\end{aligned}$$

Unfortunately, transformation matrices such as $(Q_{N_1} \otimes \bar{J}_{N_2} \otimes \bar{J}_T)$ can be memory demanding even on powerful personal computers when the dataset is large. However, these transformations only define operations on simple group means and the deviations from these. Table 2.7 provides a list of matrix transformations with their equivalent scalar operations on each observation.

Naturally, it is possible to utilise $H = (H_1, H_2, H_3, H_4)$ directly to instrument the endogenous regressors. Alternatively, we can construct an asymptotically more efficient estimator by accommodating the error component structure of the model.

Performing instrumental variable estimation on model (2.31) with instrument set

Table 2.7 *Translation of matrix operations into scalar*

Matrix	Scalar
$(Q_{N_1} \otimes Q_{N_2} \otimes Q_T)X$	$x'_{ijt} - x'_{.jt} - x'_{i.t} - x'_{ij.} + x'_{i..} + x'_{.j.} + x'_{..t} - x'_{...}$
$(\bar{J}_{N_1} \otimes Q_{N_2} \otimes Q_T)X$	$x'_{.jt} - x'_{..t} - x'_{.j.} + x'_{...}$
$(Q_{N_1} \otimes \bar{J}_{N_2} \otimes Q_T)X$	$x'_{i.t} - x'_{i..} - x'_{..t} + x'_{...}$
$(Q_{N_1} \otimes Q_{N_2} \otimes \bar{J}_T)X$	$x'_{ij.} - x'_{.j.} - x'_{i..} + x'_{...}$
$(\bar{J}_{N_1} \otimes \bar{J}_{N_2} \otimes Q_T)X$	$x'_{..t} - x'_{...}$
$(\bar{J}_{N_1} \otimes Q_{N_2} \otimes \bar{J}_T)X$	$x'_{.j.} - x'_{...}$
$(Q_{N_1} \otimes \bar{J}_{N_2} \otimes \bar{J}_T)X$	$x'_{i..} - x'_{...}$

H is identical to estimating

$$P_H y = P_H X\beta + P_H u \tag{2.32}$$

with Least Squares where P_H is the projection matrix, $H(H'H)^{-1}H'$. Given the orthogonality nature of H_p for $p = 1, \dots, 4$,

$$P_H = \sum_{p=1}^{4} P_{H_p},$$

gives the projection that is identical to the sum of individual projections of the H_p. The estimator is then defined as

$$\hat{\beta}_{HT1} = (X'P_H X)^{-1} X'P_H y. \tag{2.33}$$

Since the H_p matrices can be constructed by calculating simple group means and deviations from groups means from the original data matrix and the elements of P_{H_p} are also straightforward to calculate, the resulting estimator is computationally simple without imposing excessive burden on the memory or computation requirement.

The More Efficient Hausman-Taylor Estimator

The asymptotic efficiency of this estimator can be improved, if we exploit the error component structure. As shown in Fuller and Battese (1973), and exactly as it was done in Section 2.2, we can multiply model (2.31) by $\Omega^{-1/2}$ to get

$$\Omega^{-1/2} y = \Omega^{-1/2} X\beta + \Omega^{-1/2} u$$

where $\Omega = \mathrm{E}(uu')$. Then the same estimator as defined in equation (2.33), with the same instrument set, is asymptotically efficient. Alternatively, we can interpret this 2SLS, using Maddala's (1971) work, as a direct Least Squares on model

$$P_H^* y = P_H^* X\beta + P_H^* u$$

with

$$P_H^* = \sum_{p=1}^{4} \frac{1}{\sigma_p} P_{H_p}\,,$$

where different weights are assigned to different parts of the instruments:[5]

$$\sigma_1^2 = \sigma_\varepsilon^2 \qquad \sigma_2^2 = \sigma_\varepsilon^2 + N_2 T \sigma_\upsilon^2 \qquad \sigma_3^2 = \sigma_\varepsilon^2 + N_1 T \sigma_\zeta^2 \qquad \sigma_4^2 = \sigma_\varepsilon^2 + N_1 N_2 \sigma_\lambda^2\,.$$

The more efficient estimator can then computed as

$$\hat{\beta}_{HT2} = (X' P_H^* X)^{-1} X' P_H^* y\,. \tag{2.34}$$

In practice, these variances are unknown. Specifically, the components of the variance-covariance matrix Ω are usually unknown and must be estimated in order for (2.34) to be feasible. The first observation is that the variance of the idiosyncratic error term can always be estimated from the residuals given by the Within estimator:

$$\hat{\varepsilon} = \tilde{y} - \tilde{X}^{(1)} \hat{\beta}_W^{(1)}, \quad \text{and} \quad \hat{\sigma}_\varepsilon^2 = \hat{\tilde{\varepsilon}}' \hat{\tilde{\varepsilon}} / (N_1 N_2 T - N_1 - N_2 - T + 1)$$

where $\tilde{y}$ and $\tilde{X}$ denote y and $X^{(1)}$ after the Within transformation, respectively. $\beta_W^{(1)}$ is the Within estimates associated with $X^{(1)}$. Once we obtain consistent estimates for $\beta^{(m)}$, $m = 1 \ldots 4$, we can use them to estimate σ_υ, σ_ζ and σ_λ. Specifically,

$$(Q_{N_1} \otimes \bar{J}_{N_2 T}) \left(y - X^{(1)} \hat{\beta}_{HT}^{(1)} - X^{(2)} \hat{\beta}_{HT}^{(2)} \right) = \hat{u}_1\,,$$

and it can be shown that

$$\begin{aligned}
\operatorname*{plim}_{N_1 \to \infty} \hat{u}_1' \hat{u}_1 / (N_1 N_2 T) =& \operatorname*{plim}_{N_1 \to \infty} \frac{1}{N_2 T} \cdot \frac{N_1 - 1}{N_1} \sigma_\varepsilon^2 + \frac{N_1 - 1}{N_1} \sigma_\upsilon^2 \\
=& \frac{1}{N_2 T} \sigma_\varepsilon^2 + \sigma_\upsilon^2\,.
\end{aligned}$$

From this, we can estimate σ_υ^2 as

$$\hat{\sigma}_\upsilon^2 = \hat{u}_1' \hat{u}_1 / (N_1 N_2 T) - \frac{1}{N_2 T} \hat{\sigma}_\varepsilon^2\,.$$

Similar procedures apply to the estimation of σ_ζ^2 and σ_λ^2, where residuals are collected from

$$\begin{aligned}
(\bar{J}_{N_1} \otimes Q_{N_2} \otimes \bar{J}_T) \left(y - X^{(1)} \hat{\beta}_{HT}^{(1)} - X^{(3)} \hat{\beta}_{HT}^{(3)} \right) &= \hat{u}_2 \\
(\bar{J}_{N_1 N_2} \otimes \bar{Q}_T) \left(y - X^{(1)} \hat{\beta}_{HT}^{(1)} - X^{(4)} \hat{\beta}_{HT}^{(4)} \right) &= \hat{u}_3\,,
\end{aligned}$$

[5] The result $P_H \Omega^{-1/2} = P_H^*$ is coming from the definition of the Q and $\bar{J}$ matrices.

which can be used to estimate the variance components as

$$\begin{aligned}\hat{\sigma}^2_\zeta &= \hat{u}_2'\hat{u}_2/(N_1N_2T) - \tfrac{1}{N_1T}\hat{\sigma}^2_\varepsilon \\ \hat{\sigma}^2_\lambda &= \hat{u}_3'\hat{u}_3/(N_1N_2T) - \tfrac{1}{N_1N_2}\hat{\sigma}^2_\varepsilon\,.\end{aligned}$$

Notice, that for the consistent estimation of σ^2_υ, we need $N_1 \to \infty$. Similarly, we need N_2 and T asymptotics for the consistency of the other two variance components estimators.

The rank condition to ensure parameter identifiability is that $X'P_HX$ must have full rank. Similar to HT, there is also a set of necessary order conditions that is easier to verify. Specifically,

$$\begin{aligned}k_1^{(1)} + k_3^{(1)} + k_4^{(1)} + k_7^{(1)} &\geq k_2^{(2)} \\ k_1^{(1)} + k_2^{(1)} + k_4^{(1)} + k_6^{(1)} &\geq k_2^{(3)} \\ k_1^{(1)} + k_2^{(1)} + k_3^{(1)} + k_5^{(1)} &\geq k_2^{(4)}\end{aligned} \tag{2.35}$$

have to be satisfied jointly. Although it looks like the number of instruments far exceeds the number of endogenous regressors, it is often the case that $k_l^{(m)} = 0$ for several m and l.

These order conditions reduce to the order condition in HT in the case of standard two dimensional panel. This can be seen by assuming the presence of only a single random effect, υ_i, with $\zeta_j = \lambda_t = 0$ for all j and t. By merging indexes j and t and representing the joint index by s, we reduce the three-dimensional panel into a standard two-dimensional case. This allows us to combine $X_1^{(1)}$, $X_3^{(1)}$, $X_4^{(1)}$, $X_7^{(1)}$ as they all vary over i and s and they are uncorrelated with υ_i. If we denote the total number of their columns k_1, the first order condition in (2.35) simplifies to

$$k_1 \geq k_2^{(5)}\,. \tag{2.36}$$

Also, as $k_2^{(3)} = k_2^{(4)} = 0$, the second and third order conditions hold by construction. Thus, the three order conditions reduce to equation (2.36), which requires at least as many exogenous variables in x_{is}' as endogenous variables in x_i'. This is identical to the order condition in HT.

We can also relate our order conditions to those of the two-way panel models found in Wyhowski (1994). If we restrict $\zeta_j = 0$ for all j and $\beta_1^{(3)} = \beta_2^{(3)} = 0$ with $N_2 = 1$, then the model reduces once again to a standard two-dimensional panel. Let $k_1 = k_1^{(1)}$, $k_2 = k_4^{(1)} + k_7^{(1)}$ and $k_3 = k_2^{(1)} + k_5^{(1)}$, then the above conditions reduce to

$$k_1 + k_2 \geq k_2^{(2)} \quad \text{and} \quad k_1 + k_3 \geq k_2^{(4)}\,,$$

which is identical to the order conditions given in Wyhowski (1994).

2.5.2 Time Varying Individual Specific Effects

Let's repeat the analysis for model (2.2), and again use $\underline{x}'_{ijt}$ for x'_{ijt}:

$$y_{ijt} = \underline{x}'_{ijt}\beta + \mu_{ij} + \upsilon_{it} + \zeta_{jt} + \varepsilon_{ijt}\,, \tag{2.37}$$

Recall in the case of model (2.31), only variables with a single index, namely, x'_i, x'_j and x'_t required instruments from x'_{ijt}, as the rest of the parameters could be identified from the Within estimator. This is no longer case with model (2.37). Since the unobserved heterogeneities vary over individuals and time, the Within transformation will also eliminate all x'_{ij}, x'_{it} and x'_{jt} variables in addition to the variables with a single index. This means the parameters of x'_{ij}, x'_{it} and x'_{jt} will be unidentified as well.

Following the same approach as in the previous sections, Table 2.8 shows the partitions of the regressor vector based on the sources of endogeneity:

Table 2.8 *Sources of endogeneity on the level of partitions of the regressors for model (2.37)*

Correlated with	Partition						
None	x'_{1ijt}	x'_{1ij}	x'_{1it}	x'_{1jt}	x'_{1i}	x'_{1j}	x'_{1t}
μ_{ij}	x'_{2ijt}	x'_{2ij}	x'_{2it}	x'_{2jt}	x'_{2i}	x'_{2j}	
υ_{it}	x'_{3ijt}	x'_{3ij}	x'_{3it}	x'_{3jt}	x'_{3i}		x'_{2t}
ζ_{jt}	x'_{4ijt}	x'_{4ij}	x'_{4it}	x'_{4jt}		x'_{3j}	x'_{3t}
μ_{ij}, υ_{it}	x'_{5ijt}	x'_{5ij}	x'_{5it}	x'_{5jt}	x'_{4i}		
μ_{ij}, ζ_{jt}	x'_{6ijt}	x'_{6ij}	x'_{6it}	x'_{6jt}		x'_{4j}	
υ_{it}, ζ_{jt}	x'_{7ijt}	x'_{7ij}	x'_{7it}	x'_{7jt}			x'_{4t}
μ_{ij}, υ_{it}, ζ_{jt}	x'_{8ijt}	x'_{8ij}	x'_{8it}	x'_{8jt}			

The different group means of $x_{ijt'}$ can be used as instrumental variables for the endogenous variables in x'_{ij}, x'_{it}, x'_{jt}, x'_i, x'_j and x'_t but interestingly, groups means of x'_{ij}, x'_{it} and x'_{jt} can also be used as instruments for x'_i, x'_j and x'_t. The potential of each group of variables to be instruments of the others is illustrated in Figure 2.1.

Allowing individual specific effects to be time varying leads to some additional complications in implementing variations on the HT estimator. Extra care is required to ensure the validity of each internal instrument. For example, while the group means of x'_{2ijt} can in theory be used as instruments for x'_{3ij} and x'_{5ij}, the fact that x'_{2ijt} is correlated with μ_{ij} means that the group means must be taken over both i and j indexes which makes it invalid to be an instrument for x'_{5ij}. This argument applies more generally to model (2.37) and can be summarised in Table 2.9.

While the general HT approach is still theoretical sound in this setting, the time

10.14754/CEU.2017.05

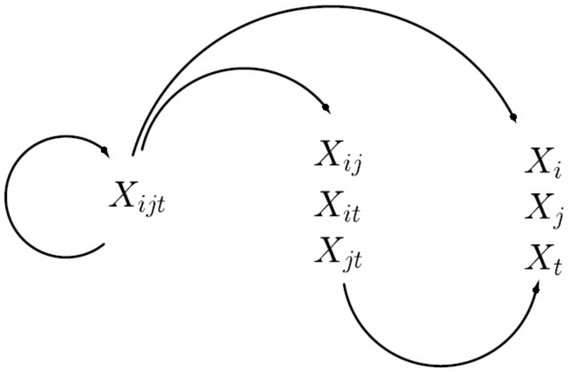

Figure 2.1 Possible Instrumental Variables

Table 2.9 *Pairs of variables needed to be instrumented jointly*

Variable	Pairs		
x'_{ij}	$(x'_{3ij}$	$x'_{5ij})$	
	$(x'_{4ij}$	$x'_{6ij})$	
	$(x'_{7ij}$	$x'_{8ij})$	
x'_{it}	$(x'_{2it}$	$x'_{5it})$	
	$(x'_{4it}$	$x'_{7it})$	
	$(x'_{6it}$	$x'_{8it})$	
x'_{jt}	$(x'_{2jt}$	$x'_{6jt})$	
	$(x'_{3jt}$	$x'_{7jt})$	
	$(x'_{5jt}$	$x'_{8jt})$	
x'_i	$(x'_{2i}$	x'_{3i}	$x'_{4i})$
x'_j	$(x'_{2j}$	x'_{3j}	$x'_{4j})$
x'_t	$(x'_{2t}$	x'_{3t}	$x'_{4t})$

varying nature of individual specific effects imposes additional restrictions on the order conditions. Following the same notation as above, define $X_l^{(m)}$ as the data matrix counterpart of row l and column m in Table 2.9 with $k_l^{(m)}$ denotes the number of columns of $X_l^{(m)}$. The instrument(s) for variable $X_l^{(m)}$ can always be expressed as linear transformations of the original variable(s). Specifically, $H_p = R_p \cdot X_p$, where X_p is a collection of variables which are used to create internal instruments and R_p represents the appropriate linear transformation. For each endogenous variable, Table 2.10 presents the instruments, H_p, the associated transformations, R_p, and the original variable set, X_p, for each group of endogenous variables.

Table 2.10 *Proposed instruments H_p for each endogenous variable*

Endogenous Variable[a]	Instrument R_p	X_p
$X^{(1)}$	$(Q_{N_1} \otimes Q_{N_2} \otimes Q_T)$	$X^{(1)}$
$(X_2^{(2)}, X_1^{(2)})$	$(Q_{N_1} \otimes Q_{N_2} \otimes \bar{J}_T)$	$(X_1^{(1)}, X_3^{(1)}, X_4^{(1)}, X_7^{(1)}, X_1^{(2)})$
$(X_3^{(2)}, X_5^{(2)})$	$(Q_{N_1} \otimes Q_{N_2} \otimes \bar{J}_T)$	$(X_1^{(1)}, X_4^{(1)})$
$(X_4^{(2)}, X_6^{(2)})$	$(Q_{N_1} \otimes Q_{N_2} \otimes \bar{J}_T)$	$(X_1^{(1)}, X_3^{(1)})$
$(X_7^{(2)}, X_8^{(2)})$	$(Q_{N_1} \otimes Q_{N_2} \otimes \bar{J}_T)$	$X_1^{(1)}$
$(X_3^{(3)}, X_1^{(3)})$	$(Q_{N_1} \otimes \bar{J}_{N_2} \otimes Q_T)$	$(X_1^{(1)}, X_2^{(1)}, X_4^{(1)}, X_6^{(1)}, X_1^{(3)})$
$(X_2^{(3)}, X_5^{(3)})$	$(Q_{N_1} \otimes \bar{J}_{N_2} \otimes Q_T)$	$(X_1^{(1)}, X_4^{(1)})$
$(X_4^{(3)}, X_7^{(3)})$	$(Q_{N_1} \otimes \bar{J}_{N_2} \otimes Q_T)$	$(X_1^{(1)}, X_2^{(1)})$
$(X_6^{(3)}, X_8^{(3)})$	$(Q_{N_1} \otimes \bar{J}_{N_2} \otimes Q_T)$	$X_1^{(1)}$
$(X_4^{(4)}, X_1^{(4)})$	$(\bar{J}_{N_1} \otimes Q_{N_2} \otimes Q_T)$	$(X_1^{(1)}, X_2^{(1)}, X_3^{(1)}, X_5^{(1)}, X_1^{(4)})$
$(X_2^{(4)}, X_6^{(4)})$	$(\bar{J}_{N_1} \otimes Q_{N_2} \otimes Q_T)$	$(X_1^{(1)}, X_3^{(1)})$
$(X_3^{(4)}, X_7^{(4)})$	$(\bar{J}_{N_1} \otimes Q_{N_2} \otimes Q_T)$	$(X_1^{(1)}, X_2^{(1)})$
$(X_5^{(4)}, X_8^{(4)})$	$(\bar{J}_{N_1} \otimes Q_{N_2} \otimes Q_T)$	$X_1^{(1)}$
$(X_2^{(5)}, X_3^{(5)}, X_4^{(5)}, X_1^{(5)})$	$(Q_{N_1} \otimes \bar{J}_{N_2} \otimes \bar{J}_T)$	$(X_1^{(1)}, X_4^{(1)}, X_1^{(2)}, X_4^{(2)}, X_1^{(3)}, X_4^{(3)}, X_1^{(5)})$
$(X_2^{(6)}, X_3^{(6)}, X_4^{(6)}, X_1^{(6)})$	$(\bar{J}_{N_1} \otimes Q_{N_2} \otimes \bar{J}_T)$	$(X_1^{(1)}, X_3^{(1)}, X_1^{(2)}, X_3^{(2)}, X_1^{(4)}, X_3^{(4)}, X_1^{(6)})$
$(X_2^{(7)}, X_3^{(7)}, X_4^{(7)}, X_1^{(7)})$	$(\bar{J}_{N_1} \otimes \bar{J}_{N_2} \otimes Q_T)$	$(X_1^{(1)}, X_2^{(1)}, X_1^{(3)}, X_2^{(3)}, X_1^{(4)}, X_2^{(4)}, X_1^{(7)})$

[a] For each row p, the instrument is obtained as $H_p = R_p \cdot X_p$. Instruments for exogenous regressors are simply themselves, and added, quite arbitrarily, to lines 2,6,10,14,15,16.

Note that exogenous variables also serve as their own instruments. Once we have all the instruments collected, it is straightforward to extend the HT estimator by following the same approach as before.

$$\hat{\beta}_{HT1} = (X'P_H X)^{-1} X'P_H y,$$

and since all H_p are orthogonal to each other,

$$P_H = \sum_{p=1}^{16} P_{H_p}.$$

The more efficient estimator which takes into account the error structure in u_{ijt}:

$$\hat{\beta}_{HT2} = (X'P_H^* X)^{-1} X'P_H^* y, \tag{2.38}$$

with

$$P_H^* = P_H \cdot \Omega^{-1/2} = \sum_{p=1}^{16} \frac{1}{\sigma_p} P_{H_p} \, ,$$

where

$$\begin{array}{lll}
\sigma_1^2 & = \sigma_\varepsilon^2 & \\
\sigma_p^2 & = \sigma_\varepsilon^2 + T\sigma_\mu^2 & p = 2, \ldots, 5 \\
\sigma_p^2 & = \sigma_\varepsilon^2 + N_2\sigma_\upsilon^2 & p = 6, \ldots, 9 \\
\sigma_p^2 & = \sigma_\varepsilon^2 + N_1\sigma_\zeta^2 & p = 10, \ldots, 13 \\
\sigma_{14}^2 & = \sigma_\varepsilon^2 + N_2\sigma_\upsilon^2 + T\sigma_\mu^2 & \\
\sigma_{15}^2 & = \sigma_\varepsilon^2 + N_1\sigma_\zeta^2 + T\sigma_\mu^2 & \\
\sigma_{16}^2 & = \sigma_\varepsilon^2 + N_1\sigma_\zeta^2 + N_2\sigma_\upsilon^2 \, . &
\end{array}$$

The set of order conditions necessary to ensure parameter identification is however, much less trivial here. It is clear that the order condition for each $X^{(m)}$ is independent from each other, as we use different group variations of the (possibly same) instruments. The same is not true for the partitions *within* $X^{(m)}$. The order condition for each has to hold not only individually, but *jointly* as well. Table 2.11 organizes these conditions which all have to be satisfied, in order to have as many instruments, as endogenous variables.

The complexity of these necessary order conditions means that the general HT approach may not be as practical in higher dimension as it is in standard two dimensional panel data model. Nevertheless, as several k_l^m are potentially null, whether the order conditions are hard to satisfy or not is a matter of the exact economic application and model, and not the result of a purely theoretical investigation.

2.5.3 Properties

While the order conditions are necessary for parameter identification it is well-known that HT-type estimators, as with other IV estimators in general, are biased in finite samples. It is therefore important to examine their asymptotic properties. In general, the estimators are consistent, if the proposed instruments are asymptotically uncorrelated with the composite disturbance term u_{ijt}. As the instruments are constructed to fulfil this particular requirement under different types of asymptotics, we have to derive the specific asymptotics that would ensure the validity of all the instrument sets. For both models (2.31) and (2.37), the $\hat{\beta}_{HT1}$ and $\hat{\beta}_{HT2}$ have asymptotic normal distributions, and are consistent only when all $N_1, N_2, T \to \infty$ jointly. This is perhaps not surprising, as instruments for x_i', x_j' and x_t' are asymptotically uncorrelated with u_{ijt} if $N_1 \to \infty$, $N_2 \to \infty$, and $T \to \infty$, respectively.

Table 2.11 *Order conditions for Model (2.37)*

Variable	Condition
$X^{(2)}$	$k_1^{(1)} \geq k_7^{(2)} + k_8^{(2)}$
	$k_1^{(1)} - k_7^{(2)} - k_8^{(2)} + k_3^{(1)} \geq k_4^{(2)} + k_6^{(2)}$
	$k_1^{(1)} - k_7^{(2)} - k_8^{(2)} + k_3^{(1)} - k_4^{(2)} - k_6^{(2)} + k_4^{(1)} \geq k_3^{(2)} + k_5^{(2)}$
	$k_1^{(1)} - k_7^{(2)} - k_8^{(2)} + k_3^{(1)} - k_4^{(2)} - k_6^{(2)} + k_4^{(1)} - k_3^{(2)} - k_5^{(2)} + k_7^{(1)} \geq k_2^{(2)}$
$X^{(3)}$	$k_1^{(1)} \geq k_6^{(3)} + k_8^{(3)}$
	$k_1^{(1)} - k_6^{(3)} - k_8^{(3)} + k_2^{(1)} \geq k_4^{(3)} + k_7^{(3)}$
	$k_1^{(1)} - k_6^{(3)} - k_8^{(3)} + k_2^{(1)} - k_4^{(3)} - k_7^{(3)} + k_4^{(1)} \geq k_2^{(3)} + k_5^{(3)}$
	$k_1^{(1)} - k_6^{(3)} - k_8^{(3)} + k_2^{(1)} - k_4^{(3)} - k_7^{(3)} + k_4^{(1)} - k_2^{(3)} - k_5^{(3)} + k_6^{(1)} \geq k_3^{(3)}$
$X^{(4)}$	$k_1^{(1)} \geq k_5^{(4)} + k_8^{(4)}$
	$k_1^{(1)} - k_5^{(4)} - k_8^{(4)} + k_2^{(1)} \geq k_3^{(4)} + k_7^{(4)}$
	$k_1^{(1)} - k_5^{(4)} - k_8^{(4)} + k_2^{(1)} - k_3^{(4)} - k_7^{(4)} + k_3^{(1)} \geq k_2^{(4)} + k_6^{(4)}$
	$k_1^{(1)} - k_5^{(4)} - k_8^{(4)} + k_2^{(1)} - k_3^{(4)} - k_7^{(4)} + k_3^{(1)} - k_2^{(4)} - k_6^{(4)} + k_5^{(1)} \geq k_4^{(4)}$
$X^{(5)}$	$k_1^{(1)} + k_4^{(1)} + k_1^{(2)} + k_4^{(2)} + k_1^{(3)} + k_4^{(3)} \geq k_2^{(5)} + k_3^{(5)} + k_4^{(5)}$
$X^{(6)}$	$k_1^{(1)} + k_3^{(1)} + k_1^{(2)} + k_3^{(2)} + k_1^{(4)} + k_3^{(4)} \geq k_2^{(6)} + k_3^{(6)} + k_4^{(6)}$
$X^{(7)}$	$k_1^{(1)} + k_2^{(1)} + k_1^{(3)} + k_2^{(3)} + k_1^{(4)} + k_2^{(4)} \geq k_2^{(7)} + k_3^{(7)} + k_4^{(7)}$

2.5.4 Incomplete Data

Incompleteness, in general, does not violate the feasibility and the validity of HT-type estimators, but can lead to computational difficulties. The transformations to be imposed on the data to construct the instruments cannot be represented as neat sample means of the variables. Instead, we have to rely on the results from Section 1.5 which derives incompleteness-robust data transformations.

Obviously this complexity is directly related to the complexity of the error component structure. For model (2.31), $(Q_{N_1} \otimes \bar{J}_{N_2T})$ can still be represented with scalar operations: first, by taking averages over j and t for each i, then de-mean the data with respect to i:

$$(Q_{N_1} \otimes \bar{J}_{N_2T})X \quad \text{in scalar form is} \quad \frac{1}{\sum_j |T_{ij}|}\sum_{j=1}^{N_2}\sum_{t\in T_{ij}} x'_{ijt} - \frac{1}{\sum_{i,j}|T_{ij}|}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\sum_{t\in T_{ij}} x'_{ijt}$$

for $i = 1 \dots N_1$.

Similar logic holds for $(\bar{J}_{N_1} \otimes Q_{N_2} \otimes \bar{J}_T)$ and $(\bar{J}_{N_1N_2} \otimes Q_T)$. The situation gets complicated, when multiple Q matrices appear in the Kronecker products, that is we remove more, than one within group variation. For such cases, unfortunately, involving matrix operations is inevitable.

2.5.5 Using External Instruments

The discussion so far focused on HT-type estimators, which utilise existing variables to generate "internal" instruments for any endogenous regressors. This section discusses briefly the more conventional IV approach, specifically, the use of "external" variables as instruments for the endogenous variables.

Consider model (2.31) with the *a priori* knowledge that x'_{ijt}, or any of its transformed counterparts are not correlated with x'_i. Obviously, we can still use the Within transform of $\tilde{x}'_{ijt}$ to instrument itself and use its different group means to instrument x'_j and x'_t but the parameters associated with x'_i cannot be identified due to the Within transformation. In this case, we can try to find a variable z'_{ijt}, which perhaps is fixed over j and/or t, such that

$$\operatorname{Corr}\left(z'_{ijt}, x'_i\right) \neq 0 \quad \text{and} \quad \operatorname{Corr}\left(z'_{i..} - z'_{...}, u_{ijt}\right) = 0 \tag{2.39}$$

Note that for the second condition in (2.39) to hold we only require $\operatorname{Corr}(z'_{ijt}, \upsilon_i) = 0$. Once we obtain z'_{ijt}, the instrument is constructed as in Section 2.5.1, where z'_{ijt} is used as instruments instead of $\tilde{x}'_{ijt}$ for x'_{2i}. In terms of identification, the additional order condition $g \geq k_2^{(2)}$, where g is the number of instrumental variables, is required. The condition requires that the number of instrumental variables must be as large as the number endogenous regressors in x'_i, which coincides with the standard result of identifiability in the instrumental variable literature.

2.6 Various Tests for Random Effects Models

This section lines up various tests for random effects models. First, a model selection tool is developed based on Baltagi et al. (1992), then two tests are discussed concerning the endogeneity issue: a simple test for endogeneity, and a test for instrument validity.

2.6.1 Testing for Model Specification

In this section we show for the all-encompassing model (2.2) how to test for the existence of different components of the unobserved heterogeneity. More specifically, we test for the nullity of the variance of some random components against the alternative, that the given variance is positive. We have to be careful, however, about what we assume about the rest of the variances. Testing $H_0: \ \sigma^2_\mu = 0$ against

$H_A:\ \sigma^2_\mu > 0$ implicitly assumes, that $\sigma^2_\upsilon = \sigma^2_\zeta = 0$, and so on. In what follows, we collect some null-, and alternative hypotheses, and present the mechanism to test them:

$$\begin{array}{llll}
\mathrm{H}_0^a: & \sigma^2_\mu = 0 \mid \sigma^2_\upsilon > 0,\ \sigma^2_\zeta > 0; & \mathrm{H}_1^a: & \sigma^2_\mu > 0 \mid \sigma^2_\upsilon > 0,\ \sigma^2_\zeta > 0 \\
\mathrm{H}_0^b: & \sigma^2_\mu = 0 \mid \sigma^2_\upsilon = 0,\ \sigma^2_\zeta > 0; & \mathrm{H}_1^b: & \sigma^2_\mu > 0 \mid \sigma^2_\upsilon = 0,\ \sigma^2_\zeta > 0 \\
\mathrm{H}_0^c: & \sigma^2_\mu = 0 \mid \sigma^2_\upsilon = 0,\ \sigma^2_\zeta = 0; & \mathrm{H}_1^c: & \sigma^2_\mu > 0 \mid \sigma^2_\upsilon = 0,\ \sigma^2_\zeta = 0 \\
\mathrm{H}_0^d: & \sigma^2_\mu = 0 \mid \sigma^2_\upsilon > 0,\ \sigma^2_\zeta > 0; & \mathrm{H}_1^d: & \sigma^2_\mu > 0 \mid \sigma^2_\upsilon = 0,\ \sigma^2_\zeta > 0 \\
\mathrm{H}_0^e: & \sigma^2_\mu = 0 \mid \sigma^2_\upsilon > 0,\ \sigma^2_\zeta = 0; & \mathrm{H}_1^e: & \sigma^2_\mu > 0 \mid \sigma^2_\upsilon = 0,\ \sigma^2_\zeta = 0
\end{array}$$

To test these hypotheses, we will invoke the ANOVA F-test, and adjust it to our purposes. In its general form, as derived in Baltagi et al. (1992),

$$F = \frac{y' M_{Z_1} D (D' M_{Z_1} D)^- D' M_{Z_1} y / (p - r)}{y' M_{Z_2} y / (N_1 N_2 T - \tilde{k} - p + r)}, \tag{2.40}$$

where both M_{Z_1} and M_{Z_2} are orthogonal projectors, and the degrees of freedom is calculated from p, r, and $\tilde{k}$. Table 2.12 captures each specific matrix and constant for all hypotheses listed above.

Although (2.40) suffices theoretically, let us not forget that in order to perform the test, we have to invert $(D' M_{Z_1} D)$, a matrix as large as the data. Instead, to avoid this computational burden, we can elaborate on (2.40), and find out what the respective projection matrices do to the data:

$$F = \frac{F_1/(p-r)}{F_2/(N_1 N_2 T - \tilde{k} - p + r)}$$

where

$$\begin{aligned}
F_1 &= (\tilde{\tilde{y}} - \tilde{\tilde{X}}(X'X)^{-1}X'y)'(I + X(\tilde{X}'\tilde{X})X')(\tilde{\tilde{y}} - \tilde{\tilde{X}}(X'X)^{-1}X'y) \\
&= (\tilde{\tilde{y}} - \tilde{\tilde{X}}\hat{\beta}_{OLS})'(I + X(\tilde{X}'\tilde{X})X')(\tilde{\tilde{y}} - \tilde{\tilde{X}}\hat{\beta}_{OLS}),
\end{aligned}$$

and

$$F_2 = (\tilde{y} - \tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}\tilde{y})'(\tilde{y} - \tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}\tilde{y}) = (\tilde{y} - \tilde{X}\hat{\beta}_w)'(\tilde{y} - \tilde{X}\hat{\beta}_w),$$

with the " $\sim$"-s on the top denoting different transformations. For H_0^a and H_A^a for example, these are

$$\tilde{y}_{ijt} = y_{ijt} - \bar{y}_{.jt} - \bar{y}_{i.t} - \bar{y}_{ij.} + \bar{y}_{..t} + \bar{y}_{.j.} + \bar{y}_{i..} - \bar{y}_{...} \tag{2.41}$$

(which is the optimal Within of model (2.2)), and

$$\tilde{\tilde{y}}_{ijt} = y_{ijt} - \bar{y}_{.jt} - \bar{y}_{i.t} + \bar{y}_{..t}\,. \tag{2.42}$$

To get an insight into the specific formula, notice, that we actually compare two models, the one where the sources of all variations are cleared (the denominator of

10.14754/CEU.2017.05

Table 2.12 *Specific functional forms of the ANOVA F-test*

Hypothesis	$Z_1{}^a$	D	Z_2	p	r	$\tilde{k}$
H^a	(X, D_2, D_3)	$(I_{N_1N_2} \otimes J_T)$	(X, D_1, D_2, D_3)	N_1N_2	1	$N_1(T-1)+N_2(T-1)+T+k$
H^b	(X, D_3)	$(I_{N_1N_2} \otimes J_T)$	(X, D_1, D_3)	N_1N_2	1	$N_2(T-1)+k$
H^c	X	$(I_{N_1N_2} \otimes J_T)$	(X, D_1)	N_1N_2	1	k
H^d	(X, D_3)	$(I_{N_1N_2} \otimes J_T, I_{N_1} \otimes J_{N_2} \otimes I_T)$	(X, D_1, D_2, D_3)	$N_1N_2+N_1T$	2	k
H^e	X	$(I_{N_1N_2} \otimes J_T, I_{N_1} \otimes J_{N_2} \otimes I_T)$	(X, D_1, D_2)	$N_1N_2+N_1T$	2	k

[a] $M_Z = I - Z(Z'Z)^{-}Z'$, $D_1 = (I_{N_1N_2} \otimes \iota_T)$, $D_2 = (I_{N_1} \otimes \iota_{N_2} \otimes I_T)$, and $D_3 = (\iota_{N_1} \otimes I_{N_2T})$.

(2.40)) with the one where all variation is cleared, but the one coming from μ_{ij} (the numerator of (2.40)). This is, because both under the null and the alternative, we assume, that $\sigma^2_\upsilon > 0$ and $\sigma^2_\zeta > 0$, that is, they are irrelevant from our point of view, we can eliminate both υ_{it} and ζ_{jt} with an orthogonal projection. Further, under the alternative, $\sigma^2_\mu > 0$ also holds, so we eliminate μ_{ij} as well, but save it under the null. The numerator and the denominator of (2.40) is then compared, and if it is sufficiently close to 1, we cannot reject the nullity of σ^2_μ.

Not much changes when the underlying data is incomplete. In principle, the orthogonal projections M_{Z_1} and M_{Z_2} now cannot be represented as linear transformations on the data, only in semi-scalar form, with the inclusion of some matrix operations, listed in Section 1.5.2. For example, (2.41) corresponds to (1.28), while (2.42) corresponds to (1.27) in case of incomplete data. Once we have the incomplete-robust $\tilde{y}$, $\tilde{\tilde{y}}$ (similarly for X) variables, the F statistic is obtained as in (2.40), with the properly computed degrees of freedom.

2.6.2 Testing for Exogeneity

While the presence of endogeneity can often be argued from theoretical grounds, there are many cases where its presence is not particularly obvious from a practical perspective. As such, tests for endogeneity are clearly useful. In the case of the one-way error component model in standard two dimensional panel data model, a simple Hausman test (see Hausman, 1978 and further Section 4.3 of Baltagi, 2013) is sufficient. In this case, the rejection of the null of exogeneity not only suggests the presence of endogeneity but also the source of endogeneity, specifically, the regressors are correlated with the unobserved heterogeneity.

The higher-dimensional case is slightly more complicated, as a standard Hausman-test, which compares the GLS to the Within estimator, can only reveal the presence of endogeneity but not the actual sources of endogeneity. The null hypothesis is

$$H_0 : \mathrm{E}(\upsilon_i + \zeta_j + \lambda_t | \underline{x}'_{ijt}) = 0 \quad \text{against} \quad H_1 : \mathrm{E}(\upsilon_i + \zeta_j + \lambda_t | \underline{x}'_{ijt}) \neq 0 \tag{2.43}$$

The GLS estimator under model (2.31) is consistent only if H_0 is true, but the Within estimator is consistent both under the null and alternative hypotheses. Following Hausman (1978), a test can be constructed with the vector $\hat{q}_1 = \hat{\beta}^{(1)}_{GLS} - \hat{\beta}^{(1)}_{Within}$,

$$\begin{aligned} m &= \hat{q}_1' \operatorname{Var}(\hat{q}_1)^{-1} \hat{q}_1 \\ &= \left(\hat{\beta}^{(1)}_{GLS} - \hat{\beta}^{(1)}_{Within}\right)' \left(\operatorname{Var}(\hat{\beta}^{(1)}_{GLS}) - \operatorname{Var}(\hat{\beta}^{(1)}_{Within})\right)^{-1} \left(\hat{\beta}^{(1)}_{GLS} - \hat{\beta}^{(1)}_{Within}\right) \end{aligned} \tag{2.44}$$

where $\mathrm{Var}(\hat{\beta}^{(1)}_{GLS} - \hat{\beta}^{(1)}_{Within}) = \mathrm{Var}(\hat{\beta}^{(1)}_{GLS}) - \mathrm{Var}(\hat{\beta}^{(1)}_{Within})$.

Note that plim $\hat{q}_1 = 0$ under the null, but plim $\hat{q}_1 =$ plim $\hat{\beta}^{(1)}_{GLS} - \beta \neq 0$ under the alternative. Therefore, if $\hat{q}_1$ is sufficiently far from 0 then there are evidence against the null of exogeneity. The test statistic m can be shown to have a χ^2_d distribution with $d = k^{(1)}$, the number of variables in x'_{ijt}.

The test of (2.44) only provides evidence against exogeneity but it does not provide any information on the sources of endogeneity. For example, it may be the case that the regressors are correlated with the error components solely through υ_i, and so both $\mathrm{E}(\zeta_j|\underline{\mathrm{x}}_{ijt}) = 0$ and $\mathrm{E}(\lambda_t|\underline{\mathrm{x}}_{ijt}) = 0$. If this is the case, only two partitions of x'_{ijt} and x'_i should be considered. Specifically, the ones that are uncorrelated and the ones that are correlated with υ_i. This obviously reduces the assumptions we have to make about the model and, at the same time, increases the number of variables available for purposes of constructing internal instruments.

In order to address this issue, a set of subsequent tests can be constructed. If we eliminate (ζ_j, λ_t) from the model with a simple transformation such that

$$\tilde{y}_{ijt} = \tilde{\underline{\mathrm{x}}}'_{ijt}\beta + \tilde{\upsilon}_i + \tilde{\varepsilon}_{ijt} \quad \text{with} \quad \tilde{\underline{\mathrm{x}}}'_{ijt} = (\underline{\mathrm{x}}'_{ijt} - \bar{\underline{\mathrm{x}}}'_{.j.} - \bar{\underline{\mathrm{x}}}'_{..t} + \bar{\underline{\mathrm{x}}}'_{...})$$

we can test

$$H_0 : \mathrm{E}(\tilde{\upsilon}_i|\tilde{\underline{\mathrm{x}}}'_{ijt}) = 0 \quad \text{against} \quad H_1 : \mathrm{E}(\tilde{\upsilon}_i|\tilde{\underline{\mathrm{x}}}'_{ijt}) \neq 0\,, \tag{2.45}$$

i.e., if the source of endogeneity is υ_i, after removing possible correlations with ζ_j and λ_t. We can repeat this test for ζ_j and λ_t on models

$$\begin{aligned} \tilde{y}_{ijt} &= \tilde{\underline{\mathrm{x}}}'_{ijt}\beta + \tilde{\zeta}_j + \tilde{\varepsilon}_{ijt} \quad \text{with} \quad \tilde{\underline{\mathrm{x}}}'_{ijt} = (\underline{\mathrm{x}}'_{ijt} - \bar{\underline{\mathrm{x}}}'_{i..} - \bar{\underline{\mathrm{x}}}'_{..t} + \bar{\underline{\mathrm{x}}}'_{...}) \\ \tilde{y}_{ijt} &= \tilde{\underline{\mathrm{x}}}'_{ijt}\beta + \tilde{\lambda}_t + \tilde{\varepsilon}_{ijt} \quad \text{with} \quad \tilde{\underline{\mathrm{x}}}'_{ijt} = (\underline{\mathrm{x}}'_{ijt} - \bar{\underline{\mathrm{x}}}'_{i..} - \bar{\underline{\mathrm{x}}}'_{.j.} + \bar{\underline{\mathrm{x}}}'_{...}) \end{aligned}$$

by testing

$$\begin{aligned} H_0 :\ & \mathrm{E}(\tilde{\zeta}_j|\tilde{\underline{\mathrm{x}}}'_{ijt}) = 0 \quad \text{against} \quad H_1 : \mathrm{E}(\tilde{\zeta}_j|\tilde{\underline{\mathrm{x}}}'_{ijt}) \neq 0 \\ H_0 :\ & \mathrm{E}(\tilde{\lambda}_t|\tilde{\underline{\mathrm{x}}}'_{ijt}) = 0 \quad \text{against} \quad H_1 : \mathrm{E}(\tilde{\lambda}_t|\tilde{\underline{\mathrm{x}}}'_{ijt}) \neq 0\,, \end{aligned} \tag{2.46}$$

respectively. All three test statistics can be constructed similar to equation (2.44) and they follow a χ^2_d distribution with $d = k^{(1)}$. Depending on the outcome of the tests, we can reformulate the partitions of the variables to obtain a set of valid instruments more efficiently.

2.6.3 Testing for Instrument Validity

The discussion in Section 2.5.1 implicitly assumes the existence of valid instruments, specifically,

$$\text{plim } H'u = 0\,.$$

Typically economic rationale are often used to argue for or against this assumption, however these arguments are usually much less credible and more difficult to justify in random effects panel models. This is largely due to the many interdependencies between the variables and the error components. Fortunately, it is possible to test for the validity of the proposed instruments, so long as the parameters are over-identified in the model. The basic idea is to compare the HT-type estimator to the Within one. This is essentially just another form of Hausman-test as proposed in Hausman (1978), where we form our null and alternative as

$$H_0 : \text{plim } H'u = 0 \quad \text{against} \quad H_1 : \text{plim } H'u \neq 0 \tag{2.47}$$

and use the fact that the Within estimator is consistent under both the null and the alternative, serving as a "reference estimator", but the HT-type estimator is consistent but efficient under the null.

A test statistic of the form, with $\hat{q}^2 = \hat{\beta}_{HT2} - \hat{\beta}_{Within}$,

$$m = \hat{q}_2' \, \text{Var}(\hat{q}_2)^{-1} \hat{q}_2$$

can be constructed, and shown to have a χ^2_d distribution with $d = \text{rank}(\text{Var}(\hat{q}_2))$. We can elaborate on $\text{Var}(\hat{q}^2)$ and find that

$$\begin{aligned} \text{Var}(\hat{q}_2) = \text{Var}(\hat{\beta}_{HT2} - \hat{\beta}_{Within}) &= \text{Var}(\hat{\beta}_{HT2}) - \text{Var}(\hat{\beta}_{Within}) \\ &= (X'P_H\hat{\Omega}^{-1}P_HX)^{-1} - (\tilde{X}'\tilde{X})^{-1}\hat{\sigma}^2_\varepsilon \,. \end{aligned}$$

Intuitively, if $\hat{q}_2$ deviates from zero raises concerns about the validity of the instruments. If the null is rejected, it does not only suggest that some variables failed as instruments, it also implies that these variables required instruments from other variables. In order words, these variables are themselves endogenous

2.7 Extensions

So far we have seen how to formulate and estimate three-way error components models. However, it is more and more typical to have data sets which require an even higher dimensional approach. As the number of feasible model formulations grow exponentially along with the dimensions, there is no point to attempt to collect all of them. Rather, we will take the 4D representation of the all-encompassing model (2.2), and show how the extension to higher dimensions can be carried out.

2.7.1 4D and beyond

The baseline 4D model we use reads as, with $i = 1 \dots N_1$, $j = 1 \dots N_2$, $s = 1 \dots N_3$, and $t = 1 \dots T$,

$$y_{ijst} = x'_{ijst}\beta + \mu_{ijs} + \upsilon_{ist} + \zeta_{jst} + \lambda_{ijt} + \varepsilon_{ijst} = x'_{ijst}\beta + u_{ijst}\,, \tag{2.48}$$

where we keep assuming, that u (and its components individually) have zero mean, the components are pairwise uncorrelated, and further,

$$\begin{aligned}
\mathrm{E}(\mu_{ijs}\mu_{i'j's'}) &= \begin{cases} \sigma_\mu^2 & i = i' \text{ and } j = j' \text{ and } s = s' \\ 0 & \text{otherwise} \end{cases} \\
\mathrm{E}(\upsilon_{ist}\upsilon_{i's't'}) &= \begin{cases} \sigma_\upsilon^2 & i = i' \text{ and } s = s' \text{ and } t = t' \\ 0 & \text{otherwise} \end{cases} \\
\mathrm{E}(\zeta_{jst}\zeta_{j's't'}) &= \begin{cases} \sigma_\zeta^2 & j = j' \text{ and } s = s' \text{ and } t = t' \\ 0 & \text{otherwise} \end{cases} \\
\mathrm{E}(\lambda_{ijt}\zeta_{i'j't'}) &= \begin{cases} \sigma_\lambda^2 & i = i' \text{ and } j = j' \text{ and } t = t' \\ 0 & \text{otherwise} \end{cases}
\end{aligned}$$

The covariance matrix of such error components formulation is

$$\begin{aligned}
\Omega = \mathrm{E}(uu') = \quad & \sigma_\mu^2(I_{N_1N_2N_3} \otimes J_T) + \sigma_\upsilon^2(I_{N_1} \otimes J_{N_2} \otimes I_{N_3T}) \\
& + \sigma_\zeta^2(J_{N_1} \otimes I_{N_2N_3T}) + \sigma_\lambda^2(I_{N_1N_2} \otimes J_{N_3} \otimes I_T) + \sigma_\varepsilon^2 I_{N_1N_2N_3T}\,.
\end{aligned} \tag{2.49}$$

Its inverse can be simply calculated, following the method developed in Section 2.3, and the estimation of the variance components can also be derived as in Section 2.4, see for details Appendix C.

The estimation procedure is not too difficult in the incomplete case either, at least not theoretically. Taking care of the unbalanced nature of the data in four dimensional panels has nevertheless a growing importance, as the likelihood of having missing and/or incomplete data increases dramatically in higher dimensions. Conveniently, we keep assuming, that our data is such, that, for each (ijs) individual, $t \in T_{ijs}$, where T_{ijs} is a subset of the index-set $\{1, \dots, T\}$, that is, we have $|T_{isj}|$ identical observations for each (ijs) pair. First, let us write up the covariance matrix of (2.48) as

$$\Omega = \mathrm{E}(uu') = \sigma_\varepsilon^2 I + \sigma_\mu^2 D_1 D_1' + \sigma_\upsilon^2 D_2 D_2' + \sigma_\zeta^2 D_3 D_3' + \sigma_\lambda^2 D_4 D_4'\,, \tag{2.50}$$

where, in the complete case,

$$\begin{aligned}
D_1 &= (I_{N_1N_2N_3} \otimes \iota_T)\,, \quad D_2 = (I_{N_1} \otimes \iota_{N_2} \otimes I_{N_3T})\,, \quad D_3 = (\iota_{N_1} \otimes I_{N_2N_3T})\,, \\
D_4 &= (I_{N_1N_2} \otimes \iota_{N_3} \otimes I_T)\,,
\end{aligned}$$

all being $(N_1N_2N_3T \times N_1N_2N_3)$, $(N_1N_2N_3T \times N_1N_3T)$, $(N_1N_2N_3T \times N_2N_3T)$, and

$(N_1N_2N_3T \times N_1N_2T)$ sized matrices respectively, but now we delete, from each D_k, the rows corresponding to the missing observations to reflect the unbalanced nature of the data. The inverse of such covariance formulation can be reached in steps, that is, one has to derive

$$\Omega^{-1}\sigma_\varepsilon^2 = P^c - P^c D_4 (R^d)^{-1} D_4' P^c \tag{2.51}$$

where P^c and R^d are obtained in the following steps:

$$\begin{aligned}
R^d &= D_4' P^c D_4 + \frac{\sigma_\varepsilon^2}{\sigma_\lambda^2}, \quad P^c = P^b - P^b D_3 (R^c)^{-1} D_3' P^b, \\
R^c &= D_3' P^b D_3 + \frac{\sigma_\varepsilon^2}{\sigma_\zeta^2}, \quad P^b = P^a - P^a D_2 (R^b)^{-1} D_2' P^a, \\
R^b &= D_2' P^a D_2 + \frac{\sigma_\varepsilon^2}{\sigma_\upsilon^2}, \quad P^a = I - D_1 (R^a)^{-1} D_1', \quad \text{and } R^a = D_1' D_1 + \frac{\sigma_\varepsilon^2}{\sigma_\mu^2}.
\end{aligned}$$

Even though the calculation above alleviates some of the "dimensionality curse",[6] to perform the inverse we still have to manipulate potentially large matrices. The last step in finishing the FGLS estimation of the incomplete 4D models is to estimate the variance components. Fortunately, this is not too difficult, however, due to the size of the formulas, the results are presented in Appendix C.

2.7.2 Mixed Fixed-Random Effects Models

As briefly mentioned in Section 2.4, when one or more of the indices are small, or the effects associated with the indices should be treated as observable parameters, rather then random draws from a population, it makes more sense to treat the corresponding effects as fixed, than random.

The Idea

As an illustration, consider a linked employee-employer data set, where we usually have numerous workers and firms, but typically annual or monthly data (low T). This means that the variance of the time random effect can not be estimated consistently, and that this inconsistency carries over to $\hat{\beta}$. While firm and worker effects can still be represented by random variables, now it makes more sense to add the year (or month) effect as fixed, arriving to a mixed effects model. All in all, model (2.10) can be rewritten as

$$y_{ijt} = x_{ijt}'\beta + \lambda_t + u_{ijt} \quad \text{with} \quad u_{ijt} = \upsilon_i + \zeta_j + \varepsilon_{ijt}, \tag{2.52}$$

or similarly, with $D_1 = (\iota_{N_1} \otimes \iota_{N_2} I_T)$,

$$y = X\beta + D_1\lambda + u.$$

[6] The higher the dimension of the panel, the larger the size of the matrices we have to work with.

Notice, that λ_t is moved out from the composite disturbance term, unlike with the pure random effects model (2.10), and added as estimable dummy. We offer two consistent estimators for model (2.52), one optimal, and one sub-optimal. The optimal estimator is the (F)GLS, which jointly estimates β and λ, and unless we want to express β out from the joint estimator, is computationally simple. The sub-optimal estimator first transforms out λ with some projection orthogonal to D_1, then estimates the transformed model with (F)GLS, using the underlying transformed composite disturbance term to construct the variance-covariance matrix. As it turns out, this sub-optimal estimator is computationally not demanding at all, and also estimates β directly. Although the presence of $D_1\lambda$, specifically its perfect multicollinearity with x_t-type (that is, purely time-varying) regressors makes parameters associated with such regressors unidentifiable, $T \to \infty$ is not required any longer to reach consistent estimators. Remember, that in case of the purely random effects model (2.10), all $N_1, N_2, T \to \infty$ is needed for consistency. This mixed effects estimator is then clearly substantial improvement over the (F)GLS estimator for short panels, and in spite of leaving some parameters unidentified, consistency conditions are much easier to sustain.

The Estimators

Let us take the two estimators under lenses, and derive specific formulas for model (2.52). The GLS (optimal) estimator reads as

$$\begin{pmatrix} \hat{\beta} \\ \hat{\lambda} \end{pmatrix} = \begin{pmatrix} X'\Omega^{-1}X & X'\Omega^{-1}D_1 \\ D_1'\Omega^{-1}X & D_1'\Omega^{-1}D_1 \end{pmatrix}^{-1} \begin{pmatrix} X'\Omega^{-1}y \\ D_1'\Omega^{-1}y \end{pmatrix},$$

with $\Omega = \mathrm{E}(uu')$, or expressed for $\hat{\beta}$ by partialling out $\hat{\lambda}$:

$$\hat{\beta} = \left(X'\Omega^{-1/2} M_{\Omega^{-1/2}D_1} \Omega^{-1/2} X \right)^{-1} X'\Omega^{-1/2} M_{\Omega^{-1/2}D_1} \Omega^{-1/2} y. \tag{2.53}$$

As both Ω and D_1 are constructed using elementary matrices, lengthy calculations lead to

$$\hat{\beta} = \left(X'\Omega^{-1} - (\bar{J}_{N_1N_2} \otimes I_T)\Omega^{-1}X \right)^{-1} X'\Omega^{-1} - (\bar{J}_{N_1N_2} \otimes I_T)\Omega^{-1}y,$$

which is the same as running an OLS on the transformed variables

$$\tilde{y}_{ijt} = \left(y_{ijt} - \bar{y}_{..t} + \sqrt{\varphi_1}\bar{y}_{.j.} + \sqrt{\varphi_2}\bar{y}_{i..} - \sqrt{(\varphi_1 + \varphi_2)}\bar{y}_{...} \right) \tag{2.54}$$

with

$$\varphi_1 = \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + N_1 T \sigma_\zeta^2} - 1; \quad \varphi_2 = \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + N_2 T \sigma_\upsilon^2} - 1,$$

similarly for X.

The suboptimal estimator first transforms out $D_1\lambda$ with the orthogonal projection M_{D_1}, then uses the covariance structure of the transformed model to construct the (F)GLS estimator for β. Formally, (F)GLS is to be performed on the transformed model

$$M_{D_1}y = M_{D_1}X\beta + M_{D_1}u \quad \text{or} \quad \tilde{y}_{ijt} = \tilde{x}'_{ijt}\beta + \tilde{u}_{ijt} \quad \text{with} \quad \tilde{y}_{ijt} = (y_{ijt} - \bar{y}_{..t})$$

which now clearly is a pure random effects model with variance-covariance matrix

$$\begin{aligned}\tilde{\Omega} = \mathrm{E}(\tilde{u}\tilde{u}') &= I\sigma^2_\varepsilon + (I_{N_1} \otimes \bar{J}_{N_2T})N_2T\sigma^2_\alpha + (\bar{J}_{N_1} \otimes I_{N_2} \otimes \bar{J}_T)N_1T\sigma^2_\gamma \\ &\quad -(\bar{J}_{N_1N_2} \otimes I_T)\sigma^2_\varepsilon - \bar{J}_{N_1N_2T}(N_2T\sigma^2_\alpha + N_1T\sigma^2_\gamma)\,.\end{aligned}$$

In spite of this covariance matrix being slightly less trivial, as it turns out, tricks used in Section 2.2 can be also used here to analytically derive its inverse:

$$\begin{aligned}\tilde{\Omega}^{-1} &= I - (\bar{J}_{N_1N_2} \otimes I_T) + \varphi_1(I_{N_1} \otimes \bar{J}_{N_2T}) \\ &\quad + \varphi_2(\bar{J}_{N_1} \otimes I_{N_2} \otimes \bar{J}_T) - (\varphi_1 + \varphi_2)\bar{J}_{N_1N_2T}\,.\end{aligned}$$

The GLS is then identical to run a Least Squares regression with $\tilde{\tilde{x}}'_{ijt}$ on $\tilde{\tilde{y}}_{ijt}$ where

$$\tilde{\tilde{y}}_{ijt} = y_{ijt} - \bar{y}_{..t} + \sqrt{\vartheta_1}\bar{y}_{i..} + \sqrt{\vartheta_2}\bar{y}_{.j.} - (\sqrt{\vartheta_1} + \sqrt{\vartheta_2})\bar{y}_{...}\,, \tag{2.55}$$

similarly for X.

The results are more than interesting. The optimal and sub-optimal data transformations (2.54) and (2.55) to be employed on the mixed effects model (2.52) are almost identical, the only difference is the coefficient of the overall mean of y: $\sqrt{\vartheta_1 + \vartheta_2}$ in the optimal case, and $\sqrt{\vartheta_1} + \sqrt{\vartheta_2}$ for the sub-optimal estimator. This suggests that the two estimators will fall close to each other, and that information held in the data will most likely outweigh this slight difference in the corresponding estimators. Further, both transformations are in fact the adjusted Within transformations, where the structure of the variance-covariance matrix is taken into account. It easy to see, that if $N_1, N_2 \to \infty$, the data transformations uniformly converge to the Within transformation to be employed on model (2.10) with fixed effects.

Alternative Scenarios and Incomplete Data

Table 2.13 summarizes alternative mixed effects scenarios for model (2.10) with their corresponding data transformations both for the optimal and for the sub-optimal estimator.

The last question concerns with the extensions of the above estimators for unbalanced data. As detailed in Section 2.4, the covariance matrix now can not be represented with the kronecker product of elementary matrices, and so its potential inverse might be costly to calculate. Further, expressing $\hat{\beta}$ as in (2.53) from the

Table 2.13 *Mixed effects model formulations*

FE[a]	RE	Optimal transformation	Sub-optimal transformation
υ_i	$\zeta_j+\lambda_t$	$y_{ijt}-\bar{y}_{i..}+\sqrt{\varphi_2}\bar{y}_{.j.}+\sqrt{\varphi_3}\bar{y}_{..t}-\sqrt{\varphi_1+\varphi_2}\bar{y}_{...}$	$y_{ijt}-\bar{y}_{i..}+\sqrt{\varphi_2}\bar{y}_{.j.}+\sqrt{\varphi_3}\bar{y}_{..t}-(\sqrt{\varphi_1}+\sqrt{\varphi_2})\bar{y}_{...}$
ζ_j	$\upsilon_i+\lambda_t$	$y_{ijt}+\sqrt{\varphi_1}\bar{y}_{i..}-\bar{y}_{.j.}+\sqrt{\varphi_3}\bar{y}_{..t}-\sqrt{\varphi_1+\varphi_3}\bar{y}_{...}$	$y_{ijt}+\sqrt{\varphi_1}\bar{y}_{i..}-\bar{y}_{.j.}+\sqrt{\varphi_3}\bar{y}_{..t}-(\sqrt{\varphi_1}+\sqrt{\varphi_3})\bar{y}_{...}$
λ_t	$\upsilon_i+\zeta_j$	$y_{ijt}+\sqrt{\varphi_1}\bar{y}_{i..}+\sqrt{\varphi_2}\bar{y}_{.j.}-\bar{y}_{..t}-\sqrt{\varphi_1+\varphi_2}\bar{y}_{...}$	$y_{ijt}+\sqrt{\varphi_1}\bar{y}_{i..}+\sqrt{\varphi_2}\bar{y}_{.j.}-\bar{y}_{..t}-(\sqrt{\varphi_1}+\sqrt{\varphi_2})\bar{y}_{...}$
$\upsilon_i+\zeta_j$	λ_t	$y_{ijt}-\bar{y}_{i..}-\bar{y}_{.j.}+\sqrt{\varphi_3}\bar{y}_{..t}-\sqrt{1+\varphi_3}\bar{y}_{...}$	$y_{ijt}-\bar{y}_{i..}-\bar{y}_{.j.}+\sqrt{\varphi_3}\bar{y}_{..t}-(1+\sqrt{\varphi_3})\bar{y}_{...}$
$\upsilon_i+\lambda_t$	ζ_j	$y_{ijt}-\bar{y}_{i..}+\sqrt{\varphi_2}\bar{y}_{.j.}-\bar{y}_{..t}-\sqrt{1+\varphi_2}\bar{y}_{...}$	$y_{ijt}-\bar{y}_{i..}+\sqrt{\varphi_2}\bar{y}_{.j.}-\bar{y}_{..t}-(1+\sqrt{\varphi_2})\bar{y}_{...}$
$\zeta_j+\lambda_t$	υ_i	$y_{ijt}+\sqrt{\varphi_1}\bar{y}_{i..}-\bar{y}_{.j.}-\bar{y}_{..t}-\sqrt{1+\varphi_1}\bar{y}_{...}$	$y_{ijt}+\sqrt{\varphi_1}\bar{y}_{i..}-\bar{y}_{.j.}-\bar{y}_{..t}-(1+\sqrt{\varphi_1})\bar{y}_{...}$

[a] Note: the pure random effects and pure fixed effects models are excluded from the table for obvious reasons. φ_1 and φ_2 are already defined, in a similar manner, $\varphi_3=\sigma_\varepsilon^2/(N_1N_2\sigma_\lambda^2+\sigma_\varepsilon^2)$.

joint estimator can not be done that easily, as a result, data transformation (2.54) is subject to (potentially computationally burdening) matrix operations.

In case of the sub-optimal estimator, some fixed effects transformations, like the one we used, $y_{ijt}-\bar{y}_{..t}$, are still suitable if the data is incomplete. This is not general, unfortunately, as argued in Chapter 1, and such Within transformations often involve matrix operations to some extent. Nevertheless, the sub-optimal estimator still seems to perform better computationally, especially with mixed models of two effects, one fixed and one random. To see this, consider that transforming out a single fixed effect can always be done with scalar operations, just as the inverse of a variance-covariance matrix with a single random effect, regardless of the form of incompleteness.

2.8 Conclusion

When observations can be considered as samples from an underlying population, random effects specifications seem to be more suited to deal with multi-dimensional data sets. FGLS estimators for three-way error components models are almost as easily obtained as for the traditional 2D panel models (especially the spectral decomposition of the covariance matrices and estimation of the variances of the random effects), however the resulting asymptotic requirements for their consistency are more peculiar. In fact, now the data can grow in three directions, and only some of the asymptotic cases are sufficient for consistency. Interestingly, for some error components specifications, consistency implies the convergence of the FGLS estimator to the Within estimator. This is utterly important, as under the Within estimation, the parameters of some fixed regressors are unidentified, which is in fact carried over to the FGLS estimation of those parameters as well. To solve this, we have shown that a simple OLS can be sufficient to get the full set of parameter estimates (of course, bearing the price of inefficiency), wherever this identification problem persists. As the strict exogeneity assumption in these multi-dimensional

10.14754/CEU.2017.05

random effects models might fail, the proposed FGLS estimators are biased and inconsistent, and the results are of no practical relevance. To overcome this, HT-type IV estimators are proposed which, under reasonable assumptions, yields consistent estimators for all model parameters. Some insights on testing are also considered, easing the decision between different model specifications (Fisher's ANOVA test), testing for endogeneity (Hausman test), or testing for instrument validity. Lastly, estimation issues with mixed effects models – models with both fixed and random effects – are also taken into account. The main results of the chapter are also extended to treat incomplete data and towards higher dimensions.

Appendix A – Rationale Behind the Normalization Factors

A.1 Example for Normalizing with 1: Model (2.12), $T \to \infty$

$$\begin{aligned} \text{plim}_{T\to\infty} V(\hat{\beta}_{OLS}) &= \text{plim}_{T\to\infty}(X'X)^{-1}X'\Omega X(X'X)^{-1} \\ &= \text{plim}_{T\to\infty}\left(\tfrac{X'X}{T}\right)^{-1}\tfrac{X'\Omega X}{T^2}\left(\tfrac{X'X}{T}\right)^{-1}. \end{aligned}$$

We assume, that $\text{plim}_{T\to\infty} X'X/T = Q_{XX}$ is a finite, positive definite matrix, and further, we use that $\Omega = \sigma_\varepsilon^2 I_{N_1N_2T} + \sigma_\mu^2(I_{N_1N_2} \otimes J_T)$. With this,

$$\text{plim}_{T\to\infty} V(\hat{\beta}_{OLS}) = Q_{XX}^{-1} \cdot \text{plim}_{T\to\infty} \frac{\sigma_\varepsilon^2 X'X + \sigma_\mu^2 X'(I_{N_1N_2} \otimes J_T)X}{T^2} \cdot Q_{XX}^{-1},$$

where we know, that $\text{plim}_{T\to\infty} \dfrac{\sigma_\varepsilon^2 X'X}{T^2} = 0$, and we assume, that

$$\text{plim}_{T\to\infty} \frac{\sigma_\mu^2 X'(I_{N_1N_2} \otimes J_T)X}{T^2} = Q_{XBX}$$

is a finite, positive definite matrix. Then the variance is finite, and takes the form

$$\text{plim}_{T\to\infty} V(\hat{\beta}_{OLS}) = Q_{XX}^{-1} \cdot Q_{XBX} \cdot Q_{XX}^{-1}.$$

Notice, that we can arrive to the same result by first normalizing with the usual $\sqrt{T}$ term, and then adjusting it with $1/\sqrt{T}$ to arrive to a non-zero, but bounded variance:

$$\begin{aligned} \text{plim}_{T\to\infty} V(\sqrt{T}\hat{\beta}_{OLS}) &= \text{plim}_{T\to\infty} T(X'X)^{-1}X'\Omega X(X'X)^{-1} \\ &= \text{plim}_{T\to\infty}\left(\tfrac{X'X}{T}\right)^{-1}\tfrac{X'\Omega X}{T}\left(\tfrac{X'X}{T}\right)^{-1}, \end{aligned}$$

which grows at $O(T)$ because of $\frac{X'\Omega X}{T}$. We have to correct for it with the $1/\sqrt{T}$ factor, leading to the overall normalization factor $\sqrt{T}/\sqrt{T} = 1$. The reasoning is the similar for all other cases and other models.

A.2 Example for Normalizing with $\sqrt{N_1N_2}/A$*: Model (2.2),* $N_1, N_2 \to \infty$

Using the standard $\sqrt{N_1N_2}$ normalization factor gives

$$\begin{aligned} \text{plim}_{N_1,N_2\to\infty} \text{Var}(\sqrt{N_1N_2}\hat{\beta}_{OLS}) &= \text{plim}_{N_1,N_2\to\infty} N_1N_2 \cdot (X'X)^{-1}X'\Omega X(X'X)^{-1} \\ &= \text{plim}_{N_1,N_2\to\infty}\left(\tfrac{X'X}{N_1N_2}\right)^{-1}\tfrac{X'\Omega X}{N_1N_2}\left(\tfrac{X'X}{N_1N_2}\right)^{-1} \\ &= Q_{XX}^{-1} \cdot \text{plim}_{N_1,N_2\to\infty}\tfrac{X'\Omega X}{N_1N_2} \cdot Q_{XX}^{-1}, \end{aligned}$$

where we assumed, that $\text{plim}_{N_1,N_2\to\infty} X'X/N_1N_2 = Q_{XX}$, is a positive definite, finite matrix. Further, we use, that

$$\Omega = \sigma_\varepsilon^2 I_{N_1N_2T} + \sigma_\mu (I_{N_1N_2} \otimes J_T) + \sigma_\upsilon^2 (I_{N_1} \otimes J_{N_2} \otimes I_T) + \sigma_\zeta^2 (J_{N_1} \otimes I_{N_2T}) \,.$$

Observe, that

$$\begin{aligned}\text{plim}_{N_1,N_2\to\infty} \quad \tfrac{X'\Omega X}{N_1N_2} &= \text{plim}_{N_1,N_2\to\infty} \tfrac{\sigma_\varepsilon^2 X'X}{N_1N_2} + \text{plim}_{N_1,N_2\to\infty} \tfrac{\sigma_\mu X'(I_{N_1N_2}\otimes J_T)X}{N_1N_2} \\ &+ \text{plim}_{N_1,N_2\to\infty} \tfrac{\sigma_\upsilon^2 X'(I_{N_1}\otimes J_{N_2}\otimes I_T)X}{N_1N_2} + \text{plim}_{N_1,N_2\to\infty} \tfrac{\sigma_\zeta^2 X'(J_{N_1}\otimes I_{N_2T})X}{N_1N_2}\end{aligned} \tag{A.56}$$

is an expression where the first two terms are finite, but the third grows with $O(N_2)$ (because of J_{N_2}), and the last with $O(N_1)$ (because of J_{N_1}), which in turns yields unbounded variance of $\hat{\beta}_{OLS}$. To obtain a finite variance, we have to normalize the variance additionally with either $1/\sqrt{N_1}$ or $1/\sqrt{N_2}$, depending on which grows faster. Let us assume, without loss of generality, that N_1 grows at a higher rate ($A = N_1$). In this way, the effective normalization factor is $\frac{\sqrt{N_1N_2}}{\sqrt{A}} = \frac{\sqrt{N_1N_2}}{\sqrt{N_1}} = \sqrt{N_2}$, under which the first three plim terms in (A.56) are zero, but the fourth is finite:

$$\text{plim}_{N_1,N_2\to\infty} \frac{\sigma_\zeta^2 X'(J_{N_1}\otimes I_{N_2T})X}{N_1^2 N_2} = Q_{XBX} \,,$$

with some Q_{XBX} finite, positive definite matrix. The same reasoning holds for other models and other asymptotics as well.

Appendix B – Proof of Formula (2.17)

Let us make the proof only for model (2.2) (so formula (2.17)), the rest is just direct application of the derivation below. The outline of the proof is based on Wansbeek and Kapteyn (1989).

First, notice, that using the Woodbury matrix identity,

$$\begin{aligned}(P^a)^{-1} &= \left(I - D_1(D_1'D_1 + I\tfrac{\sigma_\varepsilon^2}{\sigma_\mu^2})^{-1}D_1\right)^{-1} \\ &= I + D_1\left(D_1'D_1 + I\tfrac{\sigma_\varepsilon^2}{\sigma_\mu^2} - D_1'D_1\right)^{-1} D_1' \\ &= I + \tfrac{\sigma_\mu^2}{\sigma_\varepsilon^2} D_1 D_1'\end{aligned}$$

Second, using that

$$D_2'P^aD_2 = D_2'D_2 - D_2'D_1(R^a)^{-1}D_1'D_2 = R^b - \frac{\sigma_\varepsilon^2}{\sigma_\upsilon^2} I$$

gives

$$R^b - D_2' P^a D_2 = \frac{\sigma_\varepsilon^2}{\sigma_\upsilon^2} I \,.$$

Using the Woodbury matrix identity for the second time,

$$\begin{aligned}
(P^b)^{-1} &= \left(P^a - P^a D_2 (R^b)^{-1} D_2' P^a\right)^{-1} \\
&= (P^a)^{-1} + (P^a)^{-1} P^a D_2 \left(R^b - D_2' P^a (P^a)^{-1} P^a D_2\right)^{-1} D_2' P^a (P^a)^{-1} \\
&= (P^a)^{-1} + D_2 \left(R^b - D_2' P^a D_2\right)^{-1} D_2' = (P^a)^{-1} + D_2 \left(\tfrac{\sigma_\varepsilon^2}{\sigma_\upsilon^2} I\right)^{-1} D_2' \\
&= I + \tfrac{\sigma_\mu^2}{\sigma_\varepsilon^2} D_1 D_1' + \tfrac{\sigma_\upsilon^2}{\sigma_\varepsilon^2} D_2 D_2' \,.
\end{aligned}$$

Now we are almost there, we only have to repeat the last step one more time. That is,

$$D_3' P^b D_3 = D_3' D_3 - D_3' D_2 (R^b)^{-1} D_2' D_3 = R^c - \frac{\sigma_\varepsilon^2}{\sigma_\zeta^2} I \quad \text{gives} \quad R^c - D_3' P^b D_3 = \frac{\sigma_\varepsilon^2}{\sigma_\zeta^2} I \,.$$

again, and so

$$\begin{aligned}
\left(\Omega^{-1} \sigma_\varepsilon^2\right)^{-1} &= \left(P^b - P^b D_3 (R^c)^{-1} D_3' P^b\right)^{-1} \\
&= (P^b)^{-1} + (P^b)^{-1} P^b D_3 \left(R^c - D_3' P^b (P^b)^{-1} P^b D_3\right)^{-1} D_3' P^b (P^b)^{-1} \\
&= (P^b)^{-1} + D_3 \left(R^c - D_3' P^b D_3\right)^{-1} D_3' = (P^b)^{-1} + D_3 \left(\tfrac{\sigma_\varepsilon^2}{\sigma_\zeta^2} I\right)^{-1} D_3' \\
&= I + \tfrac{\sigma_\mu^2}{\sigma_\varepsilon^2} D_1 D_1' + \tfrac{\sigma_\upsilon^2}{\sigma_\varepsilon^2} D_2 D_2' + \tfrac{\sigma_\zeta^2}{\sigma_\varepsilon^2} D_3 D_3' = \Omega \sigma_\varepsilon^{-2} \,.
\end{aligned}$$

Appendix C – Inverse of (2.49), and the Estimation of the Variance Components

$$\begin{aligned}
\sigma_\varepsilon^2 \Omega^{-1} = \; & I_{N_1 N_2 N_3 T} - (1-\theta_{20})(J_{N_1} \otimes I_{N_2 N_3 T}) - (1-\theta_{21})(I_{N_1} \otimes J_{N_2} \otimes I_{N_3 T}) \\
& -(1-\theta_{22})(I_{N_1 N_2} \otimes J_{N_3} \otimes I_T) - (1-\theta_{23})(I_{N_1 N_2 N_3} \otimes J_T) \\
& +(1-\theta_{24})(J_{N_1 N_2} \otimes I_{N_3 T}) + (1-\theta_{25})(J_{N_1} \otimes I_{N_2} \otimes J_{N_3} \otimes I_T) \\
& +(1-\theta_{26})(J_{N_1} \otimes I_{N_2 N_3} \otimes J_T) + (1-\theta_{27})(I_{N_1} \otimes J_{N_2 N_3} \otimes I_T) \\
& +(1-\theta_{28})(I_{N_1} \otimes J_{N_2} \otimes I_{N_3} \otimes J_T) + (1-\theta_{29})(I_{N_1 N_2} \otimes J_{N_3 T}) \\
& -(1-\theta_{30})(J_{N_1 N_2 N_3} \otimes I_T) - (1-\theta_{31})(J_{N_1 N_2} \otimes I_{N_3} \otimes J_T) \\
& -(1-\theta_{32})(J_{N_1} \otimes I_{N_2} \otimes J_{N_3 T}) - (1-\theta_{33})(I_{N_1} \otimes J_{N_2 N_3 T}) \\
& +(1-\theta_{34}) J_{N_1 N_2 N_3 T}
\end{aligned}$$

with

$$
\begin{aligned}
\theta_{20} &= \tfrac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+N_1\sigma_\zeta} \quad \theta_{21} = \tfrac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+N_2\sigma_\upsilon} \quad \theta_{22} = \tfrac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+N_3\sigma_\lambda} \quad \theta_{23} = \tfrac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+T\sigma_\mu} \\
\theta_{24} &= \theta_{20}+\theta_{21}-\tfrac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+N_1\sigma_\zeta^2+N_2\sigma_\upsilon^2} \quad \theta_{25} = \theta_{20}+\theta_{22}-\tfrac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+N_1\sigma_\zeta^2+N_3\sigma_\lambda^2} \\
\theta_{26} &= \theta_{20}+\theta_{23}-\tfrac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+N_1\sigma_\zeta^2+T\sigma_\mu^2} \quad \theta_{27} = \theta_{21}+\theta_{22}-\tfrac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+N_2\sigma_\upsilon^2+N_3\sigma_\lambda^2} \\
\theta_{28} &= \theta_{21}+\theta_{23}-\tfrac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+N_2\sigma_\upsilon^2+T\sigma_\mu^2} \quad \theta_{29} = \theta_{22}+\theta_{23}-\tfrac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+N_3\sigma_\lambda^2+T\sigma_\mu^2} \\
\theta_{30} &= \theta_{24}+\theta_{25}+\theta_{27}-\theta_{20}-\theta_{21}-\theta_{22}+\tfrac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+N_1\sigma_\zeta^2+N_2\sigma_\upsilon^2+N_3\sigma_\lambda^2} \\
\theta_{31} &= \theta_{24}+\theta_{26}+\theta_{28}-\theta_{20}-\theta_{21}-\theta_{23}+\tfrac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+N_1\sigma_\zeta^2+N_2\sigma_\upsilon^2+T\sigma_\mu^2} \\
\theta_{32} &= \theta_{25}+\theta_{26}+\theta_{29}-\theta_{20}-\theta_{22}-\theta_{23}+\tfrac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+N_1\sigma_\zeta^2+N_3\sigma_\lambda^2+T\sigma_\mu^2} \\
\theta_{33} &= \theta_{27}+\theta_{28}+\theta_{29}-\theta_{21}-\theta_{22}-\theta_{23}+\tfrac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+N_2\sigma_\upsilon^2+N_3\sigma_\lambda^2+T\sigma_\mu^2} \\
\theta_{34} &= \theta_{20}+\theta_{21}+\theta_{22}+\theta_{23}-\theta_{24}-\theta_{25}-\theta_{26}-\theta_{27}-\theta_{28}-\theta_{29} \\
&\quad +\theta_{30}+\theta_{31}+\theta_{32}+\theta_{33}-\tfrac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+N_1\sigma_\zeta^2+N_2\sigma_\upsilon^2+N_3\sigma_\lambda^2+T\sigma_\mu^2}\,.
\end{aligned}
$$

The estimation of the variance components in case of complete data are as follows:

$$
\begin{aligned}
\hat{\sigma}_\varepsilon^2 &= \tfrac{1}{(N_1-1)(N_2-1)(N_3-1)(T-1)} \textstyle\sum_{ijst} \tilde{\hat{u}}_{ijst}^2 \\
\hat{\sigma}_\mu^2 &= \tfrac{1}{(N_1-1)(N_2-1)(N_3-1)T} \textstyle\sum_{ijst} (\tilde{\hat{u}}_{ijst}^a)^2 - \hat{\sigma}_\varepsilon^2 \\
\hat{\sigma}_\upsilon^2 &= \tfrac{1}{(N_1-1)N_2(N_3-1)(T-1)} \textstyle\sum_{ijst} (\tilde{\hat{u}}_{ijst}^b)^2 - \hat{\sigma}_\varepsilon^2 \\
\hat{\sigma}_\zeta^2 &= \tfrac{1}{N_1(N_2-1)(N_3-1)(T-1)} \textstyle\sum_{ijst} (\tilde{\hat{u}}_{ijst}^c)^2 - \hat{\sigma}_\varepsilon^2 \\
\hat{\sigma}_\lambda^2 &= \tfrac{1}{(N_1-1)(N_2-1)N_3(T-1)} \textstyle\sum_{ijst} (\tilde{\hat{u}}_{ijst}^d)^2 - \hat{\sigma}_\varepsilon^2\,,
\end{aligned}
$$

where, as before, $\hat{u}_{ijst}$ is the OLS residual, and

$$
\begin{aligned}
\tilde{u}_{ijst} = \; & u_{ijst}-\bar{u}_{ijs.}-\bar{u}_{ij.t}-\bar{u}_{i.st}-\bar{u}_{.jst}+\bar{u}_{ij..}+\bar{u}_{i.s.}+\bar{u}_{.js.} \\
& +\bar{u}_{i..t}+\bar{u}_{.j.t}+\bar{u}_{..st}-\bar{u}_{i...}-\bar{u}_{.j..}-\bar{u}_{..s.}-\bar{u}_{...t}+\bar{u}_{....} \\
\tilde{u}_{ijst}^a = \; & u_{ijst}-\bar{u}_{ij.t}-\bar{u}_{i.st}-\bar{u}_{.jst}+\bar{u}_{i..t}+\bar{u}_{.j.t}+\bar{u}_{..st}-\bar{u}_{...t} \\
\tilde{u}_{ijst}^b = \; & u_{ijst}-\bar{u}_{ijs.}-\bar{u}_{ij.t}-\bar{u}_{.jst}+\bar{u}_{ij..}+\bar{u}_{.js.}+\bar{u}_{.j.t}-\bar{u}_{.j..} \\
\tilde{u}_{ijst}^c = \; & u_{ijst}-\bar{u}_{ijs.}-\bar{u}_{ij.t}-\bar{u}_{i.st}+\bar{u}_{ij..}+\bar{u}_{i.s.}+\bar{u}_{i..t}-\bar{u}_{i...} \\
\tilde{u}_{ijst}^d = \; & u_{ijst}-\bar{u}_{ijs.}-\bar{u}_{i.st}-\bar{u}_{.jst}+\bar{u}_{i.s.}+\bar{u}_{.js.}+\bar{u}_{..st}-\bar{u}_{..s.}\,.
\end{aligned}
$$

Estimation of the variance components in case of incomplete data yields

$$
\begin{aligned}
\hat{\sigma}_\mu^2 &= \tfrac{1}{\sum_{ijs}|T_{ijs}|} \textstyle\sum_{ijst} \hat{u}_{ijst}^2 - \tfrac{1}{\bar{n}_{ijs}} \sum_{ijs} \tfrac{1}{|T_{ijs}|-1} \sum_t (\tilde{\hat{u}}_{ijst}^a)^2 \\
\hat{\sigma}_\upsilon^2 &= \tfrac{1}{\sum_{ijs}|T_{ijs}|} \textstyle\sum_{ijst} \hat{u}_{ijst}^2 - \tfrac{1}{\bar{n}_{ist}} \sum_{ist} \tfrac{1}{n_{ist}-1} \sum_j (\tilde{\hat{u}}_{ijst}^b)^2 \\
\hat{\sigma}_\zeta^2 &= \tfrac{1}{\sum_{ijs}|T_{ijs}|} \textstyle\sum_{ijst} \hat{u}_{ijst}^2 - \tfrac{1}{\bar{n}_{jst}} \sum_{jt} \tfrac{1}{n_{jst}-1} \sum_i (\tilde{\hat{u}}_{ijst}^c)^2 \\
\hat{\sigma}_\lambda^2 &= \tfrac{1}{\sum_{ijs}|T_{ijs}|} \textstyle\sum_{ijst} \hat{u}_{ijst}^2 - \tfrac{1}{\bar{n}_{ijt}} \sum_{ijt} \tfrac{1}{n_{ijt}-1} \sum_s (\tilde{\hat{u}}_{ijst}^d)^2 \\
\hat{\sigma}_\varepsilon^2 &= \tfrac{1}{\sum_{ijs}|T_{ijs}|} \textstyle\sum_{ijst} \hat{u}_{ijst}^2 - \hat{\sigma}_\mu^2 - \hat{\sigma}_\upsilon^2 - \hat{\sigma}_\zeta^2 - \hat{\sigma}_\lambda^2\,,
\end{aligned}
\tag{C.57}
$$

where $\hat{u}_{ijst}$ are the OLS residuals, and $\tilde{\hat{u}}^k_{ijst}$ are its transformations ($k = a, b, c, d$) according to

$$\begin{aligned} \tilde{u}^a_{ijst} &= u_{ijst} - \tfrac{1}{|T_{ijs}|} \textstyle\sum_t u_{ijst} \,, \quad \tilde{u}^b_{ijst} = u_{ijst} - \tfrac{1}{n_{ist}} \textstyle\sum_j u_{ijst} \,, \\ \tilde{u}^c_{ijst} &= u_{ijst} - \tfrac{1}{n_{jst}} \textstyle\sum_i u_{ijst} \,, \quad \tilde{u}^d_{ijst} = u_{ijst} - \tfrac{1}{n_{ijt}} \textstyle\sum_s u_{ijst} \,. \end{aligned}$$

Further, $|T_{ijs}|$, $\tilde{n}_{ist}$, $\tilde{n}_{jst}$, and $\tilde{n}_{ijt}$ denote the total number of observations for a given (ijs), (ist), (jst), and (ijt) pair respectively, and finally, $\tilde{n}_{ijs}$, $\tilde{n}_{ist}$, $\tilde{n}_{jst}$, and $\tilde{n}_{ijt}$ are the total number of unique (ijs), (ist), (jst), and (ijt) observations in the data.

10.14754/CEU.2017.05

10.14754/CEU.2017.05

3

The Estimation of Varying Coefficients Multi-dimensional Panel Data Models

3.1 Introduction

Over the last few years there has been an explosion in the amount of data available for economic analysis. Many of these data sets present themselves in the form of multi-dimensional panels and are used to study phenomena like international trade, capital flows between countries or regions, exchange rates between multiple currencies, and so on. Several fixed effects model specifications (along with some random effects specifications) have been put forward in the literature and used in practice to deal with these types of observations. All these specifications, however, formalize the individual and time heterogeneity of the data through simple (but multiple, due to the multi-dimensionality of the panels) individual and time effects (let them be fixed or random, see Chapters 1 and 2, and further see, for example, Egger and Pfaffermayr, 2003, Baltagi et al., 2003, Baldwin and Taglioni, 2006, Baier and Bergstrand, 2007, Matyas, 1997, Matyas, 1998, Ghosh, 1976). This in fact means, that the heterogeneity is captured through the intercept parameter in case of fixed effects (*i.e.*, via the shifts of the intercept for different individuals and time points), and is captured through unobserved random variables in case of random effects.

One of the most important statistical features of these data sets is, however, that heterogeneity is likely to take more complicated forms (like clustering, for example) which begs for more complex econometric models. Accounting for omitted variables through individual and/or time effects (or with their interactions) is not always enough. In many cases changes in economic structures or factors imply that the slope parameters may be different across entities or vary over time periods. Two of the most famous examples in the literature to promote varying coefficient models are Kuh (1963) and Swamy (1970). Kuh's study on the investment expenditure of 60 small and medium-sized firms rejects the joint hypothesis of common intercept and slope parameters for all firms, but more interestingly, it soundly rejects the hypothesis of a variable intercept and common slope parameters (that is, fixed effects models). Similarly, when Swamy undertook his project on 11 US corporations to fit the Grunfeld (1958) investment functions, the test for varying intercept parameters and common slope coefficient was strongly rejected.

In such cases tests suggest that common slopes and variable intercepts are unsatisfactory in explaining why some effects differ across individuals and/or time periods, and it makes sense to allow the slope parameters to vary across individuals and/or over time. Such formulation in its most general form can be constructed as, with $i = 1, \ldots, N_1$, $j = 1, \ldots, N_2$, and $t = 1, \ldots, T$,

$$y_{ijt} = x'_{ijt}\beta_{ijt} + \varepsilon_{ijt} \tag{3.1}$$

where y_{ijt} is the dependent variable corresponding to observation (ijt), x'_{ijt} is a $(1 \times K)$ vector of regressors, β_{ijt} is the $(K \times 1)$ vector of the slope parameters, and ε_{ijt} is assumed to be an idiosyncratic white noise disturbance term.[1]

If β_{ijt} are considered as fixed parameters, (3.1) obviously can not be regressed, as the number of slope parameters to be estimated (KN_1N_2T) well exceed the number of observations (N_1N_2T). One possibility to reduce the number of coefficients is to allow some structure on β_{ijt}, something I will encounter in Section 3.2. Alternatively, by assuming that β_{ijt} is coming from a probability distribution with some common mean:

$$\beta_{ijt} = \beta + \mu_{ijt} \, ,$$

the number of parameters to be estimated reduces highly. In this context β is estimated with GLS or Maximum Likelihood techniques and the zero mean μ_{ijt} is called the random coefficients. Giving some structure to μ_{ijt} and assuming that it is in fact well-behaved, K model parameters and some $(K \times K)$ variance-covariance matrices are to be estimated.

This latter case is called the *random coefficients* approach, something I can not take into account here due to size considerations.[2]

The choice between the two modelling frameworks, however, is usually not up to the researcher, but, as Hsiao (2003) puts it, depends on whether β_{ijt} is coming from a heterogeneous population, or is viewed as a random draw from a common population, and whether we are making inferences conditional on the individual/time characteristics, or making unconditional inferences on the population characteristics. In the former case, β_{ijt} should be considered as fixed, estimable parameter (after giving proper structures to reduce the overall number of slope parameters), invoking the fixed varying coefficient approach, whereas in the latter, the random

[1] It now makes sense not to treat the intercept parameter differently from the slope parameters (unlike in Chapters 1 and 2), and so to include it into x'_{ijt} by setting $x^1_{ijt} = 1$.

[2] The interested reader can find an extensive, fresh-from-the-oven analysis of multi-dimensional random coefficients models in Krishnakumar et al. (2017).

coefficients approach is more feasible.[3] Further, fixed- or random coefficients are not necessarily exclusive: mixed effects models, comprising both fixed and random slopes, are naturally viewed as general cases of pure fixed-coefficients or pure random-coefficients models. Mixed effects models are briefly visited in Section 3.3.3.

It is not always clear, however, which (the fixed- or the random coefficients) approach is more suitable economically. In such cases, we have to keep in mind the advantages each approach may offer, and make decisions accordingly. The fixed coefficient framework (as it is based on a standard analysis of variance approach) is computationally much simpler (if we ignore the mere size of the data for a minute), moreover, no distributional assumptions are needed for the underlying slope parameters. The drawback of the fixed coefficients framework is twofold: (i) as the differences between individuals and/or time periods are fixed and different (Hsiao and Pesaran, 2008), meaningful inferences on the population are hard to make, and (ii) the number of parameters to be estimated is radically large, as compared to random coefficients models, where estimates are only required for certain $(K \times K)$ variance-covariance matrices.

In this study I extend the well-known two-dimensional (2D) fixed coefficient model (Balestra and Krishnakumar, 2008, pp. 41–44, and Hsiao, 2003, pp. 138–140) to higher dimensions together with some extensions of the basic results, and, at the same time, I get some new insights into the 2D case by proposing a widely generalizable estimation methodology. Fixed coefficients models are almost completely absent from the literature, except for Hsiao's somewhat light theoretical foundations and a few applications of random coefficients models where fixed-coefficient-like joint estimators of β and a *single* individual-varying parameter are derived (Arellano and Bonhomme, 2012). To the best of my knowledge, no higher-dimensional models, and more importantly, no models with *multiple* varying coefficients have been considered properly in the literature, most probably due to the excessive number of parameters to estimate and due to the incomplete theoretical background. The contribution of the chapter is twofold. One, new fixed coefficients models are proposed, and proper estimators are derived which can readily be used in practice, further, can be traced back to the incomplete 2D results (like estimator (3.17) of model (3.16)). Two, using the so-called *Least Squares of no-full rank* estimator (which is a principal result of algebra/statistics), as we will see in Section 3.2, is a complete novelty in the fixed-coefficients context, and beneficial for at least three reasons: (i) Easy to apply for any fixed coefficients model regardless of parameter structure or the number of dimensions; (ii) flexibly handles multi-

[3] Hereafter, the '*fixed coefficient*' terminology will be used to refer to fixed varying coefficient models. The 'fixed' term should not confuse the reader: the slope coefficients are varying, only they are not drawn randomly from a distribution.

ple forms of identifying parameter restrictions; (iii) naturally alleviates part of the dimensionality issue if the data is large.

The chapter is structured as follows: Section 3.2 presents different types of fixed coefficients formulations for three-dimensional panels, and derives appropriate Least Squares estimators for each of them. Extensions to unbalanced data are also considered here. Section 3.3 discusses two important extensions of the main results: the case when fixed coefficients are functions of the observables, and the case of variables with index deficiencies: variables which potentially do not vary in all three directions. Section 3.4 takes a brief look at the estimation issues with fixed coefficient autoregressive models deriving a general form of the well-known Nickell-bias (Nickell, 1981). Finally, some conclusions are drawn in Section 3.5.

3.2 Fixed Coefficients Models and their Estimation

This section discusses estimation issues with a benchmark fixed coefficients model, proposes two distinct estimators, and lists possible model specifications, all being special cases of the benchmark.

3.2.1 The Benchmark Model and its Estimation with Least Squares

Model (3.1) of course in this general form is not identified, so some parameter restrictions need to be applied. To start with, let us assume that the parameter structure is the most general possible, that is, $\beta_{ijt} = (\beta + \gamma_{ij} + \alpha_{it} + \alpha^*_{jt})$. In other words, the parameter vector β_{ijt} is assumed to have a two-part effect: β captures the universal effect of x'_{ijt} on y_{ijt}, something which is the same across entities and time periods, while individual-pair, individual-time specific effects are captured by $\gamma_{ij} + \alpha_{it} + \alpha^*_{jt}$. The case of $\beta = 0$ is clearly a special case of this and so won't be investigated here. The restriction on the parameter structure gives the so called all-encompassing model (as I am going to show later on, by applying appropriate restrictions, several other useful models are in fact encompassed in model (3.2))

$$y_{ijt} = x'_{ijt}(\beta + \gamma_{ij} + \alpha_{it} + \alpha^*_{jt}) + \varepsilon_{ijt} \tag{3.2}$$

or

$$y = X_1\beta + X_2\gamma + X_3\alpha + X_4\alpha^* + \varepsilon$$

with

$$\begin{array}{ll} X_1 = \Delta(\iota_{N_1N_2T} \otimes I_K) & (N_1N_2T \times K) \\ X_2 = \Delta(I_{N_1N_2} \otimes \iota_T \otimes I_K) & (N_1N_2T \times N_1N_2K) \\ X_3 = \Delta(I_{N_1} \otimes \iota_{N_2} \otimes I_T \otimes I_K) & (N_1N_2T \times N_1TK) \\ X_4 = \Delta(\iota_{N_1} \otimes I_{N_2} \otimes I_T \otimes I_K) & (N_1N_2T \times N_2TK) \end{array}$$

where I keep assuming that I and ι are the identity matrix and the column of ones respectively with sizes on the index, and

$$\Delta = \begin{pmatrix} x'_{111} & & & \\ & x'_{112} & & \\ & & \ddots & \\ & & & x'_{N_1N_2T} \end{pmatrix} \quad (N_1N_2T \times N_1N_2TK)$$

is the diagonally arranged data matrix.

Unfortunately, β still can not be identified, as the matrix $X = (X_1,\, X_2,\, X_3,\, X_4)$ has no full column rank ($K(N_1N_2 + N_1T + N_2T - 2)$ instead of $K(N_1N_2 + N_1T + N_2T + 1)$), because the column-wise sums of X_2, X_3 and X_4 are all identical to X_1. The question is what further restrictions (a number of $3K$) to make to properly identify the model. We have many options at hand, for example the most obvious one is to set $\gamma_{ij} = \alpha_{kt} = \alpha^*_{ls} = 0$ for particular pairs.[4] This may as well be too restrictive in certain applications so a more 'even' restriction could be to normalize to the averages of the parameters,

$$\sum_{ij} \gamma_{ij} = 0; \quad \sum_{it} \alpha_{it} = 0; \quad \sum_{jt} \alpha^*_{jt} = 0 \tag{3.3}$$

as proposed in Hsiao (2003) for 2D models.[5] Then $\tilde{X} = (X_1,\, \tilde{X}_2,\, \tilde{X}_3,\, \tilde{X}_4)$ has full column rank, where the $\tilde{X}_k$-s denote the matrices of observations X_k, after imposing the proper restrictions. To proceed, the adjusted (identified) model can be estimated with straight OLS to get an estimator for the composite parameter $\delta = (\beta'\, \gamma'\, \alpha'\, \alpha^{*'})'$,

$$\hat{\delta} = \left(\tilde{X}'\tilde{X}\right)^{-1} \tilde{X}y$$

or alternatively, expressing it for $\hat{\beta}$ from the Frisch-Waugh theorem,

$$\hat{\beta} = \left(X_1' M_{\tilde{X}_2\tilde{X}_3\tilde{X}_4} X_1\right)^{-1} X_1' M_{\tilde{X}_2\tilde{X}_3\tilde{X}_4} y$$

where $M_{\tilde{X}_2\tilde{X}_3\tilde{X}_4}$ is the optimal projection orthogonal to $(\tilde{X}_2,\, \tilde{X}_3,\, \tilde{X}_4)$. Although multiplication of the model with $M_{\tilde{X}_2\tilde{X}_3\tilde{X}_4}$ removes the heterogeneous effects, one is still faced with the problem of inverting $(KN_1N_2 \times KN_1N_2)$, $(KN_1T \times KN_1T)$, and $(KN_2T \times KN_2T)$ matrices repeatedly, which can become quickly computationally forbidding. One could try to figure out what this projection (with a set of non-trivial matrices) does to a typical x'_{ijt}, and be lost in the algebra quickly. Even if the above estimators can be dealt with in small samples, we still have the inconvenience of

[4] We usually set the first of the parameters to zero and interpret the rest of the parameters as differences from the first one.

[5] With different restrictions the estimator for β will naturally be different, as well as the interpretation of the varying parameters. For example with $\gamma_{11} = 0$, $\hat{\gamma}_{ij}$ is a measure relative to a benchmark pair, γ_{11}, on the other hand with $\bar{\gamma} = 0$ $\hat{\gamma}_{ij}$ is a measure relative to the mean of the effects.

incorporating the restrictions first. Having said that, if we are uncertain about what the proper set of restriction would be, or simply there is scope for experimenting different restrictions, we would have to painfully redo the estimation each time.

3.2.2 The Least Squares of Incomplete Rank

There is, however, a more general, and useful approach to be used to derive estimators for β and for the heterogeneous parameters as well. The so-called *Least Squares of incomplete rank* (wonderfully explained in Searle, 1971, pp. 164–225) is presented in the general form

$$\hat{\delta} = \left(X'X\right)^{-} X'y + \left(\left(X'X\right)^{-} X'X - I\right) y_1 \tag{3.4}$$

with '$-$' standing for any generalized inverse, now $X = (X_2,\ X_3,\ X_4,\ X_1)$ and y_1 being an arbitrary vector satisfying some regularity conditions to be detailed later on.[6] To have an insight on the formula, notice that $\left(X'X\right)^{-} X'y$ is a general solution of the (under-identified) linear system of equations, while the second part incorporates the restriction needed for identification, through y_1. Specifically, with a linear restriction of form

$$R'\delta = r$$

y_1 is such, that it satisfies

$$R'\left((X'X)^{-}X'X - I\right) y_1 = r - R'(X'X)^{-}X'y \tag{3.5}$$

Notice, that if the model were identified, the first part of (3.4) would be a full-rank (straight) Least Squares (as the generalized inverse boils down to a 'regular' inverse), and the second part would drop out for the same reason.[7] This approach has the advantage of being able to incorporate the restriction as the last step: everything, which is model-specific can be derived strictly before arriving to the restriction, offering a flexible way to handle multiple forms of restrictions. The shortcoming of the approach lies in the derivation of the generalized inverse (hereafter, *g-inverse*) of the block matrix $X'X$. This can be cumbersome and challenging in some cases. Let me note here, that this method does not fully alleviate the dimensionality problem: to go through the g-inverse of $X'X$, using sophisticated block-inverse theory and clever matrix manipulations, one still have to invert

[6] The reason for placing X_2 to the front of X is that $X_2'X_2$ is the largest matrix, yet block-diagonal. As its inverse is the inverses of its blocks, it is easy to be computed, alleviating some of the issues with the dimensions. Such simple rearrangement of matrix elements comes very handy, and should be kept in mind whenever computational limits are close.

[7] As $(X'X)^{-}X'X - I = I - I = 0$ in case of an $X'X$ of full rank.

full matrices of order $\max\{N_1TK,\ N_2TK\}$. This is unfeasible in practice even for moderate sample sizes.[8]

Full calculations (employing some of the results of Miao, 1991 and Hartwig, 1976) are included to Appendix A, I only present the main steps and results of the estimation procedure. The data matrix to be inverted, $X'X$, is a (4×4) block matrix of order $K(N_1N_2+N_1T+N_2T+1)$. Direct inversion of such is clearly impossible even for moderate N_1 (N_2). Instead, we can proceed by first calculating the g-inverse of its second principal minor (that is, the inverse of the upper left (2×2) block), then use this g-inverse to calculate the g-inverse of $X'X$'s third principal minor (for which we need the inverse of its upper left block, which we already have), and so on. The resulting $(X'X)^-$ multiplied with $X'y$ gives the *general solutions* $\delta^0=(\gamma^{0'}\ \alpha^{0'}\ \alpha^{*0'}\ \beta^{0'})'$, where '0'-s are inserted to the superscripts to emphasise that they are *not yet* estimators. Even though $(X'X)^-$ can only be specified up to full $(KN_1T\times KN_1T)$ and $(KN_2T\times KN_2T)$ matrices, interestingly $(X'X)^-(X'X)$, necessary for the restriction, can be fully recovered (see Appendix A for the full derivations). That being said, with restrictions (3.3),

$$\left(\left(X'X\right)^{-}X'X-I\right)y_1=\begin{pmatrix} -\iota_{N_1N_2}\otimes\frac{1}{N_1N_2}\sum_{ij}\gamma_{ij}^0 \\ -\iota_{N_1T}\otimes\frac{1}{N_1T}\sum_{it}\alpha_{it}^0 \\ -\iota_{N_2T}\otimes\frac{1}{N_2T}\sum_{jt}\alpha_{jt}^{*0} \\ \frac{1}{N_1N_2}\sum_{ij}\gamma_{ij}^0+\frac{1}{N_1T}\sum_{it}\alpha_{it}^0+\frac{1}{N_2T}\sum_{jt}\alpha_{jt}^{*0} \end{pmatrix},$$

a $(K(N_1N_2+N_1T+N_2T+1)\times 1)$ matrix. This is nice, as the estimators (separately defined for each parameter) then read, as

$$\begin{aligned} \hat{\gamma}_{ij} &= \gamma_{ij}^0-\tfrac{1}{N_1N_2}\sum_{ij}\gamma_{ij}^0 && (i=1\dots N_1,\ j=1\dots N_2) \\ \hat{\alpha}_{it} &= \alpha_{it}^0-\tfrac{1}{N_1T}\sum_{it}\alpha_{it}^0 && (i=1\dots N_1,\ t=1\dots T) \\ \hat{\alpha}_{jt}^* &= \alpha_{jt}^{*0}-\tfrac{1}{N_2T}\sum_{jt}\alpha_{jt}^{*0} && (j=1\dots N_2,\ t=1\dots T) \\ \hat{\beta} &= \beta^0+\tfrac{1}{N_1N_2}\sum_{ij}\gamma_{ij}^0+\frac{1}{N_1T}\sum_{it}\alpha_{it}^0+\frac{1}{N_2T}\sum_{jt}\alpha_{jt}^{*0}\,. \end{aligned} \tag{3.6}$$

Loosely speaking, the generalized solutions are adjusted (here, de-meaned) to correct for identification. Still, we can not forget about the crucial problem of arriving to the general solutions β^0, γ^0, α^0, and α^{*0}. To have them, we have to invert potentially large matrices which makes the estimators in this present form hardly applicable for practical purposes if N_1 (N_2) is even moderately large. Unfortunately, no scalar solutions per se are available currently.

[8] A simple model with 20 regressors on a data of 10 years with 1000 individual i (and less individual j) sets up a 200000×200000 matrix to be inverted directly. This clearly exceeds all reasonable computational limits.

3.2.3 Incomplete Data

The results, so far, have assumed complete data. We know, however, that real-life panels (especially large and higher-dimensional ones) are almost exclusively incomplete of some nature. Let us see then how unbalanced data affects estimator (3.6), in order to make these fixed coefficient models more suitable for empirical analysis. Let us assume then (in line with Chapter 1), that for each (ij) pair, instead of having data over $1 \dots T$, we have a T_{ij} set of observations, where $T_{ij} \subseteq \{1, \dots, T\}$. I also assume, that $|T_{ij}| \geq 2$ for all (ij), and that for every $t = 1 \dots T$, there is at least two (ij) pairs having observations. This way I guarantee the identification of all individual parameters, also avoiding all kinds of rank issues with the data matrix X. The unbalanced version of (3.6) is then obtained in the following way. The data matrices, X_1, X_2, X_3, X_4, and y has to be adjusted to reflect the unbalanced nature of the data, *i.e.*, the rows corresponding to the missing observations has to be deleted, to get the desired number of rows, $\sum_{ij} |T_{ij}|$. Note, that column-wise, the dimensions of the respective matrices are unchanged. Using these adjusted data matrices, the unbalanced estimator of model (3.2) is formulationally identical to (3.6).[9] As this result is not model specific, the above reasoning underpins the easy application of such unbalanced estimators for all subsequent models as well. Even though the estimators are different in the underlying data and sample sizes, the estimators with the adjusted data matrices are formulationally identical to their complete versions.

3.2.4 Alternative Model Specifications

Due to the severe computational limitations with the estimation of model (3.2), I assume that some of the heterogeneous parameters are restricted over some of the indices, or even more so, completely missing. This gives us numerous new model versions, many of which carry empirical relevance. For example, if we set $\gamma_{ij} = 0$ for all (ij) pairs, we arrive to a simpler fixed coefficients marginal effect model,

$$y_{ijt} = x'_{ijt}\left(\beta + \alpha_{it} + \alpha^*_{jt}\right) + \varepsilon_{ijt} \tag{3.7}$$

with restrictions

$$\sum_{it} \alpha_{it} = 0\,; \quad \sum_{jt} \alpha^*_{jt} = 0\,.$$

[9] That is so, as $(X'X)^{-}X'X - I$ and with it, y_1 are unaffected by the unbalanced data, as long as the column space of X does not shrink.

10.14754/CEU.2017.05

The optimal estimator, very similar to estimator (3.6) takes the form

$$\begin{aligned}
\hat{\alpha}_{it} &= \alpha^0_{it} - \tfrac{1}{N_1T}\sum_{it}\alpha^0_{it} \qquad (i = 1\dots N_1\ ,t = 1\dots T)\\
\hat{\alpha}^*_{jt} &= \alpha^{*0}_{jt} - \tfrac{1}{N_2T}\sum_{jt}\alpha^{*0}_{jt} \qquad (j = 1\dots N_2\ ,t = 1\dots T)\\
\hat{\beta} &= \beta^0 + \tfrac{1}{N_1T}\sum_{it}\alpha^0_{it} + \frac{1}{N_2T}\sum_{jt}\alpha^{*0}_{jt}\,.
\end{aligned} \tag{3.8}$$

Unfortunately, the same problem arises as with model (3.2): I can not avoid to directly invert order $\max\{KN_1T,\ KN_2T\}$ non-sparse matrices. To fully get rid of this dimensionality problem, I take one step forward, and assume also, that $\alpha^*_{jt} = 0$, for all (jt) pairs, as

$$y_{ijt} = x'_{ijt}\,(\beta + \alpha_{it}) + \varepsilon_{ijt} \tag{3.9}$$

shows, or using matrix notation,

$$y = X_1\beta + X_3\alpha + \varepsilon$$

with the suitable restriction (to fix the rank deficiency of K)

$$\sum_{it}\alpha_{it} = 0\,.$$

As the upper left block of the square matrix $X'X$ (with $X = (X_3,\ X_1)$) is block diagonal, the g-inverse of $X'X$ is obtained easily. This also means, that the estimator can be further simplified to reach scalar formed estimators for all model parameters:

$$\begin{aligned}
\hat{\alpha}_{it} &= \alpha^0_{it} - \tfrac{1}{N_1T}\sum_{it}\alpha^0_{it} = \left(\sum_j x_{ijt}x'_{ijt}\right)^{-1}\left(\sum_j x_{ijt}y_{ijt}\right) - \hat{\beta}\\
\hat{\beta} &= \beta^0 + \tfrac{1}{N_1T}\sum_{it}\alpha^0_{it} = \frac{1}{N_1T}\sum_{it}\left(\sum_j x_{ijt}x'_{ijt}\right)^{-1}\left(\sum_j x_{ijt}y_{ijt}\right),
\end{aligned} \tag{3.10}$$

where again, α^0_{it} and β^0 are the generalized solutions coming from $(X'X)^-X'y$. As seen from (3.10), the largest matrix to be inverted, to get the estimators, is $(K\times K)$.

This model can, however, be considered well too restrictive. So let me turn, instead, to an other set of restrictions of the all-encompassing model (3.2), which mirrors the structure used in 2D fixed effects panel data models:

$$y_{ijt} = x'_{ijt}(\beta + \gamma_{ij} + \lambda_t) + \varepsilon_{ijt}, \tag{3.11}$$

or similarly,

$$y = X_1\beta + X_2\gamma + X_5\lambda + \varepsilon$$

with

$$X_5 = \Delta(\iota_{N_1N_2}\otimes I_T\otimes I_K) \qquad (N_1N_2T\times KT)\,.$$

For identification, I impose the restrictions

$$\sum_{ij} \gamma_{ij} = 0; \quad \sum_t \lambda_t = 0,$$

which ultimately leads to the estimators

$$\begin{aligned}
\hat{\gamma}_{ij} &= \gamma_{ij}^0 - \tfrac{1}{N_1 N_2} \sum_{ij} \gamma_{ij}^0 && (i = 1 \dots N_1,\ j = 1 \dots N_2) \\
\hat{\lambda}_t &= \lambda_t^0 - \tfrac{1}{T} \sum_t \lambda_t^0 && (t = 1 \dots T) \\
\hat{\beta} &= \beta^0 + \tfrac{1}{N_1 N_2} \sum_{ij} \gamma_{ij}^0 + \frac{1}{T} \sum_t \lambda_t^0 .
\end{aligned} \tag{3.12}$$

To get expressions for γ^0, λ^0, and β^0, however, we have to elaborate on $(X'X)^{-}X'y$. While model (3.11) already seems to be of high practical relevance, as it captures both individual and time heterogeneity (without over-complicating the parameter structure), it is also appealing computationally, as for its full estimation, I only have to invert $(KT \times KT)$ matrices. Moreover, this model can also be viewed as a 2D fixed coefficients panel data model, with the (ij) pairs being the individuals.

Model

$$y_{ijt} = x_{ijt}'(\beta + \alpha_i + \gamma_j + \lambda_t) + \varepsilon_{ijt} , \tag{3.13}$$

or alternatively,

$$y = X_1\beta + X_6\alpha + X_7\gamma + X_5\lambda + \varepsilon$$

with

$$X_6 = \Delta(I_{N_1} \otimes \iota_{N_2 T}) \qquad X_7 = \Delta(\iota_{N_1} \otimes I_{N_2} \otimes \iota_T)$$

is a special case of model (3.11), with the restriction $\gamma_{ij}^* = \alpha_i + \gamma_j$. The Least Squares of no full rank estimator for this model can easily be worked out as for the other models. With the identifying restrictions

$$\sum_i \alpha_i = 0; \quad \sum_j \gamma_j = 0; \quad \sum_t \lambda_t = 0,$$

the estimators become

$$\begin{aligned}
\hat{\alpha}_i &= \alpha_i^0 - \tfrac{1}{N_1} \textstyle\sum_i \alpha_i^0 && (i = 1 \dots N_1) \\
\hat{\gamma}_j &= \gamma_j^0 - \tfrac{1}{N_2} \textstyle\sum_j \gamma_j^0 && (j = 1 \dots N_2) \\
\hat{\lambda}_t &= \lambda_t^0 - \tfrac{1}{T} \textstyle\sum_t \lambda_t^0 && (t = 1 \dots T) \\
\hat{\beta} &= \beta^0 + \tfrac{1}{N_1} \textstyle\sum_i \alpha_i^0 + \tfrac{1}{N_2} \sum_j \gamma_j^0 + \tfrac{1}{T} \sum_t \lambda_t^0
\end{aligned} \tag{3.14}$$

The above approach is not practical for large N_1 and/or N_2 however, as to get there, I need to invert non-sparse, $(KN_1 \times KN_1)$ and/or $(KN_2 \times KN_2)$ matrices. One possible way to fix this dimensionality problem is to estimate model (3.13) as if it were model (3.11), that is, by estimating some $\gamma^*_{ij} = \alpha_i + \gamma_j$ (with the restriction $\sum_{ij} \gamma^*_{ij} = 0$), rather than α_i and γ_j separately. The estimator for β and γ^*, and λ is then identical to (3.12). As I prefer to have estimators for α_i and γ_j separately (rather than for some joint bilateral parameter γ^*_{ij}), I can do it so by applying the restrictions $\sum_i \alpha_i = 0$ and $\sum_j \gamma_j = 0$ on (3.12), leading to

$$\begin{aligned} \hat{\alpha}_i &= \tfrac{1}{N_2} \textstyle\sum_j \hat{\gamma}^*_{ij} \quad (i = 1 \ldots N_1) \\ \hat{\gamma}_j &= \tfrac{1}{N_1} \textstyle\sum_i \hat{\gamma}^*_{ij} \quad (j = 1 \ldots N_2)\,. \end{aligned} \tag{3.15}$$

It is clear, that even though these estimators are unbiased for all parameters, they are clearly not optimal: I do not derive the estimators from the Frisch-Waugh theorem any more. In this case, in fact, I do not use the information present in the restrictions for the estimation: while estimator (3.14) is optimal, (3.15) can only be suboptimal. Intuitively, the optimal estimator outperforms the suboptimal for small N_1 (N_2), as computational burdens are not yet present, but I get more precise estimates due to optimality. For large N_1 (N_2) however, the optimal estimator (3.14) becomes computationally forbidding, making the suboptimal estimator (3.15) the more attractive for estimation. I am going to assess the efficiency loss versus practicality trade off through a Monte Carlo experiment. I expect, even in small samples, the data information being so overwhelming relative to the information content of the restrictions, that the efficiency loss can be completely neglected and the suboptimal estimator recommended for use.

I conducted our experiment for various parameter values, sample sizes, and underlying data generating processes (DGP) for x'_{ijt} and ε_{ijt}. Table 3.1 summarizes some of my typical results. It can be seen, that (i) there is virtually no difference between the optimal and suboptimal estimates of β, α, and γ,[10] not even for small N_1 (N_2); (ii) differences from true parameter values are redundant for all parameter estimates, even in small sample sizes; (iii) the estimated standard errors are higher (although this difference is barely noticeable) in case of the suboptimal estimators, corresponding to the loss in efficiency. The above regularities are robust across different DGP, sample sizes, and assigned parameter values.

A further restriction on model (3.11) can be implemented when we assume that there is no time effect ($\lambda_t = 0$), arriving to one of the simplest models implementable in 3D data,

$$y_{ijt} = x'_{ijt}(\beta + \gamma_{ij}) + \varepsilon_{ijt}\,, \tag{3.16}$$

or

$$y = X_1\beta + X_2\gamma + \varepsilon$$

[10] There is no point in comparing optimal and suboptimal α_i (γ_j) for all individuals one-by-one, rather, I illustrated their closeness with some descriptive statistics, like *min*, *max*, or *mean of the differences*.

Table 3.1 *Monte Carlo simulation for assessing optimality against efficiency loss*

	$N=5$ Opt[a]	Subopt	$N=10$ Opt	Subopt	$N=20$ Opt	Subopt
		$\beta=0.5$	$x_{ijt}\sim U[-1,1]$	$\varepsilon_{ijt}\sim N(0,1)$		
$\hat\beta$	0.496	0.499	0.497	0.495	0.499	0.499
	(0.025)	(0.028)	(0.006)	(0.008)	(0.002)	(0.002)
$1/N_1\sum_i \hat\alpha_i/\alpha_i$	1.002	1.002	0.999	0.998	1.000	1.001
$\max_i\{\hat\alpha_i-\alpha_i\}$	0.009	0.007	0.014	0.015	0.010	0.014
$\min_i\{\hat\alpha_i-\alpha_i\}$	0.000	0.001	0.001	0.001	0.001	0.002
$1/N_2\sum_j \hat\gamma_j/\gamma_j$	0.999	0.999	0.999	0.999	0.999	0.999
$\max_j\{\hat\gamma_j-\gamma_j\}$	0.012	0.020	0.012	0.014	0.011	0.015
$\min_j\{\hat\gamma_j-\gamma_j\}$	0.001	0.002	0.004	0.000	0.000	0.000
		$\beta=0.5$	$x_{ijt}\sim N(0,0.5)$	$\varepsilon_{ijt}\sim N(0,10)$		
$\hat\beta$	0.419	0.499	0.499	0.483	0.495	0.499
	(4.141)	(5.554)	(0.841)	(1.256)	(0.199)	(0.321)
$1/N_1\sum_i \hat\alpha_i/\alpha_i$	1.018	1.046	1.014	1.015	1.011	0.996
$\max_i\{\hat\alpha_i-\alpha_i\}$	0.217	0.317	0.199	0.167	0.122	0.171
$\min_i\{\hat\alpha_i-\alpha_i\}$	0.036	0.042	0.005	0.010	0.005	0.001
$1/N_2\sum_j \hat\gamma_j/\gamma_j$	1.079	1.061	0.950	0.949	1.128	1.157
$\max_j\{\hat\gamma_j-\gamma_j\}$	0.129	0.265	0.120	0.149	0.111	0.175
$\min_j\{\hat\gamma_j-\gamma_j\}$	0.031	0.029	0.001	0.012	0.001	0.006
		$\beta=0.5$	$x_{ijt}\sim N(5,0.5)$	$\varepsilon_{ijt}\sim N(0,1)$		
$\hat\beta$	0.500	0.500	0.500	0.500	0.499	0.499
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
$1/N_1\sum_i \hat\alpha_i/\alpha_i$	1.000	1.000	1.000	1.000	0.999	0.999
$\max_i\{\hat\alpha_i-\alpha_i\}$	0.001	0.001	0.002	0.002	0.002	0.002
$\min_i\{\hat\alpha_i-\alpha_i\}$	0.000	0.000	0.000	0.000	0.000	0.001
$1/N_2\sum_j \hat\gamma_j/\gamma_j$	1.000	1.000	1.000	1.000	1.004	1.004
$\max_j\{\hat\gamma_j-\gamma_j\}$	0.001	0.001	0.001	0.001	0.001	0.001
$\min_j\{\hat\gamma_j-\gamma_j\}$	0.000	0.000	0.000	0.000	0.000	0.000

[a] Monte Carlo (MC) Simulation of model (3.13) with 1000 MC repetitions for various sample sizes and Data Generating Processes (DGP). α_i, γ_j, λ_t are drawn from a uniform distribution over $[-10,10]$, for $K=1$, $T=5$, $\beta=0.5$. For simplicity, $N_1=N_2=N$ is assumed. Each MC repetition uses the same defined parameters, but different underlying X data and ε disturbances. 'Opt' is for the optimal estimator, following (3.14), 'Subopt' stands for the suboptimal estimator, keep following (3.14) for β, but using (3.15) for α_i, γ_j.

with the typical identifying restriction

$$\sum_{ij}\gamma_{ij}=0\,.$$

The estimator, similarly to (3.10), simplifies to

$$\begin{aligned} \hat{\beta} &= \tfrac{1}{N_1 N_2}\sum_{ij}\left[\left(\sum_t x_{ijt}x'_{ijt}\right)^{-1}\sum_t x_{ijt}y_{ijt}\right] \\ \hat{\gamma}_{ij} &= \left(\sum_t x_{ijt}x'_{ijt}\right)^{-1}\left(\sum_t x_{ijt}y_{ijt}\right) - \hat{\beta} \quad (i = 1\dots N_1,\ j = 1\dots N_2)\,. \end{aligned} \tag{3.17}$$

Though this model is less appealing, as it probably over-simplifies the varying effect regressors can have, it corresponds to the one-fold 2D fixed coefficient model (when the (ij) pairs are treated as individuals, detailed exhaustingly in Hsiao, 2003).[11]

Conveniently, other restrictions, like $\bar{\gamma}_{ij} = c$, affect the estimators through y_1 only, so switching between different forms of restrictions is fundamentally easy:

$$\begin{aligned} \hat{\beta} &= \tfrac{1}{N_1 N_2}\sum_{ij}\left[\left(\sum_t x_{ijt}x'_{ijt}\right)^{-1}\sum_t x_{ijt}y_{ijt}\right] - c\cdot\iota_K & (K\times 1) \\ \hat{\gamma}_{ij} &= \left(\sum_t x_{ijt}x'_{ijt}\right)^{-1}\left(\sum_t x_{ijt}y_{ijt}\right) - \hat{\beta} \quad (i,j = 1\dots N_1, N_2) & (K\times 1) \end{aligned}$$

Finally, let me note, that model

$$y_{ijt} = x'_{ijt}(\beta + \alpha_i + \gamma_j) + \varepsilon_{ijt}\,, \tag{3.18}$$

or similarly,

$$y = X_1\beta + X_6\alpha + X_7\gamma + \varepsilon$$

is a restricted case of model (3.13), if I set $\lambda_t = 0$, but also a special case of model (3.16), when I let $\gamma^*_{ij} = \alpha_i + \gamma_j$. Model (3.18) can be estimated optimally, following the Least Squares of incomplete rank approach. Relying on the identifying restrictions

$$\sum_i \alpha_i = 0; \quad \sum_j \gamma_j = 0\,,$$

the estimators become

$$\begin{aligned} \hat{\alpha}_i &= \alpha_i^0 - \tfrac{1}{N_1}\textstyle\sum_i \alpha_i^0 & (i = 1\dots N_1) \\ \hat{\gamma}_j &= \gamma_j^0 - \tfrac{1}{N_2}\textstyle\sum_j \gamma_j^0 & (j = 1\dots N_2) \\ \hat{\beta} &= \beta^0 + \tfrac{1}{N_1}\textstyle\sum_i \alpha_i^0 + \tfrac{1}{N_2}\sum_j \gamma_j^0\,. \end{aligned} \tag{3.19}$$

Unfortunately, my hands are tied again with the problem of inverting order KN_1 (or KN_2) matrices. As suggested for model (3.13) however, I can use the suboptimal estimator by using (3.17) to estimate β, and γ^*_{ij} (reducing the largest matrix to be

[11] Even if its practical use is questionable (at least narrower, than, for example, model (3.2)'s), its role as being a 'reference model', to derive benchmark estimators and properties, is still distinguished.

Table 3.2 *Orders of the largest matrix to be inverted during estimation*

Model[a]	Largest Order to be Inverted		Rating
	Optimal	*Suboptimal*	
(3.2)	$\max\{KN_1T,\, KN_2T\}$	-	*[b]
(3.7)	$\min\{KN_1T,\, KN_2T\}$	-	*
(3.9)	K	-	***
(3.11)	TK	-	**
(3.13)	$\min\{KN_1,\, KN_2\}$	TK	**
(3.16)	K	-	***
(3.18)	$\max\{KN_1,\, KN_2\}$	K	***

[a] Along the way, I always assume, that $N_1 >> T$ and $N_2 >> T$, while N_1 and N_2 are expected to be of similar magnitudes.
[b] Ratings (computationally): * forbidding; ** feasible; *** first-best.

inverted to a mere order of K), then, by applying the relevant restrictions on γ^*_{ij}, reach separate (but suboptimal) estimators

$$\begin{aligned} \hat{\alpha}_i &= \tfrac{1}{N_1}\textstyle\sum_j \hat{\gamma}^*_{ij} \\ \hat{\gamma}_j &= \tfrac{1}{N_2}\textstyle\sum_i \hat{\gamma}^*_{ij}\,. \end{aligned} \tag{3.20}$$

While this section provided relevant consistent estimators (in fact appealing ones in terms of generality and flexibility) for all considered models, their empirical usage fully hinges on computer memory requirements (*i.e.*, on the order of the largest matrix to be stored an inverted).[12] As some models are also supplemented with a suboptimal estimator, it is of our best interest to collect the largest matrix orders to be inverted directly for each model, as Table 3.2 does.

3.3 Extensions

3.3.1 Varying Coefficients as Functions of Observables

Estimating any of the fixed coefficients models (3.2), (3.7), etc. as done in Section 3.2, implicitly assumes that no further knowledge is available on the slope coefficients. If, on the other hand, it is suspected what actual economic factors cause the coefficients' varying nature, the corresponding parameters might be expressed as

[12] Do not forget, that while storing a matrix of order n is $\mathcal{O}(n^2)$, inverting it is $\mathcal{O}(n^3)$.

(linear) functions of the observables. Specifically, let us assume, that it is known in model (3.16) that

$$\gamma_{ij} = Z_{ij}\nu\,,$$

where Z_{ij} is a $(K \times L)$ matrix of observed variables, and ν is an $(L \times 1)$ column of unknown parameters. The model then can be rephrased as

$$y_{ijt} = x'_{ijt}\beta + x'_{ijt}Z_{ij}\nu + \varepsilon_{ijt} = x'_{ijt}\beta + w'_{ijt}\nu + \varepsilon_{ijt}\,, \tag{3.21}$$

having a number of $K+L$ slope parameters, and a single error variance to estimate. Unless in some very special cases, (3.21) is identified as the underlying data matrix is of full rank, and an estimator is reached by regressing $(X,\,W)$, the matrix stacked versions of x'_{ijt} and w'_{ijt}, on y. This model (3.21) is then exceptionally convenient from an estimation point of view, as its fixed coefficients nature reduces to having interactions in some variables. Beyond its simplicity (remember that originally KN_1N_2 parameters were needed to be estimated), the model is also capable of identifying the exact sources of the variation in γ_{ij}: Z_{ij}. We have to be careful with the interpretation of the coefficients, however, as the marginal change in y_{ijt} as a response to a unit change in the kth regressor $x_{ijt,k}$ is $\beta + Z_{ij,k}\nu$, where $Z_{ij,k}$ is the kth row of Z_{ij}. So, just as with the original fixed coefficients model, we expect x'_{ijt} to have differential effect on y_{ijt} for different (ij) pairs of entities.

A more general form of a model where varying coefficients are rephrased as functions of observable variables are obtained from model (3.2):

$$\begin{aligned} y_{ijt} &= x'_{ijt}\beta + x'_{ijt}Z^{(1)}_{ij}\nu^{(1)} + x'_{ijt}Z^{(2)}_{it}\nu^{(2)} + x'_{ijt}Z^{(3)}_{jt}\nu^{(3)} + \varepsilon_{ijt} \\ &= x'_{ijt}\beta + w'^{(1)}_{ijt}\nu^{(1)} + w'^{(2)}_{ijt}\nu^{(2)} + w'^{(3)}_{ijt}\nu^{(3)} + \varepsilon_{ijt}\,, \end{aligned} \tag{3.22}$$

with $Z^{(1)}_{ij}$, $Z^{(2)}_{it}$ and $Z^{(3)}_{jt}$ being $(K \times L_1)$, $(K \times L_2)$ and $(K \times L_3)$ matrix of regressors, and $\nu^{(1)}$, $\nu^{(2)}$, $\nu^{(3)}$ matching (fixed, unknown) parameters. Note that $Z^{(1)}_{ij}$ is expected to be sparse in each row: it is highly unlikely that L_1 different regressors govern the variability of γ_{ij} for each $k = 1 \ldots K$. Similar statements can be made for $Z^{(2)}_{it}$ and $Z^{(3)}_{jt}$ as well.

The idea of expressing varying slope coefficients as functions of observables has been present in the literature for some time now, and been used to estimate random coefficients models in several recent empirical applications (see, *e.g.*, Wu and Lin, 2002; Huber and Stanig, 2011). Its theoretical foundations can be traced back to Amemiya (1978), who reformulates an individual random coefficient as

$$\alpha_i = Z_i\nu + \vartheta_i$$

where ϑ_i is some unknown random variable uncorrelated with the disturbance. In my application ϑ_i is ignored, as no randomness in the coefficients are ever assumed, but as seen in model (3.22), multiple varying slope coefficients are allowed to have functional dependence on the observables at the same time.

Intuitively, not much changes when indeed a disturbance in the varying coefficients is allowed, like if

$$x'_{ijt}\gamma_{ij} = x'_{ijt}(Z_{ij}\nu + \vartheta_{ij})$$

is assumed in model (3.10) with some mean zero random variable ϑ_{ij}, so long it is uncorrelated with x'_{ijt} and Z_{ij}. A Least Squares with x'_{ijt} and $w'_{ijt} = x'_{ijt}Z_{ij}$ on y'_{ijt} gives consistent estimates, yet a better, GLS estimator can be constructed by taking into account the error structure with $u_{ijt} = x'_{ijt}\vartheta_{ij} + \varepsilon_{ijt}$. This is, however, beyond the scope of the section.

3.3.2 Index Deficiency in the Variables

So far we have assumed that our data fully spans the three-dimensional (ijt) space, *i.e.*, it is heterogeneous in all directions. Most typically, however majority of the covariates fail to show variation in all three ways. Regressors of such are fixed over one (or some) of the indices, raising new identification issues. If, for example, the regressors are time invariant (which is the case in many trade models), like distance, common language, common border, we have (x'_{ij})-type observations; Population or GDP observations are (x'_{it})-type (or similarly (x'_{jt})-type). Let us illustrate the problems of such limitations on model (3.16). If the data is such, that it is fully constant over time, model (3.16) is simplified to

$$y_{ijt} = x'_{ij}(\beta + \gamma_{ij}) + \varepsilon_{ijt}$$

It is true, that the model parameters are identified (obviously after imposing the usual K restrictions), but the number of observations (KN_1N_2) exactly matches with the number of model parameters to be estimated (KN_1N_2). This raises two serious issues: one, standard errors can not be computed, so the model is unfeasible for testing, and two, the parameter estimates are extremely imprecise (as we actually use one observation to estimate one parameter).

Other data limitations, like

$$y_{ijt} = x'_{it}(\beta + \gamma_{ij}) + \varepsilon_{ijt}$$

affects the model in an other (though not less problematic) way: now γ_{ij} is not identified, as x'_{it} is fixed over j. Even though estimates for γ_{ij} can be obtained numerically, it makes no sense to formulate the model this way any more. Other models, addressed by this chapter, also have these kinds of identification problems, only even more significantly. Table 3.3 collects, for each type of data limitation,

Table 3.3 *Feasible model restrictions in response to various forms of index deficiencies*

Data Restriction	Parameter Specifications Feasible	Sample Size	Parameters to Estimate
x'_{ij}	$\beta+\alpha_i+\gamma_j$	KN_1N_2	$K(N_1+N_2-1)$
	$\beta+\alpha_i$	KN_1N_2	KN_1
	$\beta+\gamma_j$	KN_1N_2	KN_2
x'_{it}	$\beta+\alpha_i+\lambda_t$	KN_1T	$K(N_1+T-1)$
	$\beta+\alpha_i$	KN_1T	KN_1
	$\beta+\lambda_t$	KN_1T	KT
x'_{jt}	$\beta+\gamma_j+\lambda_t$	KN_2T	$K(N_2+T-1)$
	$\beta+\gamma_j$	KN_2T	KN_2
	$\beta+\lambda_t$	KN_2T	KT

the feasible restricted models to work with, along with the number of parameters to be estimated compared to the size of the data.

It is now clear how different restrictions over the data vector x'_{ijt} lead to identification issues, but I have not yet covered the case where only elements of the x'_{ijt} vector are restricted, but x'_{ijt}, as a vector, do span the (ijt) three-dimensional space. This is probably the most typical case, as there is virtually no chance of all the regressors being three dimensional. Take the following example. Let y_{ijt} be the volume of real export, with i (j) being the origin (destination) country, t being time, and let the two regressors employed be the GDP of the origin and the destination country. In this way, $x'_{ijt}=(GDP^{(1)}_{it},\ GDP^{(2)}_{jt})$ spans the whole (ijt) space, even though separately none of the regressors do. Luckily, from point of view of the estimation, such scenario does not violate the estimators detailed in Section 3.2, until none of the restricted $x^{(k)}_{ijt}$ $(k=1\dots K)$ is orthogonal to all the other regressors. To have an insight for this observation, consider the following. Even if it does not make sense at first to estimate a $\gamma^{(1)}_{ij}$ parameter for $GDP^{(1)}_{it}$, if $GDP^{(2)}_{jt}$ is also among the regressors, $\gamma^{(1)}_{ij}$ is interpreted as the effect of a unit increase of GDP of country i at time t on the export activity between county pair (ij). If $\mathrm{Corr}(GDP_{it},GDP_{jt})\neq 0$ (which we can believe easily), the unit change in GDP_{it} has also have an impact on GDP_{jt}. As a result, a change in GDP_{it} has different effects on different country pairs (ij), justifying the use of the γ_{ij} model parameters.

10.14754/CEU.2017.05

3.3.3 Mixed Models

In this chapter I have covered the cases when the varying coefficients are exclusively thought of as fixed, while Krishnakumar et al. (2017) covers the cases where these effects are represented by random variables. Now, it is not at all necessary to assume that either all are random or all are fixed. It may very well happen in practice that it makes sense to incorporate different type of effects at different levels. Such models can be specified in an analogous manner to model (3.13), where, for example the λ_t's are considered as fixed-, but α_i and γ_j are random coefficients.[13]

While this interesting concept is briefly mentioned in theoretical works (Hsiao, 2003), it is seemingly fully absent from empirical works. This is hardly surprising for two main reasons. First, there is an enormous number of possible model specifications. I have already shown multiple economically meaningful fixed coefficient models – now imagine how this number changes when each coefficient can either be fixed or random, or further, if out of the number K α_i parameters some are random, some are fixed. There is no testing tool constructed at the moment to my knowledge, which can help deciding between the unbelievably many model specifications. Second, computational difficulties are already present for many pure fixed coefficients models (as well as for pure random coefficients models), and their joint presence does not help either.

Due to the aforementioned number of models, and the scope of this chapter, I only briefly visit the essentials of estimation issues with mixed models, rather than excessively (and possibly dauntingly) carry out a full analysis.

Let us rewrite model (3.13) as

$$y_{ijt} = x'_{ijt}(\beta + \lambda_t) + u_{ijt} \quad \text{with} \quad u_{ijt} = x'_{ijt}(\alpha_i + \gamma_j) + \varepsilon_{ijt}\,. \tag{3.23}$$

Notice, that β and λ_t are still estimable, fixed parameters, but α_i and γ_j are now assumed to be zero mean pairwise uncorrelated random coefficients satisfying

$$\mathrm{E}(\alpha_i \alpha'_s) = \begin{cases} \Delta_\alpha, & \text{if } i = s \\ 0, & \text{otherwise} \end{cases} \qquad \mathrm{E}(\gamma_j \gamma'_s) = \begin{cases} \Delta_\gamma, & \text{if } j = s \\ 0, & \text{otherwise.} \end{cases}$$

The first observation to make is that just like with pure fixed coefficients models, β and λ can not be separated, and so are not identified. To be able to separate and estimate β and λ_t, the usual K parameter restrictions has to be imposed, like

$$\sum_{t=1}^{T} \lambda_t = 0 \quad (K \times 1)$$

The restricted model from this point behaves exactly as a pure random coefficients

[13] Some thoughts on mixed effects models are also published in the joint work Krishnakumar et al. (2017).

model, where β and λ_t are estimated with FGLS, performed by taking the covariance structure

$$\mathrm{E}(u_{ijt}u'_{ijt}) = \mathrm{E}\left((x'_{ijt}\alpha_i + x'_{ijt}\gamma_j + \varepsilon_{ijt})(x'_{ijt}\alpha_i + x'_{ijt}\gamma_j + \varepsilon_{ijt})'\right)$$

into account. So long as the random coefficients are uncorrelated with x'_{ijt}, the estimators for λ_t and β are consistent.

In theory, any mixed model can be estimated using the following recipe:

1. Identify the number and the form of parameter restrictions necessary to identify the model, and incorporate them
2. Derive the variance-covariance matrix Ω (more precisely its inverse)
3. Perform the GLS
4. Estimate the covariance matrix to make the GLS feasible.

Although these steps are easily formulated in theory, the size and the number of the fixed coefficients and random coefficients can strongly discourage its practical application. For example, when both α_i and λ_t are considered fixed along with β, the number of parameters to be estimated directly (in the restricted model) is $K + (N_1 - 1)K + (T - 1)K = (N_1 + T - 1)K$, whose feasibility is doubtful with any statistical package for large N_1.

A viable alternative is to estimate the incomplete rank model with FGLS first, *then* line up the parameter restriction. This might be more convenient, as long as $\Omega^{-1/2}$ can be attained at reasonable costs, as I first transform model (3.23) by premultiplying with $\Omega^{-1/2}$, then estimate the transformed model with Least Squares of incomplete rank. Intuitively,

$$\Omega^{-1/2}y = \Omega^{-1/2}X\beta + \Omega^{-1/2}X_5\lambda + \Omega^{-1/2}u$$

is simply the pure fixed coefficients model

$$\tilde{y} = \tilde{X}\beta + \tilde{X}_5\lambda + \tilde{u}\,.$$

3.4 Some Thoughts on Dynamic Models

Despite the wide applicability of the static models considered, many applications also require some autoregressive structure to the models. In order to make it operational, I have to expand fixed coefficients models to allow for lagged dependent variables as regressors. What is coming next has similarities to the static models, however, as the consistency properties of the OLS are bad, some new estimation methodology is warranted.

Take the fixed coefficient dynamic autoregressive model of order one with the simplest possible structure

$$y_{ijt} = y_{ijt-1}(\rho + \psi_{ij}) + \varepsilon_{ijt} \tag{3.24}$$

with the usual identifying restriction

$$\sum_{ij} \psi_{ij} = 0\,.$$

Luckily, estimator (3.16) can be put to my use with the only modification of replacing x'_{ijt} by y_{ijt-1}. Although there is no real modal novelty as opposed to estimators of the static models, there is a huge conceptual difference between static and dynamic models: the resulting estimators are now N_1-inconsistent (N_2-inconsistent). This falls very close to the Nickell-bias in concept (see Nickell, 1981), but falls very far from it algebraically.

Estimator (3.17) adjusted for model (3.24) reads as

$$\begin{aligned}
\hat{\rho} &= \tfrac{1}{N_1 N_2} \sum_{ij} \left(\sum_{t=2}^{T} y_{ijt-1} y_{ijt} \right)^{-1} \left(\sum_{t=2}^{T} y_{ijt-1}^2 \right) && (1 \times 1) \\
\hat{\psi}_{ij} &= \left(\sum_{t=2}^{T} y_{ijt-1} y_{ijt} \right)^{-1} \left(\sum_{t=2}^{T} y_{ijt-1}^2 \right) - \hat{\rho} \quad (i,j = 1 \dots N_1, N_2) && (1 \times 1)
\end{aligned}$$

The asymptotic properties of $\hat{\rho}$ are not surprising: the estimator is T-consistent, but inconsistent in N_1 (N_2). To see this, consider

$$\operatorname*{plim}_{T\to\infty}(\hat{\rho} - \rho) = \frac{1}{N_1 N_2} \sum_{ij} \frac{\operatorname*{plim}_{T\to\infty} \frac{1}{T} \sum_{t=2}^{T} y_{ijt-1} \varepsilon_{ijt}}{\operatorname*{plim}_{T\to\infty} \frac{1}{T} \sum_{t=2}^{T} y_{ijt-1}^2} = \frac{1}{N_1 N_2} \sum_{ij} \frac{E_t\left(y_{ijt-1}\varepsilon_{ijt}\right)}{E_t\left(y_{ijt-1}^2\right)} = 0$$

as y_{ijt-1} and ε_{ijt} are uncorrelated (and assuming the finiteness of $E_t(y_{ijt-1}^2)$). On the other hand, taking plim with respect to N_1, N_2 gives

$$\operatorname*{plim}_{N_1,N_2\to\infty}(\hat{\rho} - \rho) = \operatorname*{plim}_{N_1,N_2\to\infty} \frac{1}{N_1 N_2} \sum_{ij} \frac{\sum_{t=2}^{T} y_{ijt-1}\varepsilon_{ijt}}{\sum_{t=2}^{T} y_{ijt-1}^2} = E_{ij} \left(\frac{\sum_{t=2}^{T} y_{ijt-1}\varepsilon_{ijt}}{\sum_{t=2}^{T} y_{ijt-1}^2} \right) \neq 0$$

as the numerator and the denominator of the expression are correlated both through $\sum_t y_{ijt-1}$ and through $\sum_t \varepsilon_{ijt}$.[14]

[14] The same non-nullity of the asymptotic bias also holds, when only N_1 or N_2 grows.

10.14754/CEU.2017.05

3.5 Conclusion

This chapter has taken a step into the relatively unexplored and underdeveloped area of fixed coefficient models. A varying setup is superior to fixed effects specifications in incorporating heterogeneity not only as an *average* effect for an individual (or for its interaction with time), but *marginally*, accompanying (possibly) all covariates, resulting in several appealing model specifications. As such models by construction suffer from severe identification issues, some parameter restrictions need to be imposed. To estimate such restricted models, I take the so-called *Least Squares of no full rank* approach off the shelf and adjust it to deal with these fixed coefficient models. While the proposed method is fairly simple to implement and it flexibly handles multiple forms of restrictions, it clearly fails to offer a full solution for the dimensionality problem arising from the size of the data. Some further tricks have been introduced. First, turning to *suboptimal* estimators and bearing the negligible loss in efficiency while improving a lot on computation, with which many of the problematic models become feasible in practice. Second, expressing the varying parameters as functions of observables, and by that tracing back the model into a one with fixed slope parameters, highly reducing the number of parameters to be estimated. Finally, some insights on variables with index deficiency and on the inconsistency of autoregressive specifications have been drawn.

Two ideas of the chapter require, and deserve further elaboration. First, there is no argument against the possibility to fully reduce calculations of the Least Squares of no full rank approach (*i.e.*, give scalar representations for the estimators). If this is in fact the case, the estimation of fixed coefficient models would face no computational burden any more, no matter how large the data set is. Second, a GMM approach should be proposed along with the dynamic models, to fix the arising inconsistency, and put such models into practical use. These ideas are, however, left for future research.

Appendix A – Detailed Estimation Strategy

I am going to show the exact derivation of the Least Squares of no full rank for model (3.2); the method is analogous to all other models, and require only small modifications.

I start with data matrices y and $X = (X_2,\ X_3,\ X_4,\ X_1)$ of sizes $(N_1N_2T \times 1)$ and $(N_1N_2T \times K(N_1N_2 + N_1T + N_2T + 1))$, respectively, with the X_k $(k = 1, \ldots, 4)$ already defined.

Step 1. Get the generalized inverse of the matrix $X'X$, by applying partial inverse theory repeatedly (Miao (1991) gives an excellent guide on how to calculate the g-inverse of block matrices of various properties). First, take the second principal minor of $X'X$ (*i.e.*, its 2×2 upper left block), and calculate its g-inverse as

$$(X'X)^-_{(1)} = \begin{pmatrix} X_2'X_2 & X_2'X_3 \\ X_3'X_2 & X_3'X_3 \end{pmatrix}^- = \begin{pmatrix} A_1 & B_1 \\ B_1' & D_1 \end{pmatrix}^- =$$
$$Q_1A_1^{-1}Q_1' + \begin{pmatrix} K_1 \\ -I_{KN_1T} \end{pmatrix} Z_1^g \begin{pmatrix} K_1 \\ -I_{KN_1T} \end{pmatrix}'$$

with

$$Q_1 = \begin{pmatrix} I_{KN_1N_2} - K_1(I_{KN_1T} - Z_1^gZ_1)\tilde{K}_1^{-1}K_1' \\ (I_{KN_1T} - Z_1^gZ_1)\tilde{K}_1^{-1}K_1' \end{pmatrix}$$

$$K_1 = A_1^{-1}B_1\ , \quad \tilde{K}_1 = I_{KNT} + K_1'K_1\ , \quad \text{and} \quad Z_1^g = \tilde{K}_1^{-\frac{1}{2}}(\tilde{K}_1^{-\frac{1}{2}}Z_1\tilde{K}_1^{-\frac{1}{2}})^-\tilde{K}_1^{-\frac{1}{2}}$$

where $Z_1 = (D_1 - B_1'A_1^{-1}B_1)$ is the Schur complement. Notice, that A_1 is a nonsingular block-diagonal matrix, so its inverse is simply the inverse of its blocks, but Z_1 is a $(KN_1T \times KN_1T)$ singular full matrix, whose inversion can be computationally forbidding.

Having the inverse of the second principal minor, $(X'X)^-_{(1)} = A_2^-$, in hand, the next iterating step is the g-inverse of the third principal minor of $X'X$.

$$(X'X)^-_{(2)} = \begin{pmatrix} \begin{matrix} & A_2 & \end{matrix} & \begin{matrix} X_2'X_4 \\ X_3'X_4 \end{matrix} \\ X_4'X_2\ X_4'X_3 & X_4'X_4 \end{pmatrix}^- = \begin{pmatrix} A_2 & B_2 \\ B_2' & D_2 \end{pmatrix}^- =$$
$$Q_2A_2^-Q_2' + \begin{pmatrix} K_2 \\ -I_{KN_2T} \end{pmatrix} Z_2^g \begin{pmatrix} K_2 \\ -I_{KN_2T} \end{pmatrix}'$$

again, with

$$Q_2 = \begin{pmatrix} I_{K(N_1N_2+N_2T)} - K_2(I_{KN_2T} - Z_2^gZ_2)\tilde{K}_2^{-1}K_2' \\ (I_{KN_2T} - Z_2^gZ_2)\tilde{K}_2^{-1}K_2' \end{pmatrix}$$

$$K_2 = A_2^- B_2 \ , \quad \tilde{K}_2 = I_{KN_2T} + K_2' K_2 \ , \quad \text{and} \quad Z_2^g = \tilde{K}_2^{-\frac{1}{2}} (\tilde{K}_2^{-\frac{1}{2}} Z_2 \tilde{K}_2^{-\frac{1}{2}})^- \tilde{K}_2^{-\frac{1}{2}}$$

where $Z_2 = (D_2 - B_2' A_2^- B_2)$ is the $(KN_2T \times KN_2T)$ singular (but non-scarce) Schur complement. Notice, that as A_2^- has already been derived, only Z_2 should be inverted directly.

Lastly, picking up $(X'X)_{(2)}^- = A_3^-$, the g-inverse of the fourth principal minor of $X'X$, that is, of the matrix itself, is

$$(X'X)_{(3)}^- = \quad (X'X)^- = \begin{pmatrix} & & & X_2'X_1 \\ & A_3 & & X_3'X_1 \\ & & & X_4'X_1 \\ X_1'X_2 & X_1'X_3 & X_1'X_4 & X_1'X_1 \end{pmatrix}^- =$$

$$\begin{pmatrix} A_3 & B_3 \\ B_3' & D_3 \end{pmatrix}^- = Q_3 A_3^- Q_3' + \begin{pmatrix} K_3 \\ -I_K \end{pmatrix} Z_3^g \begin{pmatrix} K_3 \\ -I_K \end{pmatrix}'$$

with

$$Q_3 = \begin{pmatrix} I_{K(N_1N_2+N_1T+N_2T)} - K_3(I_K - Z_3^g Z_3)\tilde{K}_3^{-1} K_3' \\ (I_K - Z_3^g Z_3)\tilde{K}_3^{-1} K_3' \end{pmatrix}$$

and

$$K_3 = A_3^- B_3 \ , \quad \tilde{K}_3 = I_K + K_3' K_3 \ , \quad \text{and} \quad Z_3^g = \tilde{K}_3^{-\frac{1}{2}} (\tilde{K}_3^{-\frac{1}{2}} Z_3 \tilde{K}_3^{-\frac{1}{2}})^- \tilde{K}_3^{-\frac{1}{2}}$$

where $Z_3 = (D_3 - B_3' A_3^- B_3)$ is the Schur complement of size $(K \times K)$. Note that, if $Z_3 = 0$, $I_K - Z_3^g Z_3 = I_K$ holds, further, $Z_3^g = 0$, so the above formula reduces significantly. Multiplying $(X'X)^-$ with $X'y$ gives the generalized solution vectors γ^0, α^0, α^{*0}, β^0.

Step 2. Calculate $(X'X)^- X'X - I$ from the g-inverse of $X'X$. As $(X'X)^- X'X$ has a lot of structure, this is most easily done by 'guessing and verifying':

$$(X'X)^- X'X - I = \frac{1}{N_1N_2T + N_1 + N_2 + T} \begin{pmatrix} H_{11} & H_{12} & H_{13} & H_{14} \\ H_{12}' & H_{22} & H_{23} & H_{24} \\ H_{13}' & H_{23}' & H_{33} & H_{34} \\ H_{14}' & H_{24}' & H_{34}' & H_{44} \end{pmatrix} \otimes I_K \ ,$$

with

$$
\begin{aligned}
H_{11} &= \left(\tfrac{N_2T+1}{N_2+T} + \tfrac{N_1T+1}{N_1+T} - T\right) J_{N_1N_2} - \tfrac{N_1N_2T+N_1+N_2+T}{N_2+T}(I_{N_1} \otimes J_{N_2}) \\
&\quad - \tfrac{N_1N_2T+N_1+N_2+T}{N_1+T}(J_{N_1} \otimes I_{N_2}) \\
H_{12} &= -\tfrac{N_2T+1}{N_2+T}\iota_{N_1N_2} \otimes \iota'_{N_1T} + \tfrac{N_1N_2T+N_1+N_2+T}{N_2+T}(I_{N_1} \otimes \iota_{N_2} \otimes \iota'_T) \\
H_{13} &= -\tfrac{N_1T+1}{N_1+T}\iota_{N_1N_2} \otimes \iota'_{N_2T} + \tfrac{N_1N_2T+N_1+N_2+T}{N_1+T}(\iota_{N_1} \otimes I_{N_2} \otimes \iota'_T) \\
H_{14} &= T\iota_{N_1N_2} \\
H_{22} &= \left(\tfrac{N_2T+1}{N_2+T} + \tfrac{N_1N_2+1}{N_1+N_2} - N_2\right) J_{N_1T} - \tfrac{N_1N_2T+N_1+N_2+T}{N_2+T}(I_{N_1} \otimes J_T) \\
&\quad - \tfrac{N_1N_2T+N_1+N_2+T}{N_1+N_2}(J_{N_1} \otimes I_T) \\
H_{23} &= -\tfrac{N_1N_2+1}{N_1+N_2}J_{N_1T} + \tfrac{N_1N_2T+N_1+N_2+T}{N+T}J_{N_2} \otimes I_T \\
H_{24} &= N_2\iota_{N_1T} \\
H_{33} &= \left(\tfrac{N_1T+1}{N_1+T} + \tfrac{N_1N_2+1}{N_1+N_2} - N_1\right) J_{NT} - \tfrac{N_1N_2T+N_1+N_2+T}{N_1+T}(I_{N_1} \otimes J_T) \\
&\quad - \tfrac{N_1N_2T+N_1+N_2+T}{N_1+N_2}(J_{N_2} \otimes I_T) \\
H_{34} &= N_1\iota_{N_2T} \\
H_{44} &= -(N_1+N_2+T)
\end{aligned}
$$

Step 3. Formulate the linear restrictions

$$\sum_{ij}\gamma_{ij} = 0; \quad \sum_{it}\alpha_{it} = 0; \quad \sum_{jt}\alpha^*_{jt} = 0.$$

With $\delta = (\gamma',\ \alpha',\ \alpha^{*'}\ \beta')'$ being the composite parameter,

$$R'\delta = r$$

is written up with

$$R' = \begin{pmatrix} \iota'_{N_1N_2} & 0 & 0 & 0 \\ 0 & \iota'_{N_1T} & 0 & 0 \\ 0 & 0 & \iota'_{N_2T} & 0 \end{pmatrix} \otimes I_K \qquad r = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \otimes \iota_K\,,$$

$(3K \times K(N_1N_2+N_1T+N_2T+1))$ and $(3K \times 1)$ matrices, respectively.

Step 4. Find a y_1 of size $(K(N_1N_2+N_1T+N_2T+1)\times 1)$ satisfying condition (3.5):

$$R'\left((X'X)^-X'X-I\right)y_1 = r - R'(X'X)^-X'y\,.$$

As all matrices and vectors in (3.5) are already derived, finding y_1 is straightforward. Such candidate y_1 is

$$y_1 = \begin{pmatrix} (\frac{1+N_1N_2}{N_1^2N_2^2}\sum_{ij}\gamma_{ij}^0 + \frac{1}{N_1^2N_2T}\sum_{it}\alpha_{it}^0 + \frac{1}{N_1N_2^2T}\sum_{jt}\alpha_{jt}^{*0})\iota_{N_1N_2} \\ (\frac{1}{N_1^2N_2T}\sum_{ij}\gamma_{ij}^0 + \frac{1+N_1T}{N_1^2T^2}\sum_{it}\alpha_{it}^0 + \frac{1}{N_1N_2T^2}\sum_{jt}\alpha_{jt}^{*0})\iota_{N_1T} \\ (\frac{1}{N_1N_2^2T}\sum_{ij}\gamma_{ij}^0 + \frac{1}{N_1N_2T^2}\sum_{it}\alpha_{it}^0 + \frac{1+N_2T}{N_2^2T^2}\sum_{jt}\alpha_{jt}^{*0})\iota_{N_2T} \\ 0 \end{pmatrix} \otimes \iota_K \, .$$

Computing

$$(I-(X'X)^{-}X'X)y_1 = \begin{pmatrix} -\frac{1}{N_1N_2}\sum_{ij}\gamma_{ij}^0 \otimes \iota_{N_1N_2} \\ -\frac{1}{N_1T}\sum_{it}\alpha_{it}^0 \otimes \iota_{N_1T} \\ -\frac{1}{N_2T}\sum_{jt}\alpha_{jt}^{*0} \otimes \iota_{N_2T} \\ \frac{1}{N_1N_2}\sum_{ij}\gamma_{ij}^0 + \frac{1}{N_1T}\sum_{it}\alpha_{it}^0 + \frac{1}{N_2T}\sum_{jt}\alpha_{jt}^{*0} \end{pmatrix}$$

of size $(K(N_1N_2 + N_1T + N_2T + 1) \times 1)$ gives the correction I have to make on the generalized solutions, in order to arrive to estimator (3.6).

10.14754/CEU.2017.05

4

Empirical Applications for Multi-dimensional Panels

Section 4.3 is joint work with Janos Kollo and Istvan Boza, Sections 4.1 and 4.2 are solely my own.

4.1 Introduction

This chapter collects two favoured applications of multi-dimensional panel data. The first, a model performance assessment, takes an international trade data off the shelf, and shows how various panel estimators fare vis-a-vis each other. While Section 4.2 is primarily devoted to uncovering typical regularities of these estimators, I also shed light on some famous and controversial results of trade membership not boosting trade activity (Rose, 2004). Section 4.3 falls in line with several international attempts to measure the contemporaneous and lagged effects of foreign experience. To show the existence of and quantify such effects, we use linked employer-employee data covering half the Hungarian working-age population.

While both of these data sets are in fact three-dimensional panels, the relative position of the individual indices are fundamentally different. 3D International trade data is superior to its 2D correspondents in augmenting country pairs with the time dimension, or more typically, to substitute unilateral trade by bilateral. Either way, the individuals are at the *same* level (country), or as in many cases, they even share elements (one group of country trading with an other overlapping group). Linked employer-employee data is superior to usual 2D data on employees by adding a new cluster (employers), and by that, a rich set of information on the grouping of individuals (*i.e.*, a level *above* employees). As we will see, the models proposed in Chapters 1 and 2 are successfully applied to the problems of both Sections 4.2 and 4.3.

10.14754/CEU.2017.05

4.2 Regularities of Panel Estimators: A Trade Application

4.2.1 Introduction and Previous Results

In this section I illustrate through an empirical application how the models and estimators introduced in Chapters 1 and 2 (that is, the Within Estimator, FGLS, and the OLS) fare against each other. Along with the general performance of the models, I also hope that the estimation outcome on the introduced gravity model itself will be insightful in some ways and further supports/falsifies some earlier results.

Gravity models have enjoyed a more or less stable attention since the seminal results of Tinbergen (1962) and Pyhnen (1963). Among many others, gravity equations are capable of answering questions like the effects of trade membership on various measures of trade activity. Possibly the best example for such trade agreement is the so called '*General Agreement on Tariffs and Trade*' (hereafter, *GATT*), founded in 1947, which was later replaced by the *World Trade Organization* (*WTO*), currently comprising 153 member countries. Even though this agreement had the clear prime objective to enhance trade between member countries, Rose in his famous 2004 paper found no such promoting effect. Rose's (2004) findings naturally started a heated debate, and opened the path for several follow-up studies aiming to explain the absence of the GATT/WTO trade effect.[1]

The studies taking this debate to a new level and which this section mostly relies on, are Konya et al. (2011) and Konya et al. (2013). They attempted to overcome several shortcomings of Rose (2004) by creating a new, bilateral trade data set. They argued that as Rose employed several country-combined variables and average bilateral trade values (on a 2D data set), country-specific measures could not be traced back. Further, Rose merged *real import* and *real export* (and used them as an indicator for '*real trade activity*'), but most importantly, he did not distinguish zero trade from missing observations. The last point is of paramount importance, as the excess zeros in Rose (2004) (which are partly attributed to missing observations) can easily mitigate any positive effect of the GATT/WTO trade membership.

Though Konya et al. (2011) and Konya et al. (2013) find strong positive effects and partly explain why Rose (2004) might have failed, they suffer from two limitations which I fix in this study. First, 3D data offers a great variety of model specifications (fixed effects as well as random effects), of which many can be compared to

[1] These studies were mainly looking for reasons related to the data set or the model specification, *e.g.*, distinguishing developed and developing countries (Subramanian and Wei, 2007), reclassifying countries as being actual participants of trade agreements, rather than formal GATT/WTO members (Tomz et al., 2007), counting on relative trade barriers, not only absolute (Anderson and Wincoop, 2003), endogeneity and the issue of self-selection (Baier and Bergstrand, 2007 and Magee, 2003), *etc*.

10.14754/CEU.2017.05

give further robustness to the results. Second, as is detailed in Chapter 1, most 3D fixed effects formulations give biased and inconsistent estimators when the data is unbalanced, resulting in erroneous estimates. By using the incompleteness robust estimators for all models considered, the resulting estimators are uniformly consistent and yield better estimates then previous ones. These advantages together are expected to render the results of Konya et al. (2011) and Konya et al. (2013) more credible.

More explicitly, the questions I seek to answer in the following sections are (i) On a World trade data set which distinguishes zero trade from missing observations, and under various fixed effects and random effects model specifications, what is the effect of GATT/WTO trade membership on real trade activity? (ii) To what extent are the theoretical results of Chapters 1 and 2, specifically identification problems with fixed effects models and the convergence of FGLS to the Within estimator reflected by the empirical results?

4.2.2 The Data and Model Specifications

The new GATT/WTO data set used in the study involves 182 trading countries worldwide, observed annually over 53 years (for the period 1960–2012),[2] with over 1.2 million exporter–importer–time (*i.e.*, ijt-type) observations. Raw net import-export data were collected from IMF's Direction of Trade Statistics Yearbook, and were deflated to 2000 US $ using US CPI from IMF's International Financial Statistics Yearbook. Population and GDP measures were obtained from the World Bank's World Development Indicators. Other country- and country-pair specific demographics were collected from several sources, including World Trade Organization, CIA's Factbook and Wikipedia. The panel is highly unbalanced, as around 25% of the observations are missing (relative to a complete, fully balanced data). Finally, as mentioned earlier, the data set keeps track of the occurrence of zero trades separately from missing observations.

I estimate a fairly standard gravity model, using multiple ways to formulate heterogeneity, closely following the specification proposed by Rose (2004), and later

[2] The data set was originally embracing the 1960–2005 period, and had been updated in 2014 with an additional 7 year of data.

modified by Konya et al. (2013):

$$
\begin{aligned}
\log(\mathrm{RT})_{ijt} = \;& \beta_0 + \beta_1 \mathrm{BOTHIN}_{ijt} + \beta_2 \log(\mathrm{rGDP})_{it} + \beta_3 \log(\mathrm{rGDP})_{jt} \\
& + \beta_4 \log\left(\mathrm{rGDP}/\mathrm{POP}\right)_{it} + \beta_5 \log\left(\mathrm{rGDP}/\mathrm{POP}\right)_{jt} \\
& + \beta_6 \log(\mathrm{DIST})_{ij} + \beta_7 \log(\mathrm{LAND})_i + \beta_8 \log(\mathrm{LAND})_j \\
& + \beta_9 \mathrm{CLANG}_{ij} + \beta_{10} \mathrm{CBORD}_{ij} + \beta_{11} \mathrm{LLOCK}_i \\
& + \beta_{12} \mathrm{LLOCK}_j + \beta_{13} \mathrm{ISAND}_i + \beta_{14} \mathrm{ISLAND}_j \\
& + \beta_{15} \mathrm{EVCOL}_{ij} + \beta_{16} \mathrm{COMCOL}_{ij} + \beta_{17} \mathrm{MUN}_{ijt} \\
& + \beta_{18} \mathrm{TA}_{ijt} + \mathrm{FE(RE)}_{ijt} + \varepsilon_{ijt} \,,
\end{aligned} \tag{4.1}
$$

where RT is real trade activity (*export* or *import*), measured uniformly in 2000 US \$, BOTHIN (the main variable of interest) is the dummy variable taking 1, if country i and j are both GATT/WTO members at time t; rGDP is a country's GDP in real terms at time t; rGDP/POP is the real GDP per capita; DIST is the great circle distance in miles; LAND is land area of the country; CLANG is a dummy taking 1, if the country-pair shares language; CBORD is a dummy taking 1, if the country-pair shares border; LLOCK and ISLAND are 1, if the country is landlocked or an island, respectively; EVCOL and COMCOL are dummies taking 1, if the country has ever been colonized or if there is a common colonizer, respectively; MUNI takes 1, if the country-pair is a member of the same monetary union at time t; TA is a dummy for existing trade agreement; FE (RE) is the corresponding fixed effects (random effects) structure (any of (1.2)–(1.7) in Chapter 1, and (2.2)–(2.12) in Chapter 2, which, for more transparency, are re-collected in Table 4.1); finally, ε is the idiosyncratic disturbance term. Notice, that β_2–β_8 are all elasticities, while the rest of the parameters are all semi-elasticities. A positive BOTHIN means that trade membership promotes trade activity between member countries.

4.2.3 Results

The Pooled OLS

Equation (4.1) is estimated on the GATT/WTO trade data set using various model specifications and formulations. First, I estimate the model with Pooled OLS which in fact completely omits individual and/or time effects. The coefficient of main interest is identified if the variable BOTHIN varies at least in one dimension. Its estimate captures the effect of being a GATT/WTO member on the volume of real trade: trade activity between two GATT/WTO member countries are higher by β log points as compared to countries not sharing this membership, controlling for various, time fixed and time varying, country and country-pair characteristics. Table 4.2 contains the estimates and the estimated standard errors. All estimates are strongly statistically significant.

Table 4.1 *Fixed effects and random effects model specifications used in estimating (4.1)*

Equation Number	Model
FE Models	
(1.2)	$y_{ijt} = x'_{ijt}\beta + \alpha_i + \gamma_j + \lambda_t + \varepsilon_{ijt}$
(1.3)	$y_{ijt} = x'_{ijt}\beta + \gamma_{ij} + \varepsilon_{ijt}$
(1.4)	$y_{ijt} = x'_{ijt}\beta + \gamma_{ij} + \lambda_t + \varepsilon_{ijt}$
(1.5)	$y_{ijt} = x'_{ijt}\beta + \alpha_{jt} + \varepsilon_{ijt}$
(1.6)	$y_{ijt} = x'_{ijt}\beta + \alpha_{it} + \alpha^*_{jt} + \varepsilon_{ijt}$
(1.7)	$y_{ijt} = x'_{ijt}\beta + \gamma_{ij} + \alpha_{it} + \alpha^*_{jt} + \varepsilon_{ijt}$
RE Models	
(2.2)	$y_{ijt} = x'_{ijt}\beta + \mu_{ij} + \upsilon_{it} + \zeta_{jt} + \varepsilon_{ijt}$
(2.4)	$y_{ijt} = x'_{ijt}\beta + \upsilon_{it} + \zeta_{jt} + \varepsilon_{ijt}$
(2.6)	$y_{ijt} = x'_{ijt}\beta + \zeta_{jt} + \varepsilon_{ijt}$
(2.8)	$y_{ijt} = x'_{ijt}\beta + \mu_{ij} + \lambda_t + \varepsilon_{ijt}$
(2.10)	$y_{ijt} = x'_{ijt}\beta + \upsilon_i + \zeta_j + \lambda_t + \varepsilon_{ijt}$
(2.12)	$y_{ijt} = x'_{ijt}\beta + \mu_{ij} + \varepsilon_{ijt}$

As seen from the table, after controlling for several country-specific, country-time-specific, and country-pair specific factors, trade membership still has a strong positive effect on real export. Being a member of GATT/WTO raises annual real export significantly, by 0.601 log points. Running the regressions with real import, as dependent variable, gives similar estimates in magnitudes, therefore I exclude these from the table. Both the exporters' and importers' GDP seem to be an important factor in governing trade activity: a 1% jump of the home country's GDP (destination country's GDP) in real terms raises exports by 2.025% (1.549%). Per capita GDP has ambiguous sign after controlling for real GDP, and is only slightly different economically from zero. The rest of the estimates suggest that trade activity grows in 'commons', that is, with sharing language, border, colonizer, and with being in the same monetary union, but not surprisingly, falls sharply with distance.

Even though I control for a rich set of factors which may explain trade activity, and may account for individual differences as well, it is highly unlikely that there is no further omitted unobserved individual/time heterogeneity which in turn biases the Least Squares estimator.[3] The presence of such unobserved factor, which makes a country-pair more likely to trade (some innate similarity) by also raising the likelihood of them being members of the same trade union, would cause the Pooled

[3] Such omitted variables can be countries' attitude, historical bonds, general openness, general willingness to trade in a given year, etc.

Table 4.2 *Pooled OLS estimate of model (4.1)*

Variable[a]	$\hat{\beta}$	Se($\hat{\beta}$)
C	−54.365	0.120
BOTHIN	0.601	0.014
LNRGDP1	2.025	0.004
LNRGDP2	1.549	0.004
LNRGDPPOP1	0.076	0.005
LNRGDPPOP2	−0.031	0.006
LNDIST	−2.058	0.009
LNLAND1	−0.259	0.004
LNLAND2	−0.166	0.004
CLANG	0.588	0.014
CBORD	1.450	0.045
LLOCK1	0.233	0.018
LLOCK2	−0.642	0.018
ISLAND1	0.170	0.017
ISLAND2	0.238	0.017
EVCOL	1.266	0.070
COMCOL	1.013	0.020
MUNI	1.434	0.022
TA	0.818	0.017
R^2	0.804	

[a] Estimation is done on the GATT/WTO World data set, with LNREXPORT as dependent variable.

OLS to overestimate the true effect. For this reason, I now account for this excess individual/time heterogeneity via fixed effects first, then via random effects.

The Fixed Effects

Let me consider first the case, when the individual heterogeneity is incorporated as observable (and estimable) parameters, that is, the case of fixed effects. Table 4.4 collects Within estimates of the fixed effects models (1.2)–(1.7) (for the exact model specifications see Table 4.1).[4] As it is not always clear what is the source of variation needed for the identification of the coefficient of main interest, nor how to interpret the estimates, a supporting Table 4.3 collects these ideas for each fixed effects model specification.

Estimation procedures are detailed and discussed in Section 1.5. As can be seen from the table, several variables are not identified under Within estimation, simply

[4] Program codes written in R for the LS, Within- and the FGLS estimation of the listed fixed- and random effects models are publicly available at https://www.dropbox.com/sh/gzbixlckpqj8839/AAAeAPNnZEdt5jPJ-7k74gBIa?dl=0

Table 4.3 *The source of variation needed for identification, and the interpretation of the coefficient of main interest*

Model	Variation needed in BOTHIN[a]	Interpretation of β_1[b]
(1.2)	In two dimensions at the same time	...the average trade memberships of the exporting country with all importing countries over all time periods, the average trade memberships of the importing country with all exporting countries over all time periods, and the average total memberships in year t
(1.3)	In t for at least one (ij) pair	...the average trade memberships of the country-pair over years
(1.4)	In t for at least one (ij) pair and in i or j for at least one t period	...the average trade memberships of the country-pair over years, and the average trade memberships in year t
(1.5)	In i for at least one jt pair	...the average trade memberships of the importing country in year t
(1.6)	In j for at least one it pair and in i for at least one jt pair	...the average trade memberships of the importing country in year t, and the average trade memberships of the exporting country in year t
(1.7)	In t for at least one ij pair, in j for at least one it pair and in i for at least one jt pair	...the average trade memberships of the country-pair over years, the average trade memberships of the importing country in year t, and the average trade memberships of the exporting country in year t

[a] These conditions are only met in case of models (1.2), (1.5) and (1.6), as seen in Table 4.4.
[b] The interpretation starts with: "Difference between the real trade activity of two GATT/WTO member countries relative to the case when one or both countries are non-members, controlling for country- and country-pair specific observables, as well as for..."

because some variables are fixed in some dimensions and are eliminated by the Within transformation. For example, model (1.3) has γ_{ij} fixed effects, whereas, DIST_{ij} is fixed in a similar manner. When we eliminate γ_{ij}, we also clear the latter from the model.[5] This is even more visible in case of the all-encompassing model (1.7), where *all* variables are fixed in one way or an other, and so are eliminated.

The estimates are qualitatively the same as of the Pooled OLS's. Table 4.4 suggests that a GATT/WTO member country trades around 0.6-1.1 log points more with an other member country, than with a country of similar characteristics, but outside the trade union. This observation about the positivity of the effect of trade membership is robust across various model specifications, though it varies strongly in magnitude.

The low R^2 values are of no concern, and are simply reflecting the fact that after the Within transformation, much of the variation is removed from the data. As the left hand side variables are fixed in one way, or another, the transformations eliminate the major part of this existing variation, as opposed to the right hand side, where real export, after the Within transformation, still exhibits a considerable variation.

As much as fixed effects results are more credible in this panel data context

[5] To be fully precise, not only x_{ij}-type covariates are eliminated, but also which are 'nested' in x_{ij}: x_i and x_j, *i.e.*, all country-specific measures as well.

Table 4.4 *Within estimates of various fixed effects model formulations*

Variable[a]	Model (1.7)		Model (1.6)		Model (1.5)		Model (1.4)		Model (1.2)		Model (1.3)	
	$\hat{\beta}$	Se($\hat{\beta}$)	$\hat{\beta}$	Se($\hat{\beta}$)	$\hat{\beta}$	Se($\hat{\beta}$)	$\hat{\beta}$	Se($\hat{\beta}$)	$\hat{\beta}$	Se($\hat{\beta}$)	$\hat{\beta}$	Se($\hat{\beta}$)
BOTHIN	-	-	0.615	0.048	1.111	0.020	-	-	0.758	0.052	-	-
LNRGDP1	-	-	-	-	1.983	0.004	1.265	0.018	1.301	0.022	1.948	0.014
LNRGDP2	-	-	-	-	-	-	1.449	0.019	1.486	0.022	2.170	0.015
LNRGDPPOP1	-	-	-	-	0.100	0.005	−0.083	0.009	−0.100	0.010	−0.219	0.009
LNRGDPPOP2	-	-	-	-	-	-	−0.180	0.009	−0.181	0.011	−0.338	0.009
LNDIST	-	-	−2.287	0.009	−2.019	0.010	-	-	−2.256	0.010	-	-
LNLAND1	-	-	-	-	−0.231	0.004	-	-	-	-	-	-
LNLAND2	-	-	-	-	-	-	-	-	-	-	-	-
CLANG	-	-	0.923	0.016	0.668	0.015	-	-	0.896	0.017	-	-
CBORD	-	-	0.992	0.040	1.368	0.043	-	-	1.161	0.042	-	-
LLOCK1	-	-	-	-	0.185	0.017	-	-	-	-	-	-
LLOCK2	-	-	-	-	-	-	-	-	-	-	-	-
ISLAND1	-	-	-	-	0.184	0.017	-	-	-	-	-	-
ISLAND2	-	-	-	-	-	-	-	-	-	-	-	-
EVCOL	-	-	0.876	0.064	1.561	0.068	-	-	1.331	0.071	-	-
COMCOL	-	-	1.253	0.020	1.186	0.021	-	-	1.277	0.022	-	-
MUNI	-	-	0.575	0.023	1.331	0.023	-	-	0.576	0.024	-	-
TA	-	-	1.062	0.017	1.058	0.017	-	-	1.076	0.018	-	-
R^2	0		0.165		0.456		0.014		0.159		0.170	

[a] Estimation is done on the GATT/WTO World data set, with LNREXPORT as dependent variable. The unusual ordering of the models helps the comparison with the random effects models. No general intercept is added to the model due to the presence of fixed effects.

10.14754/CEU.2017.05

and are considerable improvements over the Pooled OLS estimator as well, many important parameters (among them the key parameter of focus) are not identified and so are non-estimable. Further, the number of model parameters is incredibly high, especially when interaction fixed effects (country–country, country–time) are added to the model, which can easily lead to a classical textbook over-specification. These fixed effect parameters can in fact capture too much from the data variation, and might actually capture some of the trade union's effect which I would like to uncover. To go around this problem, I incorporate the individual/time heterogeneity as random variables next, and compare the estimates to those of the Within estimator's.

The Random Effects

The random effects specifications, *i.e.*, models (2.2)–(2.12) from Chapter 2 (see also the second panel of Table 4.1 for the exact model specifications), are estimated optimally with FGLS. Random effects specifications in general correspond to the case where the heterogeneity is incorporated to the models as unobservable, random variables, significantly reducing the number of parameters to estimate.[6] If the data resembles more to a random sample from the underlying population, random effects is more suitable, than fixed effects.[7] Table 4.5 captures the FGLS estimates of the random effects models. While the interpretation of the coefficient of main interest is similar to fixed effects models, now BOTHIN only needs to exhibit variation in one direction in order to make its coefficient identified. This result is universal across random effects models.

From the table it is clear that although the parameter estimates vary somewhat across model specifications, they all lead to the same qualitative outcome: being a GATT/WTO member raises trade flows by 0.5–0.9 log points, a 1% jump in GDP is associated with a 1–2% increase in trade activity, whereas per capita GDP estimates are mixed and much smaller in magnitude. Similarly to Table 4.4, trade activity grows in commons, and falls with distance. R^2-s are much higher now, and imply that about 75% of the variation of real export is explained by the covariates.

Sections 2.2 and 2.3 of Chapter 2 collected the conditions needed for the consistency of the FGLS estimator (Table 2.3), and also the conditions, under which the FGLS converges to each model's specific Within estimator (Table 2.5). These tables are repeated here in Table 4.6 for transparency reasons.

In the present scenario, when all N_1, N_2 and T are considered 'large', the FGLS estimator of all models in fact converge to the Within. This means that in the limit,

[6] Instead of a number of $N_1 \cdot N_2$ γ_{ij} fixed parameters, for example, only the variance of γ_{ij} has to be estimated.

[7] Let me note here, that if, for example for model (2.12), the μ_{ij} interaction effects were incorporated as fixed effects (model (1.3)), as it has been the practice in most applied studies, it would mean the explicit or implicit estimation of $N_1 N_2$ parameters, in this case about $182 \cdot 182 = 33,124$. This would look very much like a textbook over-specification case.

Table 4.5 *FGLS estimates of various random effects model formulations*

Variable[a]	Model (2.2)		Model (2.4)		Model (2.6)		Model (2.8)		Model (2.10)		Model (2.12)	
	$\hat{\beta}$	Se($\hat{\beta}$)	$\hat{\beta}$	Se($\hat{\beta}$)	$\hat{\beta}$	Se($\hat{\beta}$)	$\hat{\beta}$	Se($\hat{\beta}$)	$\hat{\beta}$	Se($\hat{\beta}$)	$\hat{\beta}$	Se($\hat{\beta}$)
C	−54.166	0.512	−54.434	0.319	−54.463	0.183	−49.146	0.395	−40.040	0.804	−59.473	0.351
BOTHIN	0.510	0.062	0.581	0.031	0.927	0.018	0.907	0.049	0.783	0.050	0.553	0.049
LNRGDP1	1.987	0.022	2.043	0.018	1.996	0.004	1.808	0.011	1.391	0.021	2.136	0.010
LNRGDP2	1.623	0.017	1.577	0.010	1.538	0.010	1.498	0.012	1.513	0.018	1.846	0.010
LNRGDPPOP1	0.045	0.027	0.059	0.022	0.096	0.005	−0.098	0.008	−0.106	0.011	−0.191	0.008
LNRGDPPOP2	−0.145	0.019	−0.022	0.013	−0.021	0.013	−0.160	0.009	−0.174	0.011	−0.264	0.009
LNDIST	−2.084	0.036	−2.242	0.010	−2.027	0.010	−1.958	0.034	−2.252	0.011	−1.915	0.034
LNLAND1	−0.287	0.019	−0.262	0.015	−0.239	0.004	−0.152	0.012	0.063	0.050	−0.310	0.012
LNLAND2	−0.156	0.015	−0.140	0.009	−0.171	0.009	−0.128	0.012	−0.094	0.023	−0.301	0.012
CLANG	0.817	0.058	0.878	0.016	0.644	0.015	0.706	0.055	0.892	0.018	0.749	0.055
CBORD	1.180	0.173	1.139	0.042	1.387	0.044	1.224	0.173	1.164	0.045	1.324	0.173
LLOCK1	0.682	0.091	0.144	0.070	0.201	0.018	0.001	0.063	-0.690	0.277	0.350	0.062
LLOCK2	−0.705	0.072	−0.752	0.039	−0.666	0.039	−0.694	0.062	−0.734	0.122	−0.308	0.061
ISLAND1	0.330	0.092	0.276	0.067	0.178	0.017	0.358	0.066	0.162	0.299	0.370	0.065
ISLAND2	0.434	0.075	0.250	0.039	0.243	0.040	0.529	0.066	0.556	0.134	0.577	0.066
EVCOL	0.861	0.277	1.157	0.070	1.498	0.070	1.496	0.269	1.335	0.075	1.010	0.268
COMCOL	0.954	0.081	1.238	0.021	1.151	0.021	0.552	0.078	1.270	0.023	0.892	0.077
MUNI	1.406	0.093	0.684	0.024	1.352	0.023	1.798	0.090	0.586	0.026	1.688	0.089
TA	1.226	0.068	1.047	0.017	1.015	0.017	1.321	0.065	1.078	0.019	1.142	0.065
R^2	0.765		0.795		0.798		0.712		0.661		0.797	

[a] Estimation is done on the GATT/WTO World data set, with LNREXPORT as dependent variable

10.14754/CEU.2017.05

Table 4.6 *Asymptotic conditions under which the FGLS estimator converges to a Within, and conditions needed for consistency*

Model	Condition
FGLS converges to a Within	
(2.2)	$N_1 \to \infty, N_2 \to \infty, T \to \infty$
(2.4)	$N_1 \to \infty, N_2 \to \infty$
(2.6)	$N_1 \to \infty$
(2.8)	$(N_1 \to \infty,\ T \to \infty)$ or $(N_2 \to \infty,\ T \to \infty)$
(2.10)	$(N_1 \to \infty,\ N_2 \to \infty)$ or $(N_1 \to \infty,\ T \to \infty)$ or $(N_2 \to \infty,\ T \to \infty)$
(2.12)	$T \to \infty$
Consistency	
(2.2)	$(N_1 \to \infty,\ N_2 \to \infty)$ or $(N_1 \to \infty,\ T \to \infty)$ or $(N_2 \to \infty,\ T \to \infty)$
(2.4)	$(T \to \infty)$ or $(N_1 \to \infty,\ N_2 \to \infty)$
(2.6)	$(N_2 \to \infty)$ or $(T \to \infty)$
(2.8)	$(N_1 \to \infty,\ T \to \infty)$ or $(N_2 \to \infty,\ T \to \infty)$
(2.10)	$(N_1 \to \infty,\ N_2 \to \infty,\ T \to \infty)$
(2.12)	$(N_1 \to \infty)$ or $(N_2 \to \infty)$

for each model, the estimates in Tables 4.4 and 4.5 should fall very close, which is actually the case here. According to the theoretical results in Section 2.3, in case of 'exploding' data (all-asymptotics), the FGLS estimator becomes the Within, and so all severe identification issues of the Within estimator is carried over to the FGLS. This is unfortunate, and would certainly reduce the attractiveness of the random effects approach specifically in those cases, where its application is usually recommended: in large panels.[8] Luckily, as seen from Table 4.5, under FGLS, I reach full identification of the parameters, and the convergence of the estimators only results similar fixed effects and random effects estimates.[9] This comes handy in cases when the Within estimation fails (like in case of the identification of model (1.7)), as we get asymptotically identical results under the FGLS approach.

There is, however, an other (among others) way to estimate random effects models: by running a Least Squares regression to estimate the model parameters and then estimate the standard errors according to the specific random effect specification. Basic algebra shows, that the resulting estimators are consistent (as long as the

[8] Of course, the decision between fixed and random effects should not be governed by merely the size of the data (despite what we see in practice), but rather by the economic rationale.

[9] The GLS estimator can also be interpreted as a two-step procedure, in which first a linear transformation is to be employed on the variables, then the transformed variables are estimated with Least Squares. When the underlying data grows in all directions, the GLS transformation becomes the Within, but for a given data and sample sizes, it never actually reaches it. This is why fixed variables are never actually eliminated from the model (their parameters are still identified, maybe only weakly), and estimates can still be collected.

random effects and covariates are uncorrelated). Table 4.7 collects Least Squares estimates of the random effect models (2.2)–(2.12).

The parameter estimates are naturally the same model-wide, but the estimated standard errors are different, as they are calculated by based on the specific model's disturbance term structure. The parameter estimates and the overall R^2 (naturally the same as in Table 4.2) stand considerably close to the FGLS estimates (and to the Within estimates in case of the parameter estimates), but the standard errors are slightly higher now, reflecting the loss of efficiency from not using the extra information about the error components structure for the estimation. Despite its non-optimality, Least Squares gives a decent first estimate to the random effects models whilst it's low computational cost is also appealing.

4.2.4 Discussion

The conclusion to be drawn from this study is twofold. In terms of empirics, I have supported the results of Konya et al. (2011) by identifying the same positive effect of GATT/WTO membership on real trade activity (export or import) in magnitude. With that, I reject Rose's (2004) result on the nullity of the trade membership effect. Although my typical estimate of the effect (around 0.7 log points) is somewhat smaller what Konya et al. (2011) finds (who concluded with 0.9-1.4 log points), this difference is most probably attributed to differences in model specification, in the underlying data or in estimation techniques, and not to various economic factors in the background. Furthermore, the fact that both fixed and random effects model formulations are considered and estimated optimally and consistently makes it hard to argue against the applied methodology and the estimates (also keeping in mind that the employed gravity model is widely used and supported by the literature).

The fact that all 6 model specifications (estimated with fixed effects or random effects) gave the same qualitative results, although they vary somewhat in magnitude, serves some evidence against the result of Rose (2004). It is not clear, however, which model specification should we trust the most and why: an issue of high empirical relevance. Let us see first the fixed effects models. The key regressor BOTHIN shows no variation in the time dimension, and so models involving country-pair fixed effects γ_{ij} render its parameter unidentified. As much as country-pair effects are appealing in capturing any unobserved bilateral economic factor affecting trade, the excess number of parameters might in fact over-fit the model. A better way to account for country-specific differences in trade is to add country fixed effects, as in model (1.2). Not to mention that nearly all parameters are estimable now, model (1.2) is far from being over-specified (having a few hundred fixed effect instead of several thousand) yet rich enough to control for systematic differences in countries' trade. In case of random effects models the question of the

Table 4.7 *Least squares estimates of various random effects specifications*

Variable[a]		Model (2.2)	Model (2.4)	Model (2.6)	Model (2.8)	Model (2.10)	Model (2.12)
	$\hat{\beta}$	Se($\hat{\beta}$)	Se($\hat{\beta}$)	Se($\hat{\beta}$)	Se($\hat{\beta}$)	Se($\hat{\beta}$)	Se($\hat{\beta}$)
C	−54.365	0.584	0.353	0.197	0.220	1.334	0.480
BOTHIN	0.601	0.069	0.045	0.025	0.021	0.159	0.054
LNRGDP1	2.025	0.024	0.018	0.005	0.010	0.073	0.017
LNRGDP2	1.549	0.020	0.011	0.010	0.009	0.032	0.017
LNRGDPPOP1	0.076	0.030	0.023	0.005	0.016	0.088	0.020
LNRGDPPOP2	−0.031	0.024	0.013	0.013	0.017	0.040	0.021
LNDIST	−2.058	0.040	0.017	0.012	0.037	0.065	0.038
LNLAND1	−0.259	0.021	0.015	0.004	0.011	0.065	0.015
LNLAND2	−0.166	0.017	0.009	0.009	0.011	0.028	0.015
CLANG	0.588	0.066	0.031	0.019	0.059	0.126	0.060
CBORD	1.450	0.188	0.049	0.046	0.187	0.124	0.187
LLOCK1	0.233	0.099	0.072	0.018	0.061	0.305	0.070
LLOCK2	−0.642	0.079	0.041	0.041	0.063	0.132	0.070
ISLAND1	0.170	0.098	0.070	0.017	0.070	0.320	0.072
ISLAND2	0.238	0.081	0.041	0.041	0.071	0.144	0.072
EVCOL	1.266	0.315	0.129	0.071	0.295	0.518	0.297
COMCOL	1.013	0.090	0.037	0.025	0.083	0.142	0.084
MUNI	1.434	0.102	0.042	0.028	0.094	0.166	0.095
TA	0.818	0.075	0.030	0.021	0.070	0.115	0.071
R^2	0.804						

[a] Estimation is done on the GATT/WTO World data set, with LNREXPORT as dependent variable

10.14754/CEU.2017.05

best model specification is not that obvious, as the incidental parameter problem now vanishes. The all-encompassing model (2.2) appears to be the most powerful one, as all dependencies possible under three-way data are accounted for, yet the number of parameters to be estimated grows only by 3 (the variance of the respective random effects), as opposed to Pooled OLS estimates. The concern of having "more" effects is only reflected in higher standard errors: the estimate for the variance of a "missing" random effect should be close to zero, and so its weight in the parameter estimation diminishes. Yet, to find more precise estimates a test for model specification (discussed in Section 2.6.1) can be carried out, starting from the all-encompassing model. Systematically testing for the nullity of random effects variances reveals which random components should be represented in the model defining the exact specification to go with. This, however, is beyond the scope of this section.

The conclusion to be drawn in terms of the underlying theory, is that some typical regularities of different panel estimators were proven. We have seen how the identification problems arising from the presence of fixed effects were tackled by the FGLS estimators of the random effects models, and how the emerging estimates in the latter case were converging to the former. Even though the Within estimators or the FGLS are computationally forbidding if done directly, due to the techniques discussed in Chapters 1 and 2 for incomplete panels, computational times and memory requirements were highly reduced. To get a vague idea about their magnitude, note that none of the estimations used more than 2GB of RAM or took more than 2 hours on a middle-end computer (dual core Intel i5 2.6GHz processor with 8GB of RAM). This is not bad, considering that the dataset has over one million observations, underpinning the wide applicability of the incomplete estimation techniques in Chapters 1 and 2.

We should not forget, however, that the primary purpose of this section is to implement the theoretical results of Chapters 1 and 2. Although the results obtained here appear to be appealing, issues like endogeneity or selection bias induced by mostly observing non-zero trades are likely to affect the validity and most certainly affect the credibility of the section. I ignored these issues for the moment in order to not lose focus on the goal, but future research has to take them into account.

10.14754/CEU.2017.05

4.3 Contemporaneous and Lagged Wage Returns to Foreign-Firm Experience – Evidence from Linked Employer-Employee Data

4.3.1 Introduction

The presence of foreign capital generates heated debates in emerging market economies. The opponents charge foreign firms with displacing local businesses, expatriating their profits and fraudulently reducing their tax liabilities by means of transfer pricing. The proponents emphasize the influx and diffusion of novel technologies and modern corporate culture, and stress the direct and indirect productivity gains from FDI. Domestic firms can learn from multinationals by copying them and adopting technological and quality standards required on the part of local suppliers. Importantly, the wages paid by foreign companies remain and are mostly spent in the host country. Furthermore, workers leaving these enterprises can be more productive and earn higher wages in the domestic sector than their incumbent counterparts. The presence of ex-foreign workers in domestic companies can also boost the productivity of co-workers with no foreign-firm experience. These advantages can outweigh the losses caused by displacement, expatriated profit and foregone tax revenues.

This study looks at the *wage advantage* of workers in foreign-owned firms, the portability of this advantage to the domestic sector through worker mobility (resulting in *lagged returns*), and possible *spillover* effects. We do so under the assumption that multinational enterprise (hereafter MNE) employees accumulate valuable knowledge that is partly transferable and, when former MNE workers shift to the domestic sector, helps their co-workers to acquire part of that knowledge.

The process of skill accumulation and skills themselves are largely unobservable, therefore their presence can be inferred only from their effects on observed outcomes like wages and productivity.[10] This study looks at the implications for wages similar to Aitken et al. (1996), Barry et al. (2005) and Poole (2013) as opposed to similar studies by Aitken and Harrison (1999), Smarzynska Javorcik (2004), Görg and Strobl (2005) or Vera-Cruz and Dutrenit (2005) and others who rather look at the effects of previous MNE experience on firms' productivity.

Aitken et al. (1996) investigate the effects of FDI on wages paid in the US, Mexican and Venezuelan economy using a wage equation similar to ours, having log wages on the left hand side and the degree of foreign presence in a given location and industry on the right hand side. While a higher level of FDI is uniformly asso-

[10] This, of course, is not to say that research on the impact of formal training is useless but it clearly covers only a fraction of 'on-the-job training', which includes copying, informal communication and *trial-and-error*. See Loewenstein and Spletzer (1999), Görg et al. (2007) and Konings and Vanormelingen (2015) as examples of research on the differential effects of formal general and specific training.

ciated with higher wages in all three economies, they only observe a spillover of increased foreign presence to domestic wages in case of the US, which is consistent with the theory of emerging economies having high wage differentials between the domestic and foreign sectors. As much as these results suggest that the Hungarian economy may as well potentially benefit from such increased presence of foreign investments, Barry et al. (2005) draw attention that the identified higher wages might be due to the fact that MNEs poach high ability workers. In particular, they argue that such "poaching effect" can not occur under (near to) perfect competition and Cobb-Douglas-type productions, but may arise in case of higher elasticities of substitution. Sure enough, consistently with their theoretical results, estimates on a pool of Irish firms identifies that wage spillovers are absent for domestic non-exporter firms, but both skilled and unskilled workers benefit from a higher FDI in case of exporters (firms involved in international competition).

In identifying the spillover effects what ex-MNE workers may exert on domestic incumbents by labour turnover, our study is closest in spirit to Poole (2013) and mirrors his empirical strategy. Poole was the first to quantify wage increments as a result of formal or informal communications between domestic incumbent and MNE-trained coworkers. On a fixed effects wage equation with the share of ex-MNE workers on the right hand side he finds, that identical workers earn more at firms with higher presence of foreign trained colleagues, and concludes that this *spillover effect* corresponds to knowledge transfers of MNE workers when moving to the domestic sector, and preserved by domestic incumbents.

A historical starting point for a skill-accumulation and skill-transfer analysis is Becker's (1962) seminal work on the wage effects of general and firm-specific human capital. Becker's benchmark model predicts that productivity and wages move in tandem in case the worker accumulates *general skills*. Since these skills are valuable for all firms in the market, a firm risks losing the returns to its investment if it also covered the costs of accumulation (through forfeiting the direct expenses and tolerating foregone revenues) and pays less than the worker's increased marginal product afterwards. Therefore the costs are borne and the gains are collected by the worker. By contrast, wages are unaffected by the accumulation of *firm-specific skills* if the risk of voluntary separation, motivated by factors other than inter-firm wage differentials, is zero.

In both of these extreme scenarios workers can move between firms without wage losses. Workers accumulating a substantial stock of general skills can earn *higher-than-average* wages in any firm and, as far as general skills are developed by informal communication between co-workers, their presence will also have a spillover effect. *Firm-specific* skills are lost with separation without an effect on wages: pre- and post-separation wages are equal, post-separation wages do not exceed the host firm's average level and do not exert influence on the earnings of co-workers.

The benchmark model's sharp distinction between general and specific knowledge has been relaxed in several ways, by Gary Becker himself in the first place, arguing that in the likely case of non-zero risk of voluntary quits the firm will share in the costs and benefits which implies lower wages in the accumulation phase and higher wages afterwards. In this case post-training involuntary separations imply a wage loss but we continue not to expect lagged returns and spillover effects. Only voluntary quits may result in *apparent* lagged returns through the effect of high pre-separation earnings on reservation wages.

The subsequent literature has been trying to reconcile the theory of *on-the-job* training with a series of facts inconsistent with the extreme scenarios. A series of empirical findings and ample everyday experience suggest that most skills are general, or at least sector- rather than firm-specific; enterprises are willing to pay for general training, and involuntary separations typically imply a loss. Acemoglu and Pischke (1998) demonstrate that in a variety of market settings such as a compressed wage structure, substantial hiring costs, information asymmetry and other labour market imperfections, workers general skills are rewarded as if they were partly specific. The 'skill-weights' model of Lazear (2009) hypothesizes that skills are general but firms attach different weights to its components. A worker who leaves a firm will have a difficult time finding another firm that can make use of all the skills he acquired at the first firm. This imposes a cost on mobile workers so the workers are unwilling to bear the full cost of training and the costs and benefits will be shared.

In these and similar settings (i) workers accumulating general and sector specific knowledge in the modern environment of MNEs are expected to earn more, than their domestic counterparts barring the youngest of them; (ii) Earnings are expected to rise with tenure; (iii) The specific components in their skills and/or the scarcity of firms applying the same skill weights as their parent company imply that MNE workers lose a part of their wage advantage after involuntary separation; (iv) The general component in their skills give rise to wage advantages in their new (domestic) workplace and exert positive influence on the productivity of their incumbent co-workers.

In search of these symptoms of a '*knowledge transfer*' scenario we estimate the instantaneous and lagged wage effects of MNE experience then the spillover effect of having ex-MNE co-workers. All models are estimated for low skilled, middling and high-skilled workers separately.

Human capital accumulation is not the only possible source of a wage gap between MNEs and domestic firms. MNEs may try to prevent leakage of information through labour turnover by paying higher-than-average wages and/or providing deferred compensation schemes. Efficiency wages induce an instantaneous gap and/or rising tenure-wage profiles in MNEs, and also imply lagged advantages by increasing reservation wages in the course of on-the-job search. Nevertheless, they are not expected to generate spillover effects.

Wage Gap

The first research question to be answered regards a contemporaneous wage gap: "Is there a difference in wages paid for a worker, if he is employed in a domestically owned firm versus when the firm is foreign owned?". Specifically, the task is to uncover a wage gap induced *solely* by firm ownership status, and disentangle it from the wage gap caused by working for more productive firms or having higher abilities. There exists a wage w_1 paid for a worker in some month by some company, when the company is owned domestically, and there exists a wage w_2 paid for the same worker in the same month, by the same company, except that now the company is owned by a foreign entrepreneur. The $w_2 - w_1$ difference gives the wage gap for our worker in some month and for some company. The (not surprising) problem with this thought experiment is that either w_1 or w_2 is latent (unobserved). Rather, to estimate the gap we have to compare workers with similar skills (to ensure that the comparison in fact makes sense) working for similar companies (to ensure that the gap is not attributed to productivity, size, etc. differentials), where the first group works for domestic firms, the second for foreign-owned firms.

We follow the literature in estimating the *foreign-domestic wage gap* by controlling for both person and firm fixed effects. Foreign owners may take over high-wage firms (ones which pay above the market average irrespective of who momentarily owns them) and hire high quality workers (who would be paid similarly high wages elsewhere). Methods originating in Abowd et al. (1999) and further developed by Cornelissen (2008), Carneiro et al. (2008), Guimaraes and Portugal (2009) (see Section 1.4 of Chapter 1) and Balazsi et al. (2015) (see Section 1.5 of Chapter 1) can help remove the impact of unobserved, time fixed worker and firm characteristics.

The worker- and employer fixed effect models identify the wage gap using information on shifts of firms and workers between ownership categories.[11] However, many multinationals settle in the country by way of greenfield investment, remain under foreign ownership 'forever' and the majority of their employees remain with them for long periods. Rather than simply neglecting the huge wage difference between 'always foreign' enterprises and their 'always domestic' counterparts we try to remove the effect of worker quality from it by studying newly established always foreign and always domestic firms and comparing the post and pre-entry earnings

[11] An alternative is restricting attention to firms undergoing acquisition as Earle and Telegdy (2008, 2013) and Earle et al. (2006, 2010) do using Hungarian data, among others.

of their incumbent workers. This approach helps to account for ownership-specific differences in unobserved worker quality (albeit it obviously fails to answer the question of how Audi or General Motors would pay their workers in the unlikely event of takeover by a local businessman).

The findings suggest that *high-skilled workers* in foreign firms earn more by 0.23 log points than their domestic counterparts even after controlling for person fixed effects. This gap is small compared to the 0.76 log points raw differential and the 0.45 log points OLS estimate, but still a sizeable one. The gap virtually disappears after adding firm fixed effects to the model, suggesting that foreign sector workers typically work in high-wage firms (be them foreign or domestic) throughout their career while the wage changes associated with their firms' changes of ownership are small. The OLS estimate of the wage gap between newly established foreign and domestic firms is similarly wide (0.7 for skilled workers) but we also find a substantial gap (0.3) between the pre-entry earnings of their high skilled workers.

Lagged Returns

"Is there a difference in the current earnings of a domestic worker when he comes from a foreign owned firm or when he comes from a domestically owned firm?" This second research question identifies a current wage differential exerted by having foreign experience. Ideally we would observe the wage paid (w_1) by some domestic firm at some point in time to the worker when the worker came from firm f when f is owned domestically, and also a wage with all being the same, except that the sending firm f is foreign owned. The gap $w_2 - w_1$ in this case measures the *lagged wage effect* on the particular worker at the given time, solely induced by having foreign experience what he can (at least partially) preserve and transfer when moving to the domestic sector. Unfortunately as we can not observe w_1 and w_2 at the same time, we have to compare domestic workers of similar skills coming from domestic firms and coming from foreign firms, where the sending firms have similar characteristics except for the ownership status.

We start with a model which compares workers, who had arrived at their domestic employers from foreign versus other domestic firms. The wage equations estimated for them are controlled for attributes of the sending and receiving firms, time spent with the sending firm, time elapsed between the two jobs and both worker and firm fixed effects. We address endogeneity of the moves from foreign to domestic enterprises by paying special attention to workers arriving from firms which discharged nearly all of their workers, that is, employees leaving their previous jobs for reasons other than insufficient individual performance, dismissal for a cause, or a desire to achieve individual wage gains.[12]

[12] The data set we use provides no direct information on the reasons of separation.

While this model has desirable qualities, it identifies the wage effect of previous outside experience from a small sample meeting a series of special requirements (the existence and observability of a previous employer, sufficient time spent there, availability of lagged and lead variables on the sending firm, and several observations within the time window on both workers and firms in order to identify fixed effects). Therefore, we estimate a second model for a much larger sample of domestic workers (including information from left-censored employment spells), which compares domestic workers with *past and future* foreign experience to workers with *past and future* involvement in domestic enterprises other than their current employer. This approach is closest in spirit to models, which study the wage effect of incarceration by comparing past and future convicts (see Grogger, 1995, LaLonde and Cho, 2008, Pettit and Lyons, 2009 and Czafit and Köllő, 2015) under the assumption that the dates of incarceration can be treated as random. This model can also control for unobserved quality differentials but it cannot address the possibly endogenous selection of workers to separating from their employers.

We find that skilled people leaving collapsing or relocating MNEs earn significantly more than those leaving similar domestic enterprises (12.6 and 11.6 per cent after controlling for worker fixed affects and both worker and firm fixed effects, respectively). The *average* skilled ex-MNE worker's advantage is much smaller (4 per cent) that we attribute to the negative selection of those, who exit slightly downsizing or expanding firms in the well-paying foreign sector. The comparable estimate from the 'overlapping cohorts' model, which includes a much higher number of long-lasting spells at multinationals is larger, close to 7 per cent in the case of skilled workers.

Spillover Effects

"Is there a difference in the wages paid for a employee with no foreign experience when his colleague has foreign experience versus if he has not?" This last research question concerns with the transferability of the knowledge premia between workers. Ideally we would observe the wage of the domestic employee when his colleague has foreign experience (w_2) and also the wage when the colleague has always been employed by domestically owned companies (w_1). The difference $w_2 - w_1$ identifies the *spillover effect* of the colleague on the domestic worker, solely induced by foreign experience. Unfortunately, as before, both w_1 and w_2 can not be observed, so we compare the wages of domestic workers with similar colleagues when the colleagues have or do not have foreign experience. The equations to be regressed are also controlled for both person and firm fixed effects. The estimated spillover effects might seem relatively weak at first sight (the estimates fall short of a one per cent wage surplus in response to a one standard deviation difference in the share of ex-foreign co-workers) but they exceed rather than lag behind similar estimates in the literature.

The rest of the study is organized as follows. Section 4.3.2 gives a short introduction to FDI in Hungary. Section 4.3.3 introduces the data while Section 4.3.4 discusses estimation issues. Section 4.3.5 presents the results of the models estimated for low, middle and high skilled workers, and finally, Section 4.3.6 discusses the results and draws conclusions. The text is supplemented with Appendix A containing supplementary tables, and a Data Appendix, Appendix B, describing the variables.

4.3.2 FDI in Hungary

In the first decade after the start of the transition, Hungary was the most successful country within the former Soviet bloc in attracting foreign capital. By the beginning of our period of observation, cumulative FDI inflows exceeded 40 per cent of the GDP (UNECE, 2001), multinationals employed 15 per cent of the labour force (including self-employment and the public sector into the denominator) and more than 30 per cent of private sector employees. They produced 20 per cent of the GDP and delivered over 2/3 of the exports (Balatoni and Pitz, 2012). Large multinationals including Audi, General Motors and Suzuki dominated the motor industry and foreign presence was already decisive in the tobacco, leather, chemical, rubber and electronics industries, with employment shares between 50 and 80 per cent.

Almost three-fourth of the cumulative FDI inflows have arrived to sectors outside manufacturing and, as shown in Table 4.8, over 40 per cent of the workforce within the foreign sector were employed in the tertiary sector.[13] Therefore, we do not restrict the analysis to manufacturing as most papers do in the strand of the literature we follow. (See Barry et al. (2005), Görg and Strobl (2005), Lipsey and Sjöholm (2004) and Smarzynska Javorcik (2004) as opposed to Poole (2013) whose study covers all sectors in Brazil). While FDI typically boosts exports and may generate demand for domestic manufacturers producing intermediate goods, its contribution to the quality of retail trade, banking and services can be equally important for the host country. This is a particularly important aspect in the former state socialist countries, which started the transition with critically undeveloped non-tradeable sectors.

Furthermore, we believe that the scope for spillovers is wider in the tertiary sector. There are few domestic manufacturers outside a small group of local suppliers

[13] The quoted figures are the authors' calculations using the main data set of the study and relate to workers in firms employing more than 10 workers.

Table 4.8 *Foreign ownership in the estimation sample in 2003*

	The share of majority foreign-owned firms in person-months observed in the estimation sample (per cent)	Industrial distribution of person-months within the foreign-firm sector (per cent)
Agriculture[a]	4.5	0.7
Manufacturing	45.8	58.6
Construction	7.0	1.5
Energy	42.9	3.6
Trade	26.6	15.0
Finance	78.1	6.6
Services	23.0	14.0
Average/Total	33.4	100.0

[a] The data are annual averages and relate to workers in foreign and domestic firms employing more than 10 workers. The number of observations is 4,355,581 (first column) and 1,454,284 (second column)

to learn from Audi and other car manufacturers but there are tens of thousands of local tradesmen, craftsmen, guesthouse owners and shop keepers to absorb knowledge from Auchan, Accor, Strabag and other multinationals in retail trade, services, catering and construction.

4.3.3 Data

Our estimation samples are drawn from a large longitudinal data set covering a randomly chosen 50 per cent of Hungary's population aged 5–74 in 2003. The data collects information from registers of the Pension Directorate, the Tax Office, the Health Insurance Fund, the Office of Education, and the Public Employment Service. Each person in the sample is followed, on a monthly basis, from January 2003 until December 2011 or exit from the registers for reasons of death or permanent out-migration.

The data provides information on the highest paying job of a given person in a given month, days in work and amounts earned in that job, occupation and type of the employment relationship. Financial data of the employer are available for incorporated firms. Furthermore, we have data on registration at a labour office, receipt of transfers and several proxies of the person's state of health. Gender and age are observed but educational attainment is not – this is approximated with the highest occupational status achieved in 2003–2011. A detailed description of the variables used in the paper is presented in Appendix B.

Our analysis is restricted to workers employed with a labour contract at least once in a foreign or domestic private enterprise the maximum employment level of which exceeded the 10 workers limit at least once in 2003–2011. We have several reasons to set a size limit: (i) foreign firms are nearly absent in the small firm sector[14]; (ii) financial data are not available for sole proprietorships and unincorporated small businesses; (iii) the financial reports of incorporated small firms are often incomplete and erroneous; and (iv) their wage data are flawed by the massive presence of 'disguised' minimum wage earners.[15] Furthermore, the inclusion of small firms would substantially raise the risk of measurement error in the analysis of spillover effects since the probability of not observing an ex-MNE employee is much higher in small establishments. Moreover, we removed workers and firms with less than two data points, zero wages and missing covariates.

After these steps of data cleaning we are left with a sample of 92,663,887 person-months belonging to 1,762,812 workers and 218,572 firms.[16] Out of the 1.76 million workers in the sample, 41.3 per cent had at least one spell of employment in the foreign sector and 15.2 per cent did not work outside the foreign sector at all in 2003–2011. We draw special sub-samples from this starting population for the study of lagged returns and spillover effects. Descriptive statistics of the main data set are presented in Table A.1 of Appendix A.

4.3.4 Estimation Strategies

Foreign-domestic Wage Gap

Estimates on All Firms. Different specifications of our first model estimate the foreign-domestic wage gap in the following way:

$$\log w_{ijt} = \delta F_{ijt} + \beta_1 P_i + \beta_2 X_{it} + \beta_3 Y_{ijt} + \beta_4 Z_{jt} + [\alpha_i + \gamma_j] + \varepsilon_{ijt} \tag{4.2}$$

where w_{ijt} is the daily average earnings of person i at firm j and month t, F is a dummy for being employed in a majority foreign owned firm, P_i and X_{it} are time fixed and time varying individual attributes (gender, age, time spent unemployed, receiver of disability payment, etc.), Y_{ijt} stands for job-specific variables (occupation, tenure, etc.), Z_{jt} denotes time varying firm-specific covariates (firm size, exporter, capital-labour ratio, etc.), α_i, γ_j are person, firm fixed effects, respectively, and ε_{ijt} is the white noise disturbance term.

We estimate (4.2) first with pooled OLS, excluding α_i and γ_j, then with Within

[14] In 2014, foreign enterprises had a mere 4.5 per cent share in the 1–10 workers category. (Authors' calculation based on the 2014 Q4 wave of the Labor Force Survey).

[15] This term hints at the practice of paying workers the minimum wage (subject to taxation) and the rest of their remuneration in cash. Elek et al. (2012) estimate that in 2006 the share of workers paid in this way amounted to 20 per cent in firms employing 5–10 workers, 10 per cent in slightly higher firms (11–20 workers) and less than 3 per cent in larger enterprises.

[16] Hungary is a country of 10 million inhabitants. Dependent employment varied between 3.3 and 3.4 million in our period of observation. See http://www.mtakti.hu/file/download/mt_2014_hun/statisztika.pdf

estimator with worker fixed effects only, and both worker and firm fixed effects as well.

When the equation is estimated with pooled OLS, the δ parameter captures the pure ownership effect, plus the employment duration weighted average residual worker and firm effects given personal characteristics P and X (Abowd et al., 2006b). The person fixed effect absorbs the unobserved time invariant mean 'qualities' of workers but the estimated gap is still affected by the employment-duration weighted average of the firm effects for the firms in which the worker was employed. When both person and firm fixed effects are included, δ captures a pure ownership effect identified from worker flows between ownership categories, on the one hand, and changes in ownership, on the other. It shows the wage advantage of a foreign firm employee over a domestic worker with similar observable attributes, controlled for their average wages in the entire period of observation and also controlled for average wages of the firms where they worked during the period of observation. We control for unobserved shocks to productivity by including indicators of the firm's capital-labour ratio, exports and size, and adding sector-year interactions to the equations. Standard errors are adjusted for clustering by firms. The coefficient δ is identified from any variation in F when estimated with Pooled OLS; from (jt) variation in F for at least one worker i when estimated with worker fixed effects; and from (jt) variation in F for at least one worker *i and* from (it) variation in F for at least one firm j when estimated with both firm and worker fixed effects.[17]

Problems with the Model of Two Fixed Effects. The identification of fixed effects requires that workers move across ownership categories and some firms change majority owner. Table 4.9 seems to suggest that these requirements are met: we observe nearly 1 million workers, who changed employer in the period of observation, with close to half million of them crossing ownership boundaries. Furthermore, we observe 3,389 firms whose majority owner changed at least once in 2003–2011. While these figures might seem impressive, and conditions for the identification are clearly met, the table makes clear that the majority of workers and a vast majority of firms play no role in the identification of ownership-specific wage differentials when these are estimated with fixed effects models. This problem is particularly grave in the case of firm fixed effects since we observe a relatively small number of enterprises that switched from foreign to domestic ownership and vice versa (a mere 1.7% of the firms in the sample). Though this issue will not leave the model unidentified, we expect the firm fixed effect to pick up most of the wage gap.

[17] A within group variation of F with respect to i means that firm j changed ownership while i was working there, or that i moved between firms of different ownerships. A within group variation of F with respect to j means that the firm changed ownership at least once during the period of observation.

10.14754/CEU.2017.05

Table 4.9 *Firms and workers by type of mobility in the estimation sample*

Firms[a]	Number	%	Workers	Number	%
Always D	197,246	91.8	Inc. in F firms	214,075	12.2
Always F	14,171	6.6	Inc. in F sector	82,921	4.7
F to D, once	1,539	0.7	Inc. in D firms	539,393	30.7
D to F, once	1,605	0.8	Inc. in D sector	439,915	25.1
Multiple shifts	245	0.1	Switched sector	487,351	27.3
Total	214,806	100.0	Total	1,754,655	100.0

[a] F and D are abbreviations for foreign and domestic ownership, respectively. F to D (D to F) denote firms that changed majority owner once. 'Multiple shifts' stands for firm changing ownership more than once. 'Inc.' stands for incumbents, *i.e.*, workers who did not change employer (incumbents in firms), or, did change employer but did not leave their sectors of origin (incumbents in sectors).

Second, it comes as no surprise that mobile workers, on the one hand, and firms changing owners, on the other, are non-randomly selected, as suggested by a preliminary overview of average earnings by type of mobility (Table 4.10). Workers permanently employed in the foreign sector earn twice as much as incumbent domestic workers do. Those switching sectors earn more than their incumbent counterparts in the domestic sector but less than their incumbent co-workers in the foreign sector. Similarly, firms under majority foreign ownership throughout the period of observation pay 77 percentage points higher wages than their 'always domestic' counterparts. Firms switching majority owner pay only about 40 per cent higher wages than 'always domestic' firms and about 30 per cent lower wages than 'always foreign' companies.

The raw data foreshadow that wage changes associated with shifts between sectors and changes of ownership will prove to be small compared to the tremendous ownership-specific differences between incumbent workers and firms. The OLS equations capture these differences but do so without isolating the influence of unobserved differences in worker quality and enterprise wage policies. The fixed effects models can solve the latter problem but improvements in model quality come at the cost of distortions in the sample.

Estimates on Newly Established Incumbent Firms. We try to identify the sources of the large gap between incumbent firms by looking at foreign and domestic firms established after 2003, which did not change owner until the end of 2011, and their incumbent workers, who stayed with them until 2011 or disappearance of the enterprise from the data set.

We base the definition of a 'new firm' on its employment dynamics rather than its date of registration, since the latter is often associated with break-ups, mergers

Table 4.10 *Average wages of firms and workers by type of mobility (Fraction of the national average wage, group-level means in 2003-2011)*

Firms[a]		Workers[b]	
Always F	1.31	Incumbent in F firm	1.43
Always D	0.54	Incumbent in F sector	1.49
Switched from F to D once			
wage while F	0.96	Incumbent in D firm	0.70
wage while D	0.96	Incumbent in D sector	0.63
Switched from D to F once		Switched sector	
wage while F	1.00	wage while in F	1.07
wage while D	0.98	wage while in D	0.90
Multiple shifts			
wage while F	0.97		
wage while D	0.94		

[a] Firms: F and D are abbreviations for foreign and domestic ownership. 'Multiple shifts' stands for firms changing ownership more than once. The figures are unweighted group means of the enterprise average wages.
[b] Workers: 'Incumbent' stands for workers who did not change employer (incumbents in firms), or, did change employer but did not leave their sectors of origin (incumbents in sectors).

and acquisitions rather than actual birth of a new economic actor. We rely on the fact that a mid-sized or large firm's creation typically starts with hiring a small group of managers who arrange the start-up. This preparatory stage is followed by a 'big bang' when '*rank-and-file*' employees are hired. We speak of a big bang when a firm's staff jumps from an initial level of $L_{t-1} \leq 5$ to $L_t \geq 50$. We find 317 such firms (with no subsequent change of ownership) in the data.[18] Combined employment in these enterprises jumped from 6,728 one year before the big bang to 126,544 one year after the big bang (*i.e.*, an estimated growth from 13 to 253 thousand taking into account the 50 per cent sampling quota).

We estimate model (4.2) on a restricted data set to identify a foreign-domestic wage gap for incumbent workers in new firms (i) during their service in these firms; (ii) prior to being hired by these firms, and interpret the pre-entry wage differential as a signal of difference in worker quality. Since assignment to the groups compared is person-specific, and the firms do not change owner, we estimate the wage gap with pooled OLS but include firm fixed effects in one of the model variants.

[18] To be more precise, we found 311 firms meeting the above criteria and manually reclassified another 6 firms jumping from the range of $5 < L_{t-1} < 50$ to levels of hundreds or thousands.

Lagged Returns

Domestic workers arriving from MNEs earn higher wages than their colleagues by 9 per cent (14 per cent in the case of high skilled workers). These raw advantages are affected by compositional and quality differences between workers and potentially biased by the non-random selection of workers to transition from foreign to domestic firms. The direction of the bias is uncertain. The wage advantage of ex-foreign workers may arise because those, who leave the foreign sector do so in response to meeting exceptionally favorable wage offers. Alternatively, workers who quit the well-paying foreign sector may be negatively selected *i.e.*, fired for insufficient individual performance or dismissed for a cause. We leave it to the data to tell which of these patterns dominate.

Domestic Workers Arriving from Foreign Versus Domestic Firms. Our first model compares domestic workers having arrived to their current domestic employers from foreign versus domestic firms:

$$\begin{aligned}\log w_{ijt} \quad &= \delta_1 AfterF_{ijt} + \delta_2 \mathrm{d}L_{jt} + \delta_3 (AfterF_{ijt} \times \mathrm{d}L_{jt}) \\ &\quad + \beta_1 P_i + \beta_2 X_{it} + \beta_3 Y_{ijt} + \beta_4 Z_{jt} + [\alpha_i + \gamma_j] + \varepsilon_{ijt}\ .\end{aligned} \tag{4.3}$$

We regress the wages of newcomers on their personal characteristics, current and past job attributes, months between the two jobs, selected indicators of the sending and receiving firms, sector-year interactions and worker and firm fixed effects, which remove the average wages these workers earned during their career within the domestic sector and the wage levels of the domestic firms where they worked. The estimation is restricted to workers who have spent at least two years in the sending firm, a period long enough to absorb some knowledge. We pay special attention to workers whose previous employer collapsed so they had to leave involuntarily, for reasons other than their personal deficiencies or wish to earn higher wages.

After F is a dummy variable set to 1 for workers, who arrived from foreign firms and 0 for workers arriving from domestic companies. $\mathrm{d}L = L(t+1)/L(t-1)$ measures the change of employment in the sending firm between month $t-1$ and $t+1$, with t denoting the month when the worker left that firm. Controls in P, X, Y and Z are the same as for equation (4.2). The coefficient of dL, δ_2, measures how wages vary with employment dynamics of the sending domestic firms while, δ_3, the parameter of the interaction term $AfterF \times \mathrm{d}L$ captures the effect of dL among workers arriving from foreign employers. The wage advantage of workers arriving from foreign firms over workers arriving from domestic firms, conditional on employment dynamics of the sending firms and on various worker and firm characteristics, is given by $\delta_1 + \delta_3 \cdot \mathrm{d}L$. δ_1 and δ_3 are identified if (i) in case of estimation with worker fixed effects: a worker does multiple moves to the domestic sector involving domestic and foreign past employers; (ii) in case of estimation with both worker and firm fixed effects: along with the condition in (i), for at least one firm at least one worker comes from the domestic sector, at least one worker comes from the foreign sector.

As was mentioned earlier, this model severely restricts the sample for which the equations can be estimated. Members of the sample must have at least two employment spells in the domestic sector and a minimum of 3 jobs in 9 years in case of workers with foreign experience. We need to observe employment dynamics of the previous employer. To make sure that a worker had at least 2 year's tenure with his/her previous employer the time window should be narrowed to 2005–2011. We want to compare workers arriving at their current firms via employer change, so we exclude those getting to the domestic sector due to domestic buyout. These requirements are met by only 29 792 workers with 433 477 monthly observations. This specification we use would be first best in a wide time window comprising several decades but in the one we actually have identification comes from a relatively small sample of workers with short employment spells.

Domestic Workers Before and After Working for Multinationals. Our second model also relates to the employees in domestic firms. Each month t we identify workers whose previous employer was a foreign firm ($AfterF_{ijt}$) and those, whose next employer is known to be a foreign firm and had no previous experience in the foreign sector ($BeforeF_{ijt}$). Furthermore, we single out workers whose previous employer was domestic but had foreign sector experience earlier ($earlierF_{ijt}$) and those who arrived at the domestic sector via acquisition. Finally, we split the rest of the workers to groups with past and future experience in domestic firms other than the current one ($AfterD_{ijt}$ and $BeforeD_{ijt}$). The reference category consists of workers, who had no contact with other firms in 2003–2011. We plug these variables to the OLS wage equation

$$\begin{aligned} \log w_{ijt} &= \delta_1 AfterF_{ijt} + \delta_2 BeforeF_{ijt} + \delta_3 AfterD_{ijt} + \delta_4 BeforeD_{ijt} \\ &\quad + \beta_1 P_i + \beta_2 X_{it} + \beta_3 Y_{ijt} + \beta_4 Z_{jt} + \varepsilon_{ijt} \ . \end{aligned} \tag{4.4}$$

and measure the effect of foreign sector experience with the double difference $\delta = (\delta_1 - \delta_2) - (\delta_3 - \delta_4)$. δ_1 and δ_2 are identified from movements *between* sectors, while δ_3 and δ_4 are identified from movements *within* sectors. The diff-in-diff coefficient δ is interpreted as the effect of foreign experience on current wages, removing the effect of future foreign experience (a signal for higher ability), removing the effect of domestic future and past experience other then the current employer (wage increment due to experience, tenure, etc.), and controlling for earlier past foreign experience, foreign to domestic buyout, as well as for various individual and job characteristics.

We see no strong arguments against the assumption of random timing of the moves between the foreign and domestic sectors. In our sample, the share of foreign firms in employment fluctuated in a narrow range around 32 per cent in 2003–2011 and we have no information of strong shocks that could have altered the patterns of hiring and firing before the crisis. In order to account for the possibility that these patterns changed after 2009, we also estimate the model excluding the crisis period. Furthermore, we start the analysis only in 2005 in order to make sure that the new hires had sufficiently long experience in their previous jobs.

While this model is able to remove the effect of unobserved worker quality it can not control for endogeneity of the move from foreign to domestic firms since the sample includes left-censored spells at the current domestic employer.

Spillover Effects

We estimate spillover effects for the sample of domestic workers with no previous foreign experience. Their wages are regressed on a set of controls and the variable *ShareF* measuring the share of workers with previous foreign experience within the worker's company and skill category:

$$\log w_{ijt,s} = \delta ShareF_{jt,s'} + \beta_1 P_{i,s} + \beta_2 X_{it,s} + \beta_3 Y_{ijt,s} + \beta_4 Z_{jt,s} + \alpha_{i,s} + [\gamma_{j,s}] + \varepsilon_{ijt,s} \,. \tag{4.5}$$

We examine how the shares in skill categories $s' = \{1,2,3\}$ affect the wages of incumbents in skill categories $s = \{1,2,3\}$ so we arrive at a (3×3) matrix of parameter estimates. We estimate the models with both worker and worker and firm fixed effects including the controls used to estimate models (4.2), (4.3) and (4.4). The coefficient of main interest δ is identified from movements from the foreign sector to the domestic, and is interpreted as the effect of a unit change in the share of colleagues with foreign experience within skill category s' on the wages of incumbent workers within skill category s, controlling for various worker and firm characteristics, as well as for the average wages of workers (in case of the model with worker fixed effects), and for the average wages of workers and within a firm (in case of the model with both worker and firm fixed effects).

On Estimating More Than One High-dimensional Fixed Effects

The incomplete estimation method detailed in Section 1.5 of Chapter 1 (or in Balazsi et al., 2015) is designed to break down giant matrix calculations to smaller components, while constantly keeping an eye on optimality. As in case of more than one fixed effects, one still have to directly store and invert matrices as large as the number of the second largest fixed effect, dimensionality can still circumvent its application. In such cases, the suggestive solutions of Guimaraes and Portugal (2009), or Carneiro et al. (2008) are well taken, which, despite being iterative (that is, only precise after converging), completely avoid dimensionality. Obviously, both methods eventually reach the same estimates, but they do so in fundamentally different ways: rapidly, by pressing a little more on memory (Balazsi et al.) and memory-friendly, by pressing on time (Guimaraes and Portugal and Carneiro et al.).

Selecting Periods for the Estimations

While our data yields a detailed view of developments between January 2003 and December 2011, it provides no information about any event taking place before 2003. It is particularly painful that we have no data on the start date of left-censored employment spells. Therefore, we will choose shorter estimation periods (2005–2011 and 2005–2009) in cases the models require information on past and/or future states or events. We also have to consider the possibility that the patterns of wage setting changed during the economic crisis. Previous research (Köllő, 2011) suggested that Hungarian private firms kept wages and relative wages virtually constant in 2008–2009 and large scale wage adjustment started only in 2010. Therefore we treat 2010–2011 as the crisis period.

4.3.5 Results

The Foreign-Domestic Wage Gap

Results on All Firms. Estimates of the foreign-domestic wage gap on model (4.2) are summarized in Table 4.11. A full estimation outcome for the entire sample is shown in A.2 of Appendix A. We see that the estimated effects of the control variables are 'properly' signed, have reasonable magnitudes and very low standard errors resulting in two and three digit t-values. The models also fit the data well, with the R^2 statistics typically exceeding 0.4 in the fully controlled OLS regressions and 0.25 (overall R^2) in the fixed effects panel models. In Table 4.11, only the coefficients of the 'foreign' dummies and their corresponding t-statistics are presented.

As shown in the first column, the raw gap controlled only for sector-year interactions exceeds 0.75 in the case of high skilled employees while it is narrower in the case of medium and low skilled workers (0.41 and 0.35 log points, respectively).

Controlling for individual attributes (gender, age, proxies of educational attainment, state of health, risk of unemployment, receipt of care allowances and disability payment) exerts weak influence, if any, on estimates of the wage gap. Even after the inclusion of occupation dummies and tenure, the estimated gaps reach 0.44 log points in the entire sample, 0.7 for high-skilled workers and 0.3–0.4 log points for other groups of workers.

About half of these differences vanish after allowing for the effects of firm size,

Table 4.11 *Estimates of the foreign-domestic wage gap by skill levels*

Specification[a]	(1)	(2)	(3)	(4)	(5)	(6)
Controls	none	X_{it}	X_{it}, Y_{ijt}	X_{it}, Y_{ijt}, Z_{jt}	α_i	$\alpha_i + \gamma_j$
Estimator	OLS	OLS	OLS	OLS	Within	Within
2003–2011						
Entire sample	0.525	0.488	0.439	0.276	0.151	0.014
	(18.05)	(21.96)	(25.04)	(18.01)	(8.49)	(4.34)
Unskilled	0.353	0.354	0.289	0.202	0.113	0.011
	(12.96)	(16.00)	(14.84)	(9.94)	(2.92)	(1.99)
Med. skilled	0.407	0.400	0.360	0.215	0.120	0.012
	(17.24)	(20.83)	(20.75)	(13.83)	(7.44)	(3.39)
High skilled	0.755	0.753	0.696	0.448	0.234	0.020
	(22.18)	(25.44)	(29.10)	(22.82)	(11.15)	(7.54)
2005–2009						
Entire sample	0.518	0.476	0.424	0.274	0.140	0.014
	(16.51)	(20.34)	(22.44)	(17.34)	(5.11)	(3.19)
Unskilled	0.348	0.346	0.278	0.198	0.103	0.011
	(11.77)	(14.29)	(13.20)	(8.82)	(1.48)	(1.34)
Med. skilled	0.394	0.387	0.345	0.208	0.110	0.013
	(15.38)	(18.97)	(18.45)	(12.76)	(4.43)	(3.13)
High skilled	0.746	0.738	0.681	0.454	0.214	0.019
	(20.74)	(23.67)	(26.72)	(22.32)	(6.68)	(5.67)

[a] All coefficients are significant at 0.01 level. t-values are in brackets. Dependent variable: log daily earnings in the given month. Reference category: domestic firm employees. Sample: persons employed in firms employing more than 10 workers at least once in 2003–2011. The data comprise 92,118,857 worker-months belonging to 1,758, 834 workers in 218,572 firms. All equations, including (1) are controlled for sector-year dummies. Other controls include: (X): gender, age, age squared, Budapest dummy, log regional unemployment rate, receipt of care allowance and/or disability payment, log monthly expenditures on health services and medicine. (Y): job tenure, occupation. (Z): log firm size, log capital-labour ratio and the share of exports in sales revenues. Specifications (5) and (6) include only time-varying covariates and worker and firm fixed effects (α_i, γ_j). These specifications were estimated with Within estimators.

export share, the capital-labour ratio and industrial affiliation, suggesting that a large part of the foreign-domestic wage gap is explained by the bigger size and higher capital endowments of multinationals, and their much higher propensity to export. However, even after controlling for the most important firm-level observables, a sizeable advantage remains on the part of foreign firms – a gap of 0.28 log points in the whole sample and 0.45 in the case of skilled workers.

After allowing for worker fixed effects, the wage gap falls substantially, to only 13 per cent in the entire sample, 23 per cent for skilled and 11 per cent for unskilled workers, suggesting that about half of the residual wage gap is accounted for by the higher quality (generally higher wage level) of workers hired by multinationals. Finally, the wage gap virtually disappears (falls 1 per cent in the whole sample and

2 per cent for skilled workers) when we also include firm fixed effects *i.e.*, take into account the level of wages paid by foreign firms in periods when they were under domestic control and vice versa. This result suggests that the wage gain of workers from foreign ownership mostly stems from the permanently high level of their employers' average wages rather than large fluctuations in their pay levels depending on who transitorily own them. However, as was previously discussed, this result is based on information relating to a relatively small and non-randomly selected set of companies undergoing foreign or domestic acquisition during the period of observation.

Results on New Firms. Incumbent workers in newly established foreign firms earn more than their domestic counterparts by more than 0.5 log points in general and 0.7 log points in the case of skilled labour, as shown in the first row of Table 4.12.

Estimates in the lower block of the table indicate that the workers of new foreign firms also earned more than their domestic counterparts prior to their entry to the new firms, by 0.22 log points on average. The pre-entry wage gap is wider for high skilled than medium skilled employees (0.36 and 0.11 log points, respectively) and much narrower for the low skilled (0.04 log points).

The pre-entry advantage partly stems from foreign start-up's higher propensity to hire skilled workers from other foreign firms as shown by a comparison of the second and third rows. Even more importantly, foreign start-ups seem to hire skilled employees from high-wage firms in general: with the inclusion of firm fixed effects the pre-entry advantage falls to 0.09 log points for all skill categories and 0.12 log points for high skilled workers. These results suggest that those foreign enterprises, which tend to appear as incumbents in a narrow time window indeed pay significantly higher wages than similar domestic businesses and this remains true after allowing for the fact that they lure high wage workers from high wage firms. An ownership-specific wage differential of about 0.3 log points remains between incumbent foreign and domestic workers even after removing the pre-entry pay differential between them.

Lagged Returns to Foreign Firm Experience

Domestic Workers Arriving from Foreign versus Domestic Firms. Table 4.13 and 4.14 summarize the parameter estimates on model (4.3) with worker, and both worker and firm fixed effects of the key variables. The tables present the marginal effects at different levels of dL below the blocks containing the parameter estimates. Since now we work with much smaller samples than before, we also indicate significance levels.

Starting with the last column, which summarizes the results for skilled workers, the estimates suggest a 4.1 per cent advantage on the part of ex-foreign employees

Table 4.12 *Comparing incumbent workers in newly created incumbent firms before and during their service in these firms: Pooled OLS estimates of the ownership-specific wage differences*

Wage advantage[a]	Low skilled[b]	Middling	High skilled	Entire sample
After entry to the new firm	0.317	0.394	0.721	0.529
	(82.0)	(167.9)	(125.5)	(218.7)
Person-months	59,049	234,515	78,622	372,186
aR^2	0.531	0.473	0.432	0.437
Before entry to the new firm				
Control A[c]	0.039	0.114	0.360	0.223
	(10.5)	(62.0)	(84.4)	(126.8)
Person-months	94,989	429,706	136,712	661,407
aR^2	0.369	0.307	0.375	0.312
Control B[d]	0.015	0.090	0.294	0.182
	(4.1)	(49.5)	(71.0)	(107.1)
Person-months	94,989	429,706	136,712	661,407
aR^2	0.392	0.328	0.416	0.343
Control C[e]	-0.056	0.027	0.119	0.086
	(12.1)	(12.3)	(23.2)	(45.7)
Person-months	94,989	429,706	136,712	661,407
aR^2	0.774	0.701	0.817	0.706

[a] of incumbent workers employed in new foreign-owned firms over incumbent workers in new domestic firms
[b] All coefficients are significant at the 0.01 level. The equations have been controlled for gender, age, age squared, months spent non-employed in 2003–2011, log health expenditures, receipt of disability payment, receipt of care allowances, tenure, Budapest dummy, log regional unemployment rate, 63 sector-year interactions and the log size, log capital-labour ratio and export shares of the current employer.
[c] Uncontrolled for ownership of the current employer.
[d] Controlled for ownership of the current employer.
[e] Controlled for firm fixed effects.

over workers arriving from domestic firms in the model with worker fixed effects and 3.6 per cent in the model with both worker and firm fixed effects. However, the advantage varies strongly with employment dynamics of the sending firm: it amounts to 12-13 per cent at $\mathrm{d}L = 0$ and only about 2 per cent at $\mathrm{d}L = 1.2$.[19] We interpret this finding as a signal of predominantly negative selection: workers who

[19] Separate estimates by categories of dL are broadly consistent with these results. The wage gap between highly skilled ex-foreign and ex-domestic workers is 13.4 per cent in case their sending firms lost more than half of their employees, as opposed to only 1.9 per cent in the range of $0.49 < \mathrm{d}L < 0.99$ and -1.7 per cent when $\mathrm{d}L > 1$.

Table 4.13 *The wage advantage of ex-foreign workers controlled for employment change in the sending firm, 2005–2011, worker fixed effects*

	Entire sample[a]	Unskilled	Medium skilled	High skilled
After F	0.041***	0.030**	0.010**	0.126***
	(11.1)	(2.6)	(2.1)	(15.7)
dL of the sending firm	−0.021***	−0.014	−0.014***	−0.053***
	(6.7)	(1.4)	(4.0)	(8.5)
After $F \times$ dL of the sending firm	−0.019***	0.001	0.001	−0.085***
	(5.7)	(0.1)	(0.3)	(16.9)
Marginal effect of foreign experience at different levels of dL (per cent)				
0	4.1	3.0	1.0	12.6
0.3	3.5	3.0	1.0	10.1
0.6	3.0	3.1	1.1	7.5
0.9	2.4	3.1	1.1	5.0
1.2	1.8	3.1	1.1	2.4
At the group-level median of dL	2.2	3.1	1.1	4.1
R^2 overall	0.056	0.001	0.037	0.191
Person-months	462,608	46,761	273,178	142,669
Persons	30,343	3180	18,605	8588

[a] Coefficients and t-values in parentheses. Significant at the 0.1 (*), 0.05 (**), 0.01 (***) level. Sample: The sample consists of domestic firm employees with more than two year's tenure at a previous employer. Key variables: *After F* denotes workers arriving from foreign employers. dL measures the change of employment in the sending firm between year $t-1$ and $t+1$ (d$L = L(t+1)/L(t-1)$), where t is the year of the worker's separation from that firm. Reference category: domestic workers arriving from a domestic firm. Controls: The equations are controlled for age, age squared, health expenditures, receipt of disability and care allowances, completed duration of the previous employment spell, months of non-employment between exit from the previous employer and entry to the current one, tenure at the current employer, log regional unemployment rate, year dummies and the size, capital-labour ratio, export share and industrial affiliation of the sending and the receiving firms.

leave expanding or only slightly downsizing firms are likely to have left involuntarily and tend to be selected from among less productive employees.

The estimates for medium skilled and unskilled workers indicate minor advantages on the part of ex-MNE employees, which practically do not vary with employment dynamics of the sending firm.

Domestic Workers Observed Before and After Working for Multinationals. Our second model identifies the private returns to foreign firm experience by comparing the wages of domestic firm employees with past and future experience in multinational and domestic companies. We restrict the analysis to 2005–2009 in order to have sufficient observations on both *past and future* experience outside the workers' current firms. The estimates relate to nearly 31 million monthly observations

Table 4.14 *The wage advantage of ex-foreign workers controlled for employment change in the sending firm, 2005–2011, worker and firm fixed effects*

	Entire sample[a]	Unskilled	Medium skilled	High skilled
After F	0.037***	0.031***	0.008*	0.116***
	(10.4)	(2.7)	(1.8)	(15.1)
dL of the sending firm	−0.041***	−0.109***	−0.021*	0.043
	(4.2)	(3.3)	(1.9)	(1.1)
After $F \times$ dL of the sending firm	−0.017***	0.002*	0.002	−0.080***
	(5.4)	(0.3)	(0.6)	(16.2)
Marginal effect of foreign experience at different levels of dL (per cent)				
0	3.7	3.1	0.8	11.6
0.3	3.2	3.2	0.9	9.2
0.6	2.7	3.2	0.9	6.8
0.9	2.2	3.3	1.0	4.4
1.2	1.7	3.3	1.0	2.0
At the group-level mean of dL	2.0	3.3	1.0	3.6
Adjusted R^2	0.941	0.875	0.911	0.955
Person-months	462,608	46,761	273,178	142,699
Persons	30,343	3180	18,605	8858

[a] Coefficients and t-values in parentheses. Significant at the 0.1 (*), 0.05 (**), 0.01 (***) level. Sample: The sample consists of domestic firm employees with more than two year's tenure at a previous employer. Key variables: *After F* denotes workers arriving from foreign employers. dL measures the change of employment in the sending firm between year $t-1$ and $t+1$ (d$L = L(t+1)/L(t-1)$), where t is the year of the worker's separation from that firm. Reference category: domestic workers arriving from a domestic firm. Controls: The equations are controlled for age, age squared, health expenditures, receipt of disability and care allowances, completed duration of the previous employment spell, months of non-employment between exit from the previous employer and entry to the current one, tenure at the current employer, log regional unemployment rate, year dummies and the size, capital-labour ratio, export share and industrial affiliation of the sending and the receiving firms.

belonging to 1,431,706 workers in 204,064 firms. The coefficients of interest are summarized in Table 4.15.

Starting with the results on high-skilled workers (last column), we observe that those having arrived to the domestic sector from a foreign firm earn more by 14.5 per cent than their observationally similar counterparts, who arrived from a domestic enterprise. This difference clearly overestimates the returns to foreign sector experience since those domestic workers, who are on their way to a foreign firm (*Before F*) also earn more by 7.7 per cent than those, who are about to leave for a domestic employer (*Before D*): the former earn 3.5 per cent more than the incumbents while the latter earn less by 4.2 per cent. These data suggest that about half of the 14.5 per cent difference between the ex-foreign and ex-domestic entrants arises from a selection effect.

Put differently, workers having arrived to their domestic employer from a for-

Table 4.15 *The wages of workers arriving from/leaving for foreign and other domestic enterprises. 2005–2009, domestic firm employees – Pooled OLS regression coefficients*

	Entire sample[a]	Unskilled	Medium skilled	High skilled
After F	.045	.044	.025	.116
	(22.3)	(8.9)	(11.5)	(22.4)
Before F	−.006	−.035	−.014	.035
	(2.7)	(6.0)	(6.0)	(7.2)
After D	−.031	−.027	−.021	−.029
	(21.2)	(8.3)	(12.9)	(10.2)
Before D	−.062	−.069	−.065	−.042
	(38.8)	(20.4)	(35.3)	(18.2)
F to D buyout	.127	.001	.102	.178
	(112.8)	(0.2)	(30.3)	(24.4)
Previous F experience, months	.001	.001	−.000	.002
	(4.3)	(3.4)	(1.6)	(7.1)
After F − After D	.076	.071	.046	.145
Before F − Before D	.058[b]	.034	.051	.077
After F − Before F	.051	.079	.039	.081
After D − Before D	.031	.042	.044	.013
DiD	.020	.037	−.005	.068
aR^2	.351	.353	.279	.315
Number of obs.	30,789,701	4,418,245	18,942,910	7,428,546

[a] Pooled OLS regression coefficients. The reference category comprises workers with no outside work experience in 2005–2009. All the reported coefficients are significant at the 0.01 level. t-statistics in parentheses. For the full results see the online appendix. The equations have been controlled for gender, age, age squared, months spent non-employed in 2003–2011, log health expenditures, receipt of disability payment, receipt of care allowances, months elapsed between previous and current jobs, tenure, Budapest dummy, log regional unemployment rate, 63 sector-year interactions and the log size, log capital-labour ratio and export shares of the current employer. Further dummies stand for missing information on tenure and time spent non-employed prior to entry to the current employer. *After F/After D*: previous employer is foreign/domestic. *Before F/Before D*: next employer is foreign/domestic. *F to D buyout*: the worker arrived at the firm via domestic buyout of a foreign firm.

[b] Rounding error.

eign firm (*After F*) earn more by 8.1 per cent than those, who are about to leave for a foreign firm (*Before F*). A similarly signed but smaller difference (1.3 per cent) exists between those who came from, or go to, another domestic enterprise (*After D − Before D*).

Using these estimates, we can approximate the return to foreign work experience as the double difference (After F−Before F)−(After D−Before D) equal to 6.8 per cent. Alternatively, we can start the calculus with the extra value of having foreign firm experience (After F−After D) and remove the differential in the pre-separation difference (Before F−Before D) to arrive at the same, 6.8 per cent estimate.

Similarly calculated double differences are significantly lower (3.7 per cent) for unskilled workers, virtually zero for the medium skilled (-0.5 per cent) and amount to a mere 2 per cent in the entire sample.

The results of the two models, aimed at measuring lagged wage effects, are similar. The first model identified a 4 per cent advantage on the part of the average skilled worker arriving from a foreign firm over a worker arriving from a domestic company. In this case worker quality was taken into account with the help of worker fixed effects. The quality-adjusted estimate in the second, 'overlapping cohorts' model is 6.8 per cent, not very far: recall that this model included more persons working for multinationals over a long period. The two models are also similar in suggesting much lower returns for less skilled workers.

The first model drew attention to downward bias in estimates for the average worker: it seems that foreign sector employees leaving for domestic companies are negatively selected unless they had to leave because of the closure or relocation of their employers. However, it seems that even after allowing for a selection effect, the lagged private returns must fall short of the contemporaneous ones, which indicated 25 and 30 per cent advantages on the part of skilled foreign sector workers (after allowing for observed and unobserved individual attributes) and over 10 per cent for all workers.

Results on Spillover Effects

The fixed effects panel equations summarized in Table 4.16 regress the log wages of domestic workers with no foreign experience on the share of workers with foreign experience within the worker's firm and skill group. The coefficient in the upper left corner (0.031), for instance, suggests that the wages of low skilled workers would grow by 3.1 per cent if the share of low skilled coworkers with foreign-firm experience grew from 0 to 100 per cent, holding observed and time fixed unobserved personal attributes constant. The equations have been estimated for 2005–2009, in order to leave time for accumulation of a stock of newcomers and disregard the crisis period in the latter case. We also estimated the equations separately for smaller (11–50) and larger (50+) firms, taking into consideration the higher risk of measurement error in small establishments. Table 4.16 presents the result for all and larger firms in 2005–2009. The full results are available in the online appendix.

Several regularities stand out. The own effects (how newcomer's share in group s affects incumbent wages in group s are stronger than the cross effects. The own effects increase with skill level of the ex-foreign coworkers. Consistent with the expectations, the estimates for all firms are lower than the estimates for larger firms, and not always significant in the employee-employer fixed effects models. The

Table 4.16 *Spillover effects estimated with worker and firm fixed effects, 2005-2009*

	Low[a]	Middling	High	Low	Middling	High
	All firms			Firms employing > 50 workers		
Estimates with worker fixed effects						
Low	0.031***	0.014**	0.012**	0.041**	0.038***	0.011
	(3.9)	(2.9)	(2.7)	(3.0)	(3.5)	(1.6)
Middling	0.024***	0.053***	0.035***	0.034***	0.090***	0.045***
	(7.3)	(15.3)	(14.6)	(7.4)	(12.3)	(11.2)
High	0.017***	0.035***	0.085***	0.034***	0.063***	0.112***
	(2.3)	(7.3)	(12.6)	(3.7)	(5.4)	(9.3)
Estimates with worker and firm fixed effects						
Low	0.018**	0.005	0.004	0.036**	0.036***	0.015*
	(2.7)	(1.3)	(1.1)	(2.9)	(3.7)	(2.3)
Middling	0.009**	0.019***	0.008***	0.020***	0.056***	0.018***
	(3.0)	(5.8)	(3.5)	(4.5)	(7.8)	(4.9)
High	0.016**	0.007*	0.007	0.037***	0.029**	0.053***
	(2.6)	(1.8)	(1.2)	(4.2)	(2.6)	(4.9)

[a] Significant at the 0.1 (*), 0.05 (**), 0.01 (***) level. t-values in brackets. Denoting the columns of the table with s' and the rows with s, the coefficients measure the effect of the share within skill group s' of workers with foreign work experience on the log wage of domestic workers within skill group s who had no foreign work experience between January 2003 and the month of observation. The equations were estimated row by row, with Stata's *xtreg* and *reg2hdfe* models and controlled for gender, age, age squared, Budapest dummy, log regional unemployment rate, receipt of care allowance and/or disability payment, log monthly expenditures on health services and medicine, job tenure, occupation, log firm size, log capital-labour ratio, the share of exports in sales revenues, and sector-year interactions.

parameters estimated with both worker and firm fixed effects are markedly lower, suggesting that the share of ex-MNE workers is higher in high-wage firms.

The within-firm (2FE) impacts seem rather weak in view of the relevant ranges of newcomer's shares. The share of ex-foreign workers is zero for 37 per cent of the person months observed in 2005–2009, and for the rest of the cases the mean share is 13.6 per cent with a standard deviation of 11.7 per cent. Therefore, the wage effects of moving from zero share to mean share, or, shifting the share by one standard deviation are small. Thus, for instance, taking the one of the highest parameter estimates in the bottom panel of Table 4.16 (last column in the last row), we find that a one standard deviation difference in the share of high skilled ex-foreign employees shifts the wages of high skilled incumbents up by only slightly more than a half per cent.

In evaluating the cross effects one should take into account the relevant range in the share of ex-MNE workers. While a jump from zero to 50 or 100 per cent in the share of ex-foreign workers within the medium skilled category is beyond the realm of reality, which renders the spillover effect to be weak, this can easily happen in the high skilled category. Domestic firms employing 50 workers have 7 high skilled workers on average. Hiring two managers or professionals with foreign sector experience can increase the ex-foreign share from zero to almost 30 per cent overnight, which implies a 0.6 per cent wage growth for the medium skilled incumbents and 1.6 per cent for their high skilled counterparts.

The estimated spillover effects might seem economically insignificant but, in fact, they are stronger rather than weaker than those found in the literature. Aitken and Harrison (1999) show that foreign ownership is associated with higher wages in Mexico and Venezuela but they do not find evidence of wage spillovers leading to higher wages for domestic firms. Keller and Yeaple (2009) detect significant worker-level wage spillovers only in high-skill-intensive industries in US manufacturing. The study of Poole (2013) based on Brazilian data – that is closest to ours in terms of approach, sample characteristics and industry coverage – estimates that at the average wage for a typical domestic worker, a 10 percentage point increase in the share of former MNE workers in his establishment increases wages by 23 $ per year, a little more than the price of one *Starbucks espresso* a month in Rio de Janeiro. The comparable estimate for skilled domestic workers in our sample is 0.53 per cent of the skilled average wage, the equivalent of about 69 $ a year, which can buy about four cups of *Starbucks espresso* a month in Budapest.

4.3.6 Conclusion

In all of the models presented in the study we have found contemporaneous and lagged residual wage gaps between foreign and domestic employees, unexplained by observed and unobserved attributes of the compared workers.

The raw 0.5-0.7 log point contemporaneous wage differential between domestic and foreign workers was halved once we have controlled for firm and worker attributes, and went down to a mere 13% for the entire sample, and 23% for skilled workers, when we have introduced worker fixed effects. The gap virtually disappeared if we have also controlled for firm fixed effects (1% for the entire sample and 2% for skilled workers), suggesting that the vast majority of the wage gap is caused by the non-random selection of ownerships, rather than by the actual changes in ownerships. The fixed effects approach has one limitation though: most firms and many employees have not changed owner or moved across sectors, and with that a large part of the sample has played no role in the identification of the wage gap. In order to utilize that part of the data as well, we have carried out an analysis on 'newly established' firms and have compared pre- and in-the-process wage differences. The 0.3 log point wage gap suggested that only part of the preexisting wage gap between foreign and domestic incumbents may be attributed to quality differences between workers and to their non-random assignments.

Lagged wage effects have been identified from comparing domestic workers coming from domestic firms versus foreign firms. The non-random separation of workers from firms have been controlled by focusing on collapsing firms. The 4% and 12% gaps for the average skilled and the high-skilled people, respectively, suggested that workers accumulate reasonable general skills at foreign employers, which they can transfer when moving across sectors.

We have also found evidence that the presence of ex-foreign coworkers raised the wages of incumbent domestic workers. One standard deviation increase in the fraction of high skilled ex-foreign workers was associated with 0.5% raise in the annual wages of high skilled co-workers. Even though this spillover effect seems economically insignificant, it is much higher what other studies usually find.

The results cast doubt on some censorious explanations of the foreign-domestic wage gap. The existence of a residual gap calls into question that the wage difference simply reflects the crowding of high productivity workers in foreign firms. The post-separation gains reprehend the hypothesis that the gap is purely firm-specific *i.e.*, vanishes without a trace when a worker leaves the foreign sector.

It seems that the residual gains rather stem from the matching of better-than-average workers and modern technologies, and the resulting gains are portable and transferable to some extent. High skilled MNE workers earn more than their domestic counterparts. Their earnings rise faster with tenure and/or experience. They lose a part of their wage advantage after leaving the foreign sector but, even so, they earn more than their domestic colleagues with no MNE experience. Their presence in a domestic firm exerts positive influence on the wages of incumbent co-workers. The gaps are larger and the spillover effects are stronger in the tertiary sector where FDI brought previously unknown technologies and corporate culture. The finding of much narrower gaps and weaker effects in the case of unskilled and medium skilled workers yields further support to a 'knowledge transfer' explanation of the results.

10.14754/CEU.2017.05

Appendix A – Supplementary Tables

Table A.1 *Descriptive statistics for the estimation sample of (4.2)*

Variable[a]	Mean	Var	Min	Max
Male	.583		0	1
Age	38.49	11.0	7	82
Low skilled	.147			
Middling	.602		0	1
High skilled	.251		0	1
Months of non-employment in 2003–2011	31.1	29.2	0	108
Log health expenditures/national average wage	−2.08	1.8	−12.08	7.14
Receives disability pension/payment	.013		0	1
Receives care benefit	.012		0	1
Log regional unemployment rate	−2.59	.390	−3.32	−1.73
Central Hungary including Budapest	.307		0	1
Tenure is unobserved	.386		0	1
Tenure (months)	12.87	18.6	0	108
Top manager	.037			
Other manager	.075		0	1
Professional	.073		0	1
Other white collar	.206		0	1
Skilled blue collar	.356		0	1
Assembler, machine operator	.169		0	1
Elementary occupation	.101		0	1
Unspecified occupation	.020		0	1
Agriculture	.044			
Manufacturing	.362		0	1
Construction	.066		0	1
Trade	.188		0	1
Finance	.035		0	1
Energy	.023		0	1
Services	.281		0	1
Foreign	.326			
Domestic	.616		0	1
Public sector	.046		0	1
Other, unspecified	.012		0	1
Firm size (log)	4.94	2.25	−.693	10.88
Fixed assets per worker (log)	7.78	1.86	−5.01	17.59
Share of exports in sales revenues	.212	.349	0	1

[a] Each variable covers 92,663,887 person months. The spells belong to workers employed at least once in a firm, the size of which exceeded the 10 workers limit at least once in 2003–2011. Note that other samples used in the paper have been drawn from this source file.

Table A.2 *Pooled OLS results for (4.2). Entire sample, 2003–2011*

	Coefficient[a]	t-value	Beta[b]
Majority owner			
Foreign	.2762	265.5	
Other	.0266	17.6	
Personal characteristics			
Male	.2236	271.9	
Age	.0350	139.3	.5179
Age squared	−.0003	−119.4	−.4589
Months spent non-employed in 2003–2011	−.0034	−263.2	−.1347
Receipt of disability payment	−.3057	−99.5	
Receipt of care allowance	−.3615	−101.2	
Health expenditures	−.0159	−113.3	−.0386
Highest occupational status in 2003–2011			
Middling	.0274	25.7	
High	.2260	137.1	
Job characteristics			
Tenure	.0006	36.2	.0167
Tenure is unobserved	.0997	112.5	.0647
Manager	−.0342	−3.9	
Professional	.0496	5.6	
Other white collar	−.2600	−30.0	
Skilled blue collar	−.5522	−63.7	
Assembler, machine operator	−.5589	−64.2	
Labourer in elementary occupation	−.7082	−81.5	
Unspecified	−1.2561	−132.9	
Regional unemployment rate (log)	−.0753	−59.4	−.0392
Budapest	.1171	107.8	
Firm characteristics			
Firm size (log)	.0511	247.0	.1535
Capital-labour ratio (log)	.0392	195.6	.0977
Exports/sales revenues	.1095	78.8	.0510
Constant	−1.7556	161.0	
R^2	0.4933		
$F(87, 1758833)$	21,949.87		
Number of observations	92,118,857		

[a] Dependent variable: log daily earnings. For the exact definition of the variables see the Data Appendix B. The data relate to 1,762,812 workers in 218,572 firms. The data comprise workers ever employed in a firm employing more than 10 workers. The coefficients of 63 sector-year dummies are not shown. The standard errors are adjusted for clustering by persons.

[b] Effect of a one standard deviation difference in the respective continuous variable.

10.14754/CEU.2017.05

Appendix B – Data Appendix

Starting Sample. 50 per cent random sample drawn from Social Security Numbers (SSN, Hungarian TAJ) valid on January 1, 2003. SSN holders aged 5–74 were retained. Data held by the Pension Directorate (ONYF), the Tax Office (NAV), the Health Insurance Fund (OEP), the Office of Education (OH), and the Public Employment Service (NMH) were merged and anonymized by the National Information Service (NISZ). The original data consisted of payment records with start and end dates, a type-of-payment code and amounts received by the person. Employers were identified by ONYF and their annual financial data were provided by NAV. The data was transformed to a fixed format monthly panel data set by the Databank of the Institute of Economics of the Hungarian Academy of Sciences. The costs of building the data base were financed by the Hungarian Academy of Sciences.

Estimation Sample. Workers employed with a labour contract at least once in a foreign or domestic private enterprise the maximum employment level of which exceeded the 10 workers limit at least once in 2003–2011. We removed workers and firms with less than two data points, zero wages and missing covariates. 98.5 per cent of the workers belong to a single connected group.[20]

Wage. The daily wage figure used in the paper was calculated as monthly earnings divided by the number of days covered by pension insurance ('working days' henceforth) in the given month. Multiple payments made by the same employer to the same person in a month were summed up. Working days belonging to these payments were also summed up but capped at 30 or 31 days. In the case of multiple job holders the wage figure belongs to the highest paying job. We normalized the wage figures by dividing them with the national average wage in the given month, as measured in the starting sample. Source: ONYF.

Skill Levels. Skill levels are inferred from the 'highest' occupational status held by the person in 2003–2011. The classification is based on one-digit occupational codes: 1 Top managers, 2 Other managers, 3 Professionals, 4 Other white collars, 5 Skilled blue collars, 6 Assemblers and machine operators, 7 Elementary occupations. Persons employed in occupations 1–3 at least once are classified as high

[20] "When a group of persons and firms is connected, the group contains all the workers who ever worked for any of the firms in the group and all the firms at which any of the workers were ever employed. In contrast, when a group of persons and firms is not connected to a second group, no firm in the first group has ever employed a person in the second group, nor has any person in the first group ever been employed by a firm in the second group. From an economic perspective, connected groups of workers and firms show the realized mobility network in the economy. From a statistical perspective, connected groups of workers and firms block-diagonalize the normal equations and permit the precise statement of identification restrictions on the person and firm effects." (Abowd et al., 2006a).

skilled. Persons never employed outside occupations 6 and 7 are classified as low skilled. Other persons are classified as medium skilled. Source: ONYF.

Total Time Spent Non-employed. The number of months out of employment in 2003–2011. Source: ONYF.

Disability Payment. Dummy variable, with 1 standing for any kind of transfer (pension or allowance) received on the basis of permanent disability (rokkantnyugdíj, rokkantsági járadék). Monthly data. Source: ONYF.

Care Allowances. Dummy variable, with 1 standing for any kind of benefits received by the observed person on the basis of raising children (tgyás, gyed, gyes, gyet) or taking care of relatives (ápolási segély). Monthly data. Sources: OEP, ONYF.

Health Expenditures. Expenditures and costs registered by the National Health Insurance Fund (OEP). The items include total amount paid for OEP-supported medicine and the costs of OEP-supported services/treatment provided by district doctors, specialists and hospitals. We normalized the nominal figures by dividing them with the national average wage in the given month, as measured in the starting sample. Annual data. Source: OEP.

Unemployment Rate. Seasonally adjusted ILO-OECD unemployment rate in the given month and NUTS-2 region. The worker's region is identified on the basis of his/her postal code in 2003. Source: the author's calculation using the Labor Force Survey.

Foreign-owned Firm, MNE. Dummy variable set to 1 for firms majority owned by one or more foreign owners. Ownership shares are measured as fractions of subscribed capital. Source: NAV.

Firm Size. Average number of employees. Annual data. Source: NAV.

Export Share. Ratio of export revenues to sales revenues. Annual data. Source: NAV.

Capital-labour Ratio. Net value of fixed assets per worker. Annual data. Source: NAV.

Industry, Sector. NACE 2 and NACE1, respectively. Source: NAV.

References

Abowd, J. M., Kramarz, F., and Margolis, D. N. 1999. High Wage Workers and High Wage Firms. *Econometrica*, **67**, 251–333.

Abowd, J. M., Kramarz, F., and Woodcock, S. 2006a. Econometric Analyses of Linked Employer-Employee Data. Pages 727–760 of: *The Econometrics of Panel Data*, vol. 46. Springer.

Abowd, J. M., Kramarz, F., and Roux, S. 2006b. Wages, Mobility and Firm Performance: Advantages and Insights from Using Matched Worker-Firm Data. *Economic Journal, Royal Economic Society*, **116**, F245–F285.

Acemoglu, D., and Pischke, J.-S. 1998. Beyond Becker: Training in Imperfect Labour Markets. *The Economic Journal*, **109**, F112–F142.

Aitken, B., Harrison, A., and Lipseyc, R. E. 1996. Wages and foreign ownership: A comparative study of Mexico, Venezuela, and the United States. *Journal of International Economics*, **40**, 345–371.

Aitken, B. J., and Harrison, A. E. 1999. Do Domestic Firms Benefit from Direct Foreign In-vestment? Evidence from Venezuela. *The American Economic Review*, **89**, 605–618.

Amemiya, T. 1978. A Note on a Random Coefficients Model. *International Economic Review*, **19**, 793–796.

Anderson, J. E., and Wincoop, v. E. 2003. Gravity with Gravitas: A Solution to the Border Puzzle. *American Economic Review*, 170–192.

Arellano, M., and Bond, S. 1991. Some Tests of Specification for Panel Data: Monte-Carlo Evidence and an Application to Employment Equations. *Review of Economic Studies*, **58**, 277–297.

Arellano, M., and Bonhomme, S. 2012. Identifying Distributional Characteristics in Random Coefficients Panel Data Models. *The Review of Economic Studies*, **79**, 987–1021.

Baier, S. L., and Bergstrand, J. H. 2007. Do Free Trade Agreements Actually Increase Members International Trade? *Journal of International Economics*, **71**, 72–95.

Balatoni, A., and Pitz, M. 2012. A m'Huködőtőke hatása a bruttó nemzeti jövedelemre Magyarországon. *Közgazdasági Szemle*, **LIX**, 1–30.

Balazsi, L., Matyas, L., and Wansbeek, T. 2015. The Estimation of Multi-dimensional Fixed Effects Panel Data Models. *forthcoming in Econometric Reviews*.

Balazsi, L., Bun, M., Chan, F., and Harris, M. 2017. Models with Endogenous Regressors. In: Matyas, L. (ed), *The Econometrics of Multi-dimensional Panels – Theory and Applications*. Springer Verlag (forthcoming).

Baldwin, R., and Taglioni, D. 2006. Gravity for Dummies and Dummies for the Gravity Equations. *NBER Working Paper 12516*.

Balestra, P., and Krishnakumar, J. 2008. Fixed Effects and Fixed Coefficients Models. Pages 23–48 of: Matyas, L., and Sevestre, P. (eds), *The Econometrics of Panel Data*, 3 edn. Springer Verlag.

Baltagi, B. H. 1987. On estimating from a more general time-series cum cross-section data structure. *The American Economist*, **31**, 69–71.

Baltagi, B. H. 2005. *Econometric Analysis of Panel Data*. 3 edn. John Wiley & Sons.

Baltagi, B. H. 2013. *Econometric Analysis of Panel Data*. 5 edn. Chichester, UK: John Wiley & Sons.

Baltagi, B. H., and Chang, Y.-J. 1994. Incomplete panels: A comparative study of alternative estimators for the unbalanced one-way error component regression model. *Journal of Econometrics*, **62**, 67–89.

Baltagi, B. H., Chang, Y.-J., and Li, Q. 1992. Monte Carlo results on several new and existing tests for the error component model. *Journal of Econometrics*, **54**, 95–120.

Baltagi, B. H., Song, S. H., and Jung, B. C. 2001. The unbalanced nested error component regression model. *Journal of Econometrics*, **101**, 357–381.

Baltagi, B. H., Song, S. H., and Jung, B. C. 2002. A comparative study of alternative estimators for the unbalanced two-way error component regression model. *Econometrics Journal*, **5**, 480–493.

Baltagi, B. H., Egger, P., and Pfaffermayr, M. 2003. A generalized Design for Bilateral Trade Flow Models. *Economics Letters*, **80**, 391–397.

Baltagi, B. H., Matyas, L., and Sevestre, P. 2008. Error Components Models. Pages 49–88 of: Matyas, L., and Sevetre, P. (eds), *The Econometrics of Panel Data*. Springer Verlag.

Barry, F., Görg, H., and Strobl, E. 2005. Foreign direct investment and wages in domestic firms in Ireland: Productivity spillovers versus labour-market crowding out. *International Journal of the Economics of Business*, **12**, 67–84.

Becker, G. S. 1962. Investment in Human Capital: A Theoretical Analysis. *Joumal of Political Economy*, **70**, 9–49.

Bryk, A. S., and Raudenbush, S. W. 1992. *Hierarchical Linear Models: Applications and Data Analysis Methods*. 1 edn. Newbury Park, CA: Sage Publications.

Bun, M. J. G., and Klaassen, F. J. G. M. 2007. The Euro Effect on Trade is not as Large as Commonly Thought. *Oxford Bulletin of Economics and Statistics*, **69**, 473–496.

Carneiro, A., Guimaraes, P., and Portugal, P. 2008. Real Wages and the Business Cycle: Accounting for Worker and Firm Heterogeneity. *Unpublished manuscript*.

Cheng, I.-H., and Wall, H. J. 2005. Controlling for Heterogeneity in Gravity Models of Trade and Integration. *Federal Reserve Bank of St. Louis Review*, **87**, 49–63.

Cornelissen, T. 2008. The Stata command felsdvreg to fit a linear model with two high-dimensional fixed effects. *The Stata Journal*, **8**, 170–189.

Czafit, B., and Köllő, J. 2015. Employment and wages before and after incarceration; evidence from Hungary. *IZA Journal of European Labor Studies 2015*, **4**.

Davis, P. 2002. Estimating multi-way error components models with unbalanced data structures using instrumental variables. *Journal of Econometrics*, **106**, 67–95.

De Mol, C., Gautier, E., Giannone, D., Mullainathan, S., Reichlin, L., van Dijk, H., and Wooldridge, J. 2017. Big Data in Economics: Evolution or Revolution? In: Matyas, L., Blundell, R., Cantillon, E., Chizzolini, B., Ivaldi, M., Leininger, W., Marimon, R., and Steen, F. (eds), *Economics without Borders*. Cambridge University Press.

Earle, J. S., and Telegdy, A. 2008. Ownership and Wages: New Evidence from Linked Employer-Employee Data in Hungary, 1986-2003. *Analysis of Firms and Employees – Qualitative and Quantitative Approaches, NBER*.

Earle, J. S., and Telegdy, A. 2013. Privatization, Employment, and Wages: Evidence from Hungary in Comparative Perspective. *The Hungarian Labor Market: Review and Analysis, Institute of Economics HAS*.

Earle, J. S., Brown, J. D., and Telegdy, A. 2006. The Productivity Effects of Privatization: Longitudinal Estimates from Hungary, Romania, Russia and Ukraine. *Journal of Political Economy*, **114**.

Earle, J. S., Brown, J. D., and Telegdy, A. 2010. Employment and Wage Effects of Privatization: Evidence from Hungary, Romania, Russia and Ukraine. *Economic Journal*, **120**.

Ebbes, P., Bockenholt, U., and Wedel, M. 2004. Regressor and Random-effects Dependencies in Multilevel Models. *Statistica Neerlandica*, **58(2)**, 161–178.

Egger, P., and Pfaffermayr, M. 2003. The Proper Econometric Specification of the Gravity Equation: 3-Way Model with Bilateral Interaction Effects. *Empirical Economics*, **28**, 571–580.

Elek, P., Köllő, J., Reizer, B., and Szabó, P. A. 2012. Detecting wage under-reporting using a double hurdle model. Chap. 4, pages 135–166 of: Lehmann, H., and Tatsiramos, K. (eds), *Research in Labor Economics Vol 34: Informal Employment in Emerging and Transition Economies*. Emerald Group Publishing Ltd.

Fuller, W. A., and Battese, G. E. 1973. Transformations for Estimation of Linear Models with Nested-Error Structure. *Journal of the American Statistical Association*, **68**, 626–632.

Ghosh, S. K. 1976. Estimating from a More General Time-Series cum Cross-Section Data Structure. *The American Economist*, **20**, 15–21.

Goldstein, H. 1995. *Multilevel Statistical Models, 2nd Edition*. John Wiley & Sons, Ltd, Chichester, UK.

Görg, H., and Strobl, E. 2005. Spillovers from Foreign Firms through Worker Mobility: An Empirical Investigation. *The Scandinavian Journal of Economics*, **107**, 693–709.

Görg, H., Strobl, E., and Walsh, F. 2007. Why Do Foreign-Owned Firms Pay More? The Role of On-the-Job Training. *Review of World Economics*, **143**.

Grogger, J. 1995. The effect of arrests on the employment and earnings of young men. *Quarterly Journal of Economics*, 51–71.

Grogger, J., and Hanson, G. H. 2011. Income maximization and the selection and sorting of international migrants. *Journal of Development Economics*, **95**, 42–57.

Grunfeld, Y. 1958. The Determinants of Corporate Investment. *Unpublished PhD thesis. University of Chicago*.

Guimaraes, P., and Portugal, P. 2009. A Simple Feasible Alternative Procedure to Estimate Models with High-Dimensional Fixed Effects. *Stata Journal*, **10**, 628–649.

Harris, M., Konya, L., and Matyas, L. 2002. Modelling the Impact of Environmental Regulations on Bilateral Trade Flows: OECD, 1990-1996. *The World Economy*, **25**(3), 387–405.

Harris, M. N., Matyas, L., and Sevestre, P. 2008. Dynamic Models for Short Panels. In: Matyas, and Sevestre (eds), *The Econometrics of Panel Data, 3rd edition*. Springer Verlag.

Hartwig, R. E. 1976. Block generalized inverses. *Archive for Rational Mechanics and Analysis*, **61**, 197–251.

Hausman, J., and Taylor, W. 1981. Panel data and unobservable individual effects. *Econometrica*, **49**, 1377–1398.

Hausman, J. A. 1978. Specification Tests in Econometrics. *Econometrica*, **46**, 1251–1271.

Hornok, C. 2011. Gravity or Dummies? The Limits of Identification in Gravity Estimations. *CeFiG Working Papers*, **No. 15**.

Hsiao, C. 2003. *Analysis of Panel Data*. 2 edn. Econometric Society Monographs, Cambridge University Press.

Hsiao, C., and Pesaran, M. H. 2008. Random Coefficient Models. In: Matyas, L., and Sevestre, P. (eds), *The Econometrics of Panel Data*. Springer Verlag.

Huber, J. D., and Stanig, P. 2011. Church-state Separation and Redistribution. *Journal of Public Economics*, **95**, 828–836.

Hubler, O. 2006. Multilevel and Nonlinear Panel Data Models. *Allgemeines Statistisches Archiv*, **90**, 121–136.

Keller, W., and Yeaple, S. R. 2009. Multinational Enterprises, International Trade, and Technology Diffusion: A Firm-level Analysis of the Productivity Effects of Foreign Competition in the United States. *Review of Economics and Statistics*, **91**, 821–831.

Köllő, J. 2011. Employment, unemployment and wages in the first year of the crisis. Pages 43–68 of: Fazekas, K., and Molnár, G. (eds), *The Hungarian labour market, 2011*. Review and analysis Budapest, Institute of Economics, HAS – National Employment Foundation.

Konings, J., and Vanormelingen, S. 2015. The Impact of Training on Productivity and Wages: Firm-Level Evidence. *The Review of Economics and Statistics*, **97**, 485–497.

Konya, L., Matyas, L., and Harris, M. 2011. GATT/WTO Membership Does Promote International Trade After All – Some New Empirical Evidence. *Working paper, Munich Personal RePEc Archive, 34978, online at http://mpra.uni-muenchen.de/34978/*.

Konya, L., Matyas, L., and Harris, M. N. 2013. Modelling Trade Relationships: The Random Effects vs. Fixed Effects Approach. *CEU Economics Working Papers*.

Krishnakumar, J., Marquez, M. A., and Balazsi, L. 2017. Random Coefficient Models. In: Matyas, L. (ed), *The Econometrics of Multi-dimensional Panels – Theory and Applications*. Springer Verlag (forthcoming).

Kuh, E. 1963. *Capital Stock Growth: A Micro-Econometric Approach*. Amsterdam: North-Holland.

Laird, N. M., and Ware, J. H. 1982. Random-Effects Models for Longitudinal Data. *Biometrics*, **38(4)**, 963–974.

LaLonde, R. J., and Cho, R. 2008. The impact of incarceration in state prison on the employment prospects of women. *Journal of Quantitative Criminology*, **24**, 243–267.

Lazear, E. P. 2009. Firm-Specific Human Capital: A Skill-Weights Approach. *Journal of Political Economy*, **117**, 914–940.

Leeuw, J. D., and Kreft, I. 1986. Random Coefficient Models for Multilevel Analysis. *Journal of Educational and Behavioral Statistics*, **11(1)**, 57–85.

Lipsey, R. E., and Sjöholm, F. 2004. FDI and wage spillovers in Indonesian manufacturing. *Review of World Economics*, **140**, 321–332.

Loewenstein, M. A., and Spletzer, J. R. 1999. General and specific training – Evidence and implications. *The Journal of Human Resources*, **34**, 710–733.

Luke, D. A. 2004. *Multilevel Modeling*. Thousand Oaks, CA: SAGE Publications.

Maddala, G. 1971. The use of variance components models in pooling cross section and time series data. *Econometrica*, **39**, 341–358.

Magee, C. S. P. 2003. Endogenous Preferential Trade Agreements: An Empirical Analysis. *Contributions to Economic Analysis & Policy*, **2**.

Matyas, L. 1997. Proper Econometric Specification of the Gravity Model. *The World Economy*, **20**, 363–369.

Matyas, L. 1998. The Gravity Model: Some Econometric Considerations. *The World Economy*, **21**, 397–401.

Matyas, L., Konya, L., and Harris, M. 1997. Modelling Export Activity in a Multicountry Economic Area: The APEC Case. *Monash University, Department of Econometrics, WorkingPaper 1*.

Miao, J.-M. 1991. General Expressions for the Moore-Penrose Penrose Inverse of a 2×2 Block Matrix. *Linera Algebra and its Applications*, **151**, 1–15.

Nerlove, M., Sevestre, P., and Balestra, P. 2008. Introduction. In: Matyas, L., and Sevestre, P. (eds), *The Econometrics of Panel Data*, 3 edn. Springer Verlag.

Nickell, S. 1981. Biases in Dynamic Models with Fixed Effects. *Econometrica*, **49**, 1417–1426.

Pettit, B., and Lyons, C. J. 2009. Incarceration and the Legitimate Labor Market: Examining Age-Graded Effects on Employment and Wages. *Law & Society Review*, **43**, 725–756.

10.14754/CEU.2017.05

Poole, J. P. 2013. Knowledge Transfers from Multinational to Domestic Firms: Evidence from Worker Mobility. *Review of Economic Statistics*, **95**, 393–406.

Pyhnen, P. 1963. A Tentative Model for the Volume of Trade between Countries. *Weltwirtschaftliches Archiv*, **90**, 93–99.

Rao, R., and Kleffe, J. 1980. Handbook of Statistics 1. Pages 1–40 of: Krishnaiah, P. (ed), *Estimation of Variance Components*. North-Holland Publisher.

Rose, A. K. 2004. Do We Really Know That the WTO Increases Trade? *American Economic Review*, **94**, 98–114.

Scott, M. A., Simonoff, J. S., and Marx, B. D. 2013. *The SAGE handbook of multilevel modeling*. London: SAGE Publications Ltd.

Searle, S. R. 1971. *Linear Models*. John Wiley & Sons.

Smarzynska Javorcik, B. 2004. Does Foreign Direct Investment Increase the Productivity of Domestic Firms? *American Economic Review*, **94**, 605–627.

Subramanian, A., and Wei, S.-J. 2007. The WTO Promotes Trade, Strongly but Unevenly. *Journal of International Economics*, **72**, 151–175.

Swamy, P. A. V. B. 1970. Efficient Inference in a Random Coefficient Regression Model. *Econometrica*, **38**, 311–323.

Tinbergen, J. 1962. *Shaping the World Economy – Suggestions for an International Economic Policy*. The Twentieth Century Fund, New York.

Tomz, M., Goldstein, J. L., and Rivers, D. 2007. Do We Really Know that the WTO Increases Trade? Comment. *American Economic Review*, **97**, 2005–2018.

UNECE. 2001. *Economic growth and foreign direct investment in the transition economies*. Tech. rept. Economic Survey of Europe, United Nations Economic Commission for Europe, Geneva.

Varian, H. R. 2014. Big Data: New Tricks for Econometrics. *Journal of Economic Perspectives*, **28**, 3–28.

Vera-Cruz, A. O., and Dutrenit, G. 2005. Spillovers from MNCs through worker mobility and technological and managerial capabilities of SMEs in Mexico. *Innovation: Management, Policy & Practice*, **7**.

Wansbeek, T. 1991. Singular value decomposition of design matrices in balanced ANOVA models. *Statistics & Probability Letters*, **11**, 33–36.

Wansbeek, T., and Kapteyn, A. 1982. A Simple Way to Obtain the Spectral Decomposition of Variance Components Models for Balanced Data. *Communications in Statistics – Theory and Methods*, **11**, 2105–2112.

Wansbeek, T., and Kapteyn, A. 1989. Estimation of the Error-Components Model with Incomplete Panels. *Journal of Econometrics*, **41**, 341–361.

Wu, W.-C., and Lin, S.-J. 2002. Housing Demand with Random Group Effects. *International Real Estate Review, Asian Real Estate Society*, **5**, 133–145.

Wyhowski, D. J. 1994. Estimation of a Panel Data Model in the Presence of Correlation between Regressors and a Two-Way Error Component. *Econometric Theory*, **10**, 130–139.