Advanced Studies in Theoretical and Applied Econometrics 50

Laszlo Matyas *Editor*

The Econometrics of Multi-dimensional Panels

Theory and Applications

Springer

Advanced Studies in Theoretical and Applied Econometrics

Volume 50

More information about this series at http://www.springer.com/series/5667

Laszlo Matyas
Editor

The Econometrics of Multi-dimensional Panels

Theory and Applications

Springer

Editor
Laszlo Matyas
Department of Economics
Central European University
Budapest, Hungary

ISSN 1570-5811 ISSN 2214-7977 (electronic)
Advanced Studies in Theoretical and Applied Econometrics
ISBN 978-3-319-60782-5 ISBN 978-3-319-60783-2 (eBook)
DOI 10.1007/978-3-319-60783-2

Library of Congress Control Number: 2017946838

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Foreword

This volume of essays on the theory and applications of multi-dimensional panel data is a welcome addition to the burgeoning literature on panel data econometrics, which has been fueled by the growing availability of large data sets. This is the first volume to deal with higher-dimensional panels in a unified and comprehensive way. Multi-dimensional panel data do not neatly fit the paradigm of a survey of the same individuals, periodically, over an interval of time: for example, of business expectations, plans and realizations, quarterly, for the years, say, 2000–2010, or monthly total expenditures of a panel of households over a period of time. In contrast, we find in this volume five essays reviewing and extending, including many new results, the extensive literatures on trade or on migration flows between pairs of nations; on prices of identical products and of close substitutes in a cross-section of supermarkets of different sizes daily, over a year; on the impact on productivity of various industries, in different OECD countries, of R&D investments and of investments in information and communications technology (ICT); on temporal and spatial variation in house prices, residential mobility and location choice.

The volume begins with ten chapters dealing primarily with methodological issues, including many new results. There are not only chapters on the standard linear fixed-effects and random-effects models, albeit in a multi-dimensional context, but also chapters on various estimation and testing issues, which arise because of the increasing sources of unobserved heterogeneity as the dimensionality of the model increases. More specifically, for example: potential correlation of some of the regressors with components of unobserved heterogeneity; dynamic and/or nonlinear relationships; random coefficients; semi-parametric models; discrete response models; and quantile models. The particular issues associated with spatial data and models are dealt with in a full chapter.

What exactly are multi-dimensional panel data? In the case of familiar two-dimensional panels, each observation is typically a vector of values of a dependent variable and one or more independent variables, and comes with two labels attached, one is frequently time and the other an individual person, business or nation. When the panel is multi-dimensional, each observation comes with many labels, for example, time, individual employee, firm, and industry. The labels may themselves have labels, thus allowing an arrangement of observations in a hierarchy. In principle, an observation could consist of values of multiple endogenous variables and multiple

exogenous or predetermined variables, labeled with at least time and one other label. For example, consider a macro-econometric model for each nation in one of several regions of the world. One could imagine such a collection of models as having been estimated from a multi-dimensional panel of nations, over various time periods, and in several regions of the world.

All of the problems and issues which arise for two-dimensional panels also exist, sometimes in more complicated form, for multi-dimensional panels. Incomplete panels may be incomplete because not all the vectors of observations contain the same number of components. Or some labels may be missing all together. Such issues are dealt with throughout the book.

But perhaps the most important difference between multi-dimensional and two-dimensional panels is the many dimensions in which asymptotic expansions of the distribution of estimates may be considered. For the standard case of a cross-section of individuals observed at discrete points in time, one can consider a sequence of estimates of a parameter based on successively larger samples of individuals or of time points. Sometimes, as in the case of fixed-effects panel data models, we consider asymptotics when both the number of time points and the number of individuals are increasing, but at different rates relative to one another. In the case of multi-dimensional models, however, there are an embarrassing number of possibilities.

Should we be concerned? In his often cited text on asymptotic statistics, van der Vaart writes, "In fact, strictly speaking, most asymptotic results that are currently available are logically useless. This is because most asymptotic results are limit results, rather than approximations consisting of an approximating formula plus an accurate error bound. . . . Because it may be theoretically very hard to ascertain that approximation errors are small, one often takes recourse to simulation studies to judge the accuracy of a certain approximation."[1] In the case of multi-dimensional models, comprehensive simulation results are difficult to present and hard to interpret. While some simulations are presented in the book, how to deal informatively with multi-dimensional data represents an important direction for future research.

Formulation of multi-dimensional panel models presents serious issues. In addition to those above, because very large panel data sets are now available, nonparametric and nonlinear models are now feasible, opening up additional difficult questions of which model to choose.

This book is much more than a comprehensive state-of-the-art introduction to multi-dimensional panel modeling and estimation. It will be required reading for all those who wish to go further.

Maryland, April, 2017

Marc Nerlove
Distinguished University Professor, Emeritus
University of Maryland

[1] van der Vaart, A. W., *Asymptotic Statistics*, Cambridge: Cambridge University Press, 1998, p. 3.

Preface

The last couple of decades has seen the use of panel data become a standard in many areas of economic analysis as large numbers of such data sets have been compiled and made public. The available model formulations have become more complex, the estimation and hypothesis testing methods more sophisticated. The interaction between economics and econometrics has resulted in a huge publication output, deepening and widening immensely our knowledge and understanding of both.

Traditional panel data sets, by nature, are two-dimensional. Lately, however, as part of the big data revolution, there has been a rapid emergence of three, four and even higher dimensional panel data sets. These arose by extending or dividing the observed individuals (like household and/or firm area data, etc.), by matching different cross sectional data (e.g., matched employer-employee, doctor-patient, etc. data), by origin destination flow type data (e.g., trade, migration, investment, etc.), by cross-sectional data grouped according to some discrete variables (e.g., new college graduates' job market offerings and wage rates for different occupations, industries, regions etc.), by multi-dimensional interactive data (e.g., social networking data with a large number of social groups and group members), and so on.

Oddly, applications have rushed ahead of theory in this field. This book is aimed at filling this widening gap. The first ten chapters of the volume provide the econometric foundations to deal with these new high-dimensional panel data sets. They not only synthesize our current knowledge, but mostly present new research results. Chapters 11-15 provide in-depth insights into some relevant empirical applications in this area. These chapters are a mixture of surveys and new results, always focusing on the econometric problems and feasible solutions. They deepen our understanding on how econometrics can be applied to different kinds of data and economics problems.

Higher dimensional panel data sets have some common characteristics. For example:

- These data sets are in most cases incomplete and unbalanced, often by design, with a high percentage of missing or zero observations.

- The number of observations can be very large in some dimensions, in cases in the tens of millions, while in other dimensions quite limited.
- Emerging from economic theory, frequently, there is a right hand side index deficit in the models. For example, often, when the dependent variable is of three dimensions, the explanatory variables are only of two, etc.
- The number of theoretically available models, possible effect specifications, taking into account the interaction effects as well, can be dauntingly large.

This book aims at dealing with all the above issues, providing a helping hand to practitioners, but also pointing to unsolved problems, encouraging further research. The editor has tried to standardize to some extent the notation and language of the volume in order to present a coherent book. However, each chapter is able to stand on its own as a reference in its own area.

Acknowledgments

First of all, I would like to thank the authors of the volume for the excellent chapters produced in a timely manner. I appreciate their patience in dealing with the requests and frequent "harassments" of the editor. I would also like to thank the Central European University for their institutional support. I am also grateful for the hospitality of Curtin University, Perth and Monash University, Melbourne, Australia during the editing process. Editorial assistance by Robin Bellers and Eszter Timar added much to the quality of this volume and is kindly acknowledged. Administrative assistance by Melinda Molnar and personal assistance by Theo and Rosetta Batsakis is also much appreciated.

Laszlo Matyas
Budapest, April, 2017

Contents

List of Contributors

Laszlo Matyas
Central European University, 1051 Nador utca 9, Budapest, Hungary, e-mail: matyas@ceu.edu

Monika Avila Márquez
Institute of Economics and Econometrics, Geneva School of Economics and Management (GSEM), University of Geneva, Geneva, Switzerland, e-mail: monika.avila@unige.ch

Laszlo Balazsi
Central European University, 1051 Nador utca 9, Budapest, Hungary, e-mail: lac.balazsi@gmail.com

Badi H. Baltagi
Syracuse University Department of Economics and Center for Policy Research, 426 Eggers Hall, Syracuse, New York 13244–1020, USA, e-mail: bbaltagi@maxwell.syr.edu

Nicoletta Berardi
Banque de France, 31 rue Croix des Petits Champs - 75001 Paris, France, e-mail: nicoletta.berardi@banque-france.fr

Georges Bresson
Université Paris II, 12 place du Panthéon, 75231 Paris cedex 05, France, e-mail: bresson-georges@orange.fr

Maurice J.G. Bun
University of Amsterdam, Postbus 15867 1001 NJ, Amsterdam, the Netherlands, e-mail: M.J.G.Bun@uva.nl

Felix Chan
Curtin University, School of Economics and Finance, Bentley Campus, Perth, Western Australia, e-mail: Felix.Chan@cbs.curtin.edu.au

Peter H. Egger
ETH Zurich, CEPR, and CESifo; Leonhardstr. 21, 8092 Zurich, Switzerland, e-mail: egger@kof.ethz.ch

Katharina Erhardt
ETH Zurich, Leonhardstr. 21, 8092 Zurich, Switzerland, e-mail: erhardt@kof.ethz.ch

Antonio F. Galvao
Department of Economics, University of Iowa, W284 Pappajohn Business Building, 21 E. Market Street, Iowa City, IA 52242, USA, e-mail: antonio-galvao@uiowa.edu

Julie Le Gallo
Université Bourgogne Franche-Comté, CESAER, AgroSup Dijon, INRA, 26, Boulevard Petitjean, 21000 Dijon, France, e-mail: julie.le-gallo@agrosupdijon.fr

Mark N. Harris
Curtin University, School of Economics and Finance, Bentley Campus, Perth, Western Australia, e-mail: Mark.Harris@curtin.edu.au

Balazs Kertesz
Central European University, 1051 Nador utca 9, Budapest, Hungary, e-mail: kertesz_balazs@phd.ceu.edu

Qi Li
International School of Economics and Management, Capital University of Economics and Business, Beijing, 100070, PR China, Department of Economics, Texas A&M University, College Station, TX, 77843, USA, e-mail: qi-li@tamu.edu

Wei Lin
International School of Economics and Management, Capital University of Economics and Business, Beijing, 100070, PR China, e-mail: weilin_econ@163.com

Jimmy Lopez
Université de Bourgogne (LEDI-CNRS) and Banque de France, e-mail: jimmy.lopez@u-bourgogne.fr

George Kapetanios
School of Management & Business, King's College London, Franklin Wilkins Building, London, United Kingdom, SE19NH, e-mail: george.kapetanios@kcl.ac.uk

Jaya Krishnakumar
Institute of Economics and Econometrics, Geneva School of Economics and Management (GSEM), University of Geneva, Geneva, Switzerland, e-mail: Jaya.Krishnakumar@unige.ch

Jacques Mairesse
CREST-ENSAE, Maastricht University, and NBER, e-mail: Jacques.Mairesse@ensae.fr

Camilla Mastromarco
Dipartimento di Scienze dell'Economia, University of Salento, Ecotekne, Lecce, Italy, 73100, e-mail: camilla.mastromarco@unisalento.it

Gabriel V. Montes-Rojas
Departament d'Economia Aplicada, Universitat Autònoma de Barcelona, Edifici B Campus de la UAB, Cerdanyola del Vallès, 08193, Barcelona, Spain, e-mail: gabriel.montes@uab.cat

Alain Pirotte
CRED, Université Paris II Panthéon-Assas and TEPP (CNRS), 12 Place du Panthéon, 75231 Paris Cedex 05, France, e-mail: alain.pirotte@u-paris2.fr

Daria Pus
The University of Texas at Austin, 2225 Speedway BRB 1.116, C3100 Austin, Texas 78712, USA, e-mail: daria.pus@gmail.com

Raul Ramos
AQR-IREA and IZA, Department of Econometrics, Statistics and Applied Economics, Universitat de Barcelona, Av. Diagonal 690, 08034 Barcelona, Spain, e-mail: rramos@ub.edu

Laura Serlenga
Dipartimento di Scienze Economiche e Metodi Matematici, University of Bari, Bari, Italy, 70100, e-mail: laura.serlenga@uniba.it

Patrick Sevestre
Greqam-AMSE, Université Aix-Marseille, Chateau Lafarge - route des Milles - 13290 Les Milles, France, e-mail: patrick.sevestre@univ-amu.fr

Yongcheol Shin
Department of Economics and Related Studies, University of York, York, United Kingdom, YO105DD, e-mail: yongcheol.shin@york.ac.uk

Yiguo Sun
Department of Economics and Finance, University of Guelph, Guelph, ON N1G2W1, Canada, e-mail: yisun@uoguelph.ca

Jonathan Thébault
French Embassy in Brazil, e-mail: thebault.jonathan@gmail.com

Tom Wansbeek
University of Groningen, Nettelbosje 2, 9747 AE Groningen, the Netherlands, e-mail: t.j.wansbeek@rug.nl

Chapter 1
Fixed Effects Models

Laszlo Balazsi, Laszlo Matyas, and Tom Wansbeek

Abstract In recent years the massive emergence of multi-dimensional panels has led to an increasing demand for more sophisticated model formulations with respect to the well known two-dimensional ones to address properly the additional heterogeneity in the data. This chapter deals with the most relevant three-dimensional fixed effects model specifications and derives appropriate Least Squares Dummy Variables and Within estimators for them. The main results of the chapter are also generalized for unbalanced panels, cross-sectional dependence in the error terms, and higher dimensional data. Some thoughts on models with varying slope coefficients are also presented.

1.1 Introduction

Model formulations in which individual and/or time heterogeneity factors are considered fixed parameters, rather than random variables (see Chap. 2), 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. This approach can then be extended to a varying coefficients framework, where heterogeneity is not picked up by the constant term, but rather by the slope coefficients.

The vast majority of the empirical studies conducted on multi-dimensional panels involve fixed effects models of some form. Chapters 11–15 of this volume visit

Laszlo Balazsi
Central European University, Budapest, Hungary

Laszlo Matyas
Central European University, Budapest, Hungary

Tom Wansbeek
University of Groningen, Groningen, the Netherlands

L. Matyas (ed.), *The Econometrics of Multi-dimensional Panels*, Advanced Studies in Theoretical and Applied Econometrics 50, DOI 10.1007/978-3-319-60783-2_1

Table 1.1 Examples of empirical studies for multi-dimensional fixed effects models, as appearing in the empirical chapters of this volume

Study	Topic	Indices (*i-j-t*)	Sample Size	Fixed Effects	Balanced
Chapter 11 – Trade					
Glick and Rose (2002)	Currency Union	origin country - destination country - year	220 000	γ_{ij}	No
Head et al. (2010)	Colonial Trade Linkages		618 000	γ_{ij}	No
Baier and Bergstrand (2002)	Endogeneity of Trade Flows		1 400	$\alpha_i + \gamma_j$	No
Baier and Bergstrand (2009)	Trade Agreements		19 000	$\alpha_i + \gamma_j$	No
Egger and Pfaffermayr (2011)	Path Dependence		57 000	$\alpha_i + \gamma_j$	No
Egger et al. (2011)	Endogenous Trade Agreements		16 000	$\alpha_i + \gamma_j$	No
Matyas (1997)	Gravity Model Spec.		1 700	$\alpha_i + \gamma_j + \lambda_t$	No
Egger (2000)	Gravity Model Spec.		2 500	$\alpha_i + \gamma_j + \lambda_t$	No
Rose and van Wincoop (2001)	Currency Union		31 000	$\alpha_i + \gamma_j + \lambda_t$	No
Magee (2003)	Preferential Trade Agreements		90 000	$\alpha_i + \gamma_j + \lambda_t$	No
Egger (2001)	Exports and Outward FDI		1 000	$\gamma_{ij} + \lambda_t$	No
Bun and Klaassen (2002)	Importance of Dynamics		10 000	$\gamma_{ij} + \lambda_t$	No
Cheng and Wall (2005)	Trade Integration		3 200	$\gamma_{ij} + \lambda_t$	No
Shin and Serlenga (2007)	Intra-EU Trade		3 800	$\gamma_{ij} + \lambda_t$	No
Martin et al. (2008)	Military Conflicts and Trade		225 000	$\gamma_{ij} + \lambda_t$	No
Egger and Pfaffermayr (2003)	Gravity Model Spec.		2 000	$\alpha_i + \gamma_j + \lambda_t + \gamma^*_{ij}$	No
Baldwin and Taglioni (2006)	Gravity Model Spec.		2 500	$\alpha_i + \gamma_j + \lambda_t; \gamma_{ij} + \lambda_t$	No
Romalis (2007)	NAFTA's, CUSFTA's Impact	country - commodity - year	1 116 000	$\gamma_{ij} + \lambda_{it}$	No
Olivero and Yotov (2012)	Trade Agreements	origin country - destination country - year	5 500	$\alpha_{it} + \alpha^*_{jt}$	No
Baltagi et al. (2003)	Gravity Model Spec.		10 000	$\gamma_{ij} + \alpha_{it} + \alpha^*_{jt}$	No
Baier and Bergstrand (2007)	Trade Agreements		36 000	$\alpha_{it} + \alpha^*_{jt}; \gamma_{ij} + \alpha_{it} + \alpha^*_{jt}$	No
Nuroglu and Kunst (2014)	Factors Explaining Trade		150 000	$\gamma_{ij} + \alpha_{it} + \alpha^*_{jt}$	No
Bergstrand et al. (2015)	Border Effects		24 000	$\gamma_{ij} + \alpha_{it} + \alpha^*_{jt}$	No
Chapter 12 – Housing and Prices					
Fu et al. (2015)	Housing Tenure Choices	household - prefecture - type	2 500 000	λ_t	No
Syed et al. (2008)	House Prices Indices	house - region - quarter	418 000	α_{jt}	No
Gayer et al. (2000)	Risks from Superfund Sites	house - city - year	17 000	$\gamma_j + \lambda_t$	No
Turnbull and van der Vlist (2015)	Uninformed House Buyers	house - block - year	115 000	$\alpha_i + \gamma_j + \lambda_t$	No
Bayer et al. (2016)	Demand for Houses	household - neighbourhood - time	1 000 000	$\alpha_i + \alpha^*_{jt}$	No
Baltagi et al. (2015)	Neighbor's Prices	year - 'arrondissement' - quartier - block - flat	157 000	$\alpha_{ta} + \gamma_{taq} + \lambda_{taqi}$	No

some of the major fields in which multi-dimensional panels are used.[1] Tables 1.1-1.2 collect the fixed effects specifications relied upon in these empirical chapters. Just by itself, Matyas's (1997) seminal paper has a tremendous number of citations, which can dramatically be expanded by considering other popular fixed effects formulations. A representative selection of such publications, in addition to the ones in Tables 1.1-1.2, is presented in Table 1.3. While these collections are far from being comprehensive in terms of topics or even the kind of observations the data sets

[1] Further, see Koren and Hornok (2017) for a review on recent advances in trade and comprehensive three-dimensional data sets.

may comprise, it gives a decent picture of how fruitfully fixed effects models can be applied.

Table 1.2 Examples of empirical studies for multi-dimensional fixed effects models, as appearing in the empirical chapters of this volume, cont.

Study	Topic	Indices (*i*-*j*-*t*)	Sample Size	Fixed Effects	Balanced
Chapter 13 – Migration					
Perkins and Neumayer (2014)	International Student Flows	origin country - destination country - year	85 000	λ_t	No
Belot and Ederveen (2012)	Cultural Barriers		2 700	$\alpha_i + \gamma_j$	No
Czaika and Hobolth (2016)	Asylum and Visa Policies		9 000	$\alpha_i + \gamma_j$	No
Beine and Parsons (2015)	Climatic Factors		62 000	$\alpha_i + \alpha^*_{jt}$	No
Bertoli and Fernández-Huertas Moraga (2013)	Multilateral Resistance	origin country - quarter - year	2 700	$\gamma_j + \alpha_{it}$	No
Bertoli et al. (2016)	European Crisis	origin country - month - year	2 200	$\gamma_{ij} + \lambda_t$	Yes
Echevarria and Gardeazabal (2016)	Refugee Migration	origin country - destination country - year	700 000	$\gamma_{ij} + \lambda_t$	No
Poot et al. (2016)	Intranational Migration	origin region - destination region - year	1 200	$\gamma_{ij} + \lambda_t$	No
Eilat and Einav (2004)	International Tourism	origin country - destination country - year	5 500	$\gamma_{ij} + \alpha_i + \gamma^*_j + \lambda_t$	No
Abbott and Silles (2016)	International Student Flows		2 200	$\alpha_i + \gamma_j + \lambda_t$	No
Adserà and Pytliková (2015)	Language		95 000	$\alpha_i + \gamma_j + \lambda_t$	No
Figueiredo et al. (2016)	Migration and Regional Trade Agreements		63 000	$\alpha_{it} + \alpha^*_{jt}$; $\alpha_i + \gamma_j + \lambda_t$	No
Llull (2016)	Understanding International Migration		7 300	$\alpha_i + \gamma_j + \lambda_t$; $\alpha_{it} + \gamma_j$; $\alpha_i + \alpha^*_{jt}$; $\gamma_{ij} + \lambda_t$	No
Ortega and Peri (2013)	Immigration Policies		40 000	$\alpha_i + \gamma_j + \lambda_t + \alpha^*_{it} + \gamma^*_{ij}$	No
Barthel and Neumayer (2015)	Asylum Migration		29 000	$\gamma_{ij} + \alpha_{it} + \alpha^*_{jt}$	No
Chapter 14 – Country-Industry Productivity					
	R&D and Productivity	country - industry - time	4 000	$\alpha_i + \gamma_j + \lambda_t$; $\gamma_{ij} + \alpha_{it} + \alpha^*_{jt}$	No
	Non-Manufacturing Regulations		4 000	$\alpha_i + \gamma_j + \lambda_t$; $\gamma_{ij} + \alpha_{it} + \alpha^*_{jt}$	No
Chapter 15 – Consumer Price Heterogeneity					
	Consumer Price Dispersion	product - store - wholesaler - week	37 130 000	$\gamma_{ij} + \alpha_{ist}$	No
Gorodnichenko et al. (2014)	Price Setting in Online Markets	good - seller - time	17 700	$\alpha_i + \gamma_j$	No
Dubois and Perrone (2015)	Price Dispersion	product - store - year	445 000	$\alpha_i + \lambda_t$; $\gamma_j + \lambda_t$	No
Gorodnichenko and Talavera (2017)	Price Setting in Online Markets	good - country - time	21 700	$\gamma_j + \lambda_t$	No
Biscourp et al. (2013)	Retails Regulations	product - type - fascia	42 000	$\alpha_i + \gamma_j + \lambda_t$	No
Borenstein and Rose (1994)	US Airline Industry	airport - airport - carrier	1 000	γ_{ij}(FE) + λ_t(RE)	No
Gerardi and Shapiro (2009)	Price Dispersion	carrier - route - time	27 000	$\gamma_{ij} + \lambda_t$	No

A few regularities stand out from Tables 1.1-1.2 and 1.3. First, most models are not too sophisticated from the point of view of the kind of fixed effects used (column 5); in fact they can usually be traced back to two-dimensional (2D) models by re-

Table 1.3 Further examples of empirical studies for multi-dimensional fixed effects models, grouped by model complexity

Study	Topic	Indices (*i*-*j*-*t*)	Sample Size	Fixed Effects	Balanced
Berthelemy (2006)	Donor's Assistance	donor country - recipient country - year	36 000	α_i	No
Thompson and Pendell (2016)	Poultry Trade	country pairs - poultry product - year	2 200	α_i	Yes
Hirsch (2013)	Gender Wage Gap	employee - employer - year	1 200 000	γ_j; γ_{ij}	No
Hur et al. (2010)	Trade Agreements	origin country - destination country - year	56 000	γ_{ij}	No
Smith and Yetman (2007)	Multivariate Forecasts	forecaster - forecast horizon - quarter	15 000	γ_{ij}; $\alpha_i + \gamma_j$; α_i; γ_j	No
Horrace and Schnier (2010)	Mobile Product Technologies	vessel - spatial location - year	1 500	α_{it}	Yes
Parsley and Wei (1999)	Border Effect	traded goods - cities - quarter	228 000	$\alpha_i + \gamma_j$	Yes
Haller and Cotterill (1996)	Share-Price Measures	brand - market - quarter	3 500	$\alpha_i + \gamma_j$	No
Crozet et al. (2016)	Domestic Trade Regulations	origin country - destination country - firm - time	115 000	$\alpha_s + \lambda_t$; α_{st}	No
Bussiere et al. (2005)	Trade Integration	origin country - destination country - year	50 000	$\gamma_{ij} + \lambda_t$	No
Bellak et al. (2008)	Labour Costs and FDI Flows		400	$\alpha_i + \gamma_j + \lambda_t$	No
Fourie and Santana-Gallego (2011)	Tourist Flows		91 000	$\alpha_i + \gamma_j + \lambda_t$	No
Harris et al. (2000)	Environmental Regulations		3 800	$\alpha_i + \gamma_j + \lambda_t$	No
Heyman et al. (2007)	Foreign Ownership Wage Premium	employee - employer - year	1 600 000	$\alpha_i + \gamma_j + \lambda_t$	No
Parsley (2003)	Exchange Rate Pass Through	import goods - importing country - year	1 300	$\alpha_i + \gamma_j + \lambda_t$	No
Melitz and Toubal (2012)	Linguistic Factors of Trade	origin country - destination country - year	209 000	$\alpha_{it} + \alpha^*_{jt}$	No
Aghion et al. (2008)	Indian Trade Liberalization	industry - state - year	18 000	$\gamma_{ij} + \alpha_{it} + \alpha^*_{jt}$	Yes

placing pairs of indices with a single index. Second, as the estimation of models with a complex fixed effects structure might be problematic on large data sets, more complex models are usually applied on data with moderate sample sizes, spanning from a few thousands to "only" tens of thousand of observations. More importantly, each index also tends to be short: a few dozen countries, a handful of product categories, or annual periods of ten-twenty years, etc. Third, almost all data sets collected are unbalanced, some closer to a fully complete panel (flow-type data with a few countries), some more heavily (employer-employee matched data). From these, it seems clear that studies typically rely on simpler models, not particularly exploiting the possible interaction effects and the higher-dimensionality of the data, especially as larger data sets and more complicated models together come at the price of heavy computational burdens. This chapter provides solutions to most of these issues by proposing estimation techniques for various "truly" three-dimensional (3D) fixed effects models, feasible even under unbalanced data sets of extreme sizes. The models considered are exclusively static. Dynamic models are visited in Chap. 4.

In Sect. 1.2 we introduce the most relevant models in a three-dimensional panel data setup. Section 1.3 deals with the Least Squares estimation of these models, while Sect. 1.4 analyses the behaviour of this estimator for incomplete/unbalanced

data. Section 1.5 studies the properties of the so-called Within estimator. Section 1.6 extends the original models to account for eventual heteroscedasticity and cross-correlation. Section 1.7 generalizes the models presented to four and higher dimensional data sets, while Sect. 1.8 deals with some varying coefficients specifications. Sections 1.2, 1.5 and 1.7 rely heavily on Balazsi et al. (2015).

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}, $i = 1,\dots,N_1$, $j = 1,\dots,N_2$, and $t = 1,\dots,T$, and the observations have the same ordering: index i goes the slowest, then j, and finally t the fastest,[2] such as

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

We assume in general that the index sets $i \in \{1,\dots,N_1\}$ and $j \in \{1,\dots,N_2\}$ 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 two-dimensional 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 only consider the empirically most meaningful ones.

Throughout the chapter, we follow standard ANOVA notation, that is I and J denote the identity matrix, and the square matrix of ones respectively, with the size indicated in the subscript, $\bar{J}$ denotes the normalized J (each element is divided by the number in the subscript), 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 in the place of that index. When discussing unbalanced data, a plus sign in 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 expressed 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

[2] Note that the N_1, N_2 notation does not mean, by itself, that the data is unbalanced.

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 for more, for example, Baltagi, 2013, Balestra and Krishnakumar, 2008). The specification of this model is

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

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), $\pi = (\alpha' \ \gamma' \ \lambda')'$ with $\alpha' = (\alpha_1, \ldots, \alpha_{N_1})$, $\gamma' = (\gamma_1, \ldots, \gamma_{N_2})$ and $\lambda' = (\lambda_1, \ldots, \lambda_T)$), and ε_{ijt} are the i.i.d. $(0, \sigma^2_\varepsilon)$ idiosyncratic disturbance terms. We also assume that the x_{ijt} covariates and the disturbance terms are uncorrelated (this assumption is relaxed in Chap. 3). Equation (1.2) has been the model applied in several studies in trade, migration, as well as in labour economics (see e.g., Egger, 2000; Harris et al., 2000; Rose and van Wincoop, 2001; Magee, 2003; Parsley, 2003; Heyman et al., 2007; Bellak et al., 2008; Fourie and Santana-Gallego, 2011; Ortega and Peri, 2013; Adserà and Pytliková, 2015; Turnbull and van der Vlist, 2015).

A model has been proposed by Egger and Pfaffermayr (2003), popular in the trade literature, forecasting and labour economics (see e.g., Glick and Rose, 2002; Smith and Yetman, 2007; Head et al., 2010; Hur et al., 2010; Hirsch, 2013), which takes into account bilateral interaction effects. The model specification is

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

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 (see also Egger, 2001; Bun and Klaassen, 2002; Eilat and Einav, 2004; Bussiere et al., 2005; Romalis, 2007; Shin and Serlenga, 2007; Martin et al., 2008; Bertoli et al., 2016 or Bertoli and Fernández-Huertas Moraga, 2013; Beine and Parsons, 2015) is

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

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

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

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

where we allow the individual effect to vary over time (see e.g., Syed et al., 2008; Horrace and Schnier, 2010; Crozet et al., 2016). It is reasonable to present the sym-

metric version of this model (with α_{it} fixed effects); however, as it has exactly the same properties, we consider the two models together.[3]

A variation of this model is

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

(Olivero and Yotov, 2012; Baier and Bergstrand, 2007), whereas the model that encompasses all the above effects is

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

typically used in explaining trade flows (see e.g., Baltagi et al., 2003; Baier and Bergstrand, 2007; Aghion et al., 2008; Melitz and Toubal, 2012; Nuroglu and Kunst, 2014; Bergstrand et al., 2015). Each model with its specific D matrix from formulation (1.1) is summarized in Table 1.4.

Table 1.4 Model specific D matrices

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

It is interesting to see that our collection of models is exhaustive, apart from the permutation of indices. This is summarized in Table (1.5). Out of the five distinct models two are technically for 2D data (rows two and three), and only the rest are truly three-dimensional.

Table 1.5 The exhaustive grouping of indices

Indices	Model
i, j, t	(1.2)
(ij)	(1.3)/(1.5)
$(ij), t$	(1.4)
$(it), (jt)$	(1.6)
$(ij), (it), (jt)$	(1.7)

[3] Strictly speaking, models (1.3) and (1.5) are also the same from a mathematical point of view. Nevertheless, as it is usually the case that i and j are entities and t is time, it makes sense from an economics point of view to distinguish (ij) from (jt), but to take (it) and (jt) under one hat.

1.3 Least Squares Estimation of the Models

If the matrix $(X,\ D)$ has full column rank,[4] 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

$$\hat{\beta} = (X'M_DX)^{-1}X'M_Dy \tag{1.8}$$

$$\hat{\pi} = (D'D)^{-1}D'(y-X\hat{\beta})\,, \tag{1.9}$$

where the idempotent and symmetric matrix $M_D = I - D(D'D)^{-1}D'$ is the so-called *Within projector*. This follows directly from

$$D'D\hat{\pi}+D'X\hat{\beta} = D'y \tag{1.10}$$

$$X'D\hat{\pi}+X'X\hat{\beta} = X'y\,. \tag{1.11}$$

The first equation gives (1.9). $D\hat{\pi} = (I-M_D)(y-X\hat{\beta})$, a rearrangement, which in turn can be substituted back to (1.11) gives (1.8)

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 filling out the concrete form of D and M_D, specific to that given model. Table 1.6 captures these different projection matrices for all models discussed. Furthermore, 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 $\hat{\beta}$.

The estimation of the fixed effects parameters is captured by (1.9) if D has full column rank. This, however, only holds for models of one fixed effect, that is, for (1.3) and (1.5). Estimation of β is not affected since it is based on the projection matrices M_D. The estimators for the fixed effects read as

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

for model (1.3), and

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

[4] Since $\iota_{N_1N_2T}$ is spanned by all given specifications for D, there is intercept in X.

Table 1.6 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)	$I-(I_{N_1}\otimes \bar{J}_{N_2}\otimes I_T)-(\bar{J}_{N_1}\otimes I_{N_2T})-(I_{N_1N_2}\otimes \bar{J}_T)$ $+(\bar{J}_{N_1N_2}\otimes I_T)+(\bar{J}_{N_1}\otimes I_{N_2}\otimes \bar{J}_T)+(I_{N_1}\otimes \bar{J}_{N_2T})$ $-\bar{J}_{N_1N_2T}$	$(N_1-1)(N_2-1)(T-1)-K$

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 intuitive, as for example for model (1.2) the sum of the α_i, the sum of the γ_j and the sum of the λ_t parameters all give the general constant. To make them identified, we have to impose some restrictions on the fixed effects parameters. The two most widely used are either to normalize the fixed effects, i.e., to set their average to zero, or to leave out the parameters belonging to the last (or first) individual or time period. We will follow this latter approach. Staying with the example of 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)-(1.9) works perfectly fine 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 of these estimators if we express them separately for each fixed effects parameter. 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),

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

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, the 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^* read in a 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

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

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

$$\hat{\tilde{\varepsilon}}^2_{ijt} = (\tilde{y}_{ijt} - \tilde{x}'_{ijt}\hat{\beta})^2 \tag{1.12}$$

is the transformed residual square, and $(\mathrm{rank}(M_D) - K)$ is collected in the last column of Table 1.6 for all models.

As multi-dimensional panel data are usually large in one or more directions, it is important to also have a closer look at the asymptotic properties. 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 may 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, $N_1 \mathrm{Var}(\hat{\beta})$ is finite in the limit, as

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

where Q_{XMX} is assumed to be a finite, positive semi-definite matrix, further, using the central limit theorem,

$$\sqrt{N_1}(\hat{\beta} - \beta) \xrightarrow{d} N\left(0, \sigma_\varepsilon^2 Q_{XMX}^{-1}\right) .$$

The estimator of a fixed effect is consistent only if at least one of the indexes with which the fixed effect does not vary 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, to test for $\beta_k = 0$, the t-statistic is given in the usual form

$$\hat{\beta}_k / \sqrt{\widehat{\mathrm{Var}}(\hat{\beta}_k)}$$

where $\widehat{\mathrm{Var}}(\hat{\beta}_k)$ is the k-th diagonal element of $\widehat{\mathrm{Var}}(\hat{\beta})$. The degrees of freedom has to be adjusted accordingly, for each model, as Table 1.6 shows. In principle, it is possible, but not typical to also test for the significance of some fixed effects parameters with the usual t-tests, 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 $\alpha_1 = \alpha_2 = \ldots = \alpha_{N_1}$. Using model (1.2) for illustration, the statistic for the F-test (assuming normality) is obtained as in

$$F = \frac{(R_{\mathrm{U}}^2 - R_{\mathrm{R}}^2)/(N_1 - 1)}{(1 - R_{\mathrm{U}}^2)/(N_1 N_2 T - N_1 - N_2 - T + 1 - K)}$$

where $\mathrm{R}_{\mathrm{U}}^2$ is the R^2 of the *unrestricted model* (that is the full model (1.2)), while $\mathrm{R}_{\mathrm{R}}^2$ 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 $(N_1N_2T - N_1 - N_2 - T + 1 - K)$. This statistic then has an F-distribution with $(N_1 - 1, N_1N_2T - N_1 - N_2 - T + 1 - K)$ 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, 2013, 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 (ij) 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. We are saving, however, the no self-flow issue to Sect. 1.5, and consider the general form of incompleteness in this section.

In the case of incomplete data, the models can still be cast as in (1.1), but now D cannot be represented nicely by kronecker products, as done in Table 1.4. 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)-(1.9). 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 1.6. As $\tilde{D}$ has a known form only if we know exactly which observations are missing, $M_{\tilde{D}} = I - \tilde{D}(\tilde{D}'\tilde{D})^{-}\tilde{D}'$ cannot be analytically defined element-wise 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 cannot usually avoid large computational burdens when carrying out (1.8)-(1.9) in case of incompleteness (as opposed to no computational burden when the data is complete).[5] 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 $\tilde{D}$ usually has 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). This is why we have to turn to

[5] 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 is a promising research topic, however, beyond the scope of the text.

generalized inverses for the former, but it is enough to work with "simple" inverses for the latter dummy matrices.

Incompleteness is less of an issue in the case of 2D models, where T is usually small, and N_1 is large (so we only have to invert a $(T \times T)$ matrix (see Wansbeek and Kapteyn, 1989), but is generally present in the case of 3D data, where typically along with N_1, N_2 is also large. In practice, to alleviate the issue with the size of the individual indexes, the best approach seems to be to turn to iterative solutions to find the Least Squares estimators. One of the most widely used is based on the work of Guimaraes and Portugal (2010) and Carneiro et al. (2012). Let us show the procedure on model (1.2), the rest is a direct analogy. Model (1.2) in matrix form reads as

$$y = X\beta + \tilde{D}_1\alpha + \tilde{D}_2\gamma + \tilde{D}_3\lambda + \varepsilon\,, \tag{1.13}$$

where tildes indicate two things. First, the data is possibly incomplete: from the original $D_1 = (I_{N_1} \otimes \iota_{N_2 T})$, $D_2 = (\iota_{N_1} \otimes I_{N_2} \otimes \iota_T)$, and $D_3 = (\iota_{N_1 T} \otimes I_T)$, the rows matching with the missing observations are deleted. Second, to make all model parameters estimable, we leave out α_{N_1} and γ_{N_2} from the model. The normal equations from (1.13) are then

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

which suggests the Gauss-Seidel, or as often called, the "zigzag" algorithm. This means that 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.[6]

1.5 The Within Estimator

1.5.1 The Equivalence of the LSDV and the Within Estimator

As seen, LSDV estimates all parameters of the fixed effects models in one step. There is, however, another appealing way to approach the estimation problem. The

[6] The STATA program command `reg2hdfe` implements these results and is found in the STATA Documentation. The code is designed to tackle two fixed effects, however, it can be improved to treat three, or even more fixed effects at the same time.

idea is that by 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, note that as M_D is idempotent, (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

$$\tilde{y}_{ijt} = y_{ijt} - \bar{y}_{i..} - \bar{y}_{.j.} - \bar{y}_{..t} + 2\bar{y}_{...} \ . \tag{1.14}$$

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

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

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

$$\tilde{y}_{ijt} = y_{ijt} - \bar{y}_{ij.} \ . \tag{1.16}$$

For model (1.4), the optimal Within transformation is in fact (1.15). Note that model (1.2) is a special case of model (1.4) (with the restriction $\gamma_{ij} = \alpha_i + \gamma_j$), so while transformation (1.15) 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

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

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

$$\tilde{y}_{ijt} = y_{ijt} - \bar{y}_{.jt} - \bar{y}_{i.t} + \bar{y}_{..t} \ , \tag{1.18}$$

and

$$\tilde{y}_{ijt} = y_{ijt} - \bar{y}_{ij.} - \bar{y}_{.jt} - \bar{y}_{i.t} + \bar{y}_{..t} + \bar{y}_{.j.} + \bar{y}_{i..} - \bar{y}_{...} \ , \tag{1.19}$$

respectively.

It can be seen that the Within transformation works perfectly in wiping out the fixed effects. However, frequently in empirical applications, some explanatory vari-

ables, (i.e., some elements of the vector x'_{ijt}) do not span the whole (ijt) data space, that is, they have 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 the 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 (1.14). Clearly, parameters associated with such regressors then cannot be estimated. This is most visible for model (1.7), as in this case all regressors fixed at least in one dimension are excluded from the model automatically after the Within transformation (1.19).

1.5.2 Incomplete Panels and the Within Estimator

We have briefly covered incompleteness in Sect. 1.3 already, but the Within estimators and the underlying transformations, open a new way to deal with it.

1.5.2.1 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 Sect. 1.2, the scalar transformations introduced 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:

$$\begin{aligned}\tilde{y}_{ijt} = y_{ijt} - \tfrac{N-1}{N(N-2)T}(y_{i++} + y_{+j+}) - \tfrac{1}{N(N-2)T}(y_{j++} + y_{+i+}) \\ - \tfrac{1}{N(N-1)}y_{++t} + \tfrac{2}{N(N-2)T}y_{+++}\end{aligned} \tag{1.20}$$

for model (1.2), and

$$\tilde{y}_{ijt} = y_{ijt} - \frac{1}{T}y_{ij+} \tag{1.21}$$

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

$$\tilde{y}_{ijt} = y_{ijt} - \frac{1}{T}y_{ij+} - \frac{1}{N(N-1)}y_{++t} + \frac{1}{TN(N-1)}y_{+++}\,, \tag{1.22}$$

and

$$\tilde{y}_{ijt} = y_{ijt} - \frac{1}{N-1}y_{+jt}\,, \tag{1.23}$$

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

$$\begin{aligned}\tilde{y}_{ijt} = y_{ijt} &- \tfrac{N-1}{N(N-2)}\left(y_{i+t}+y_{+jt}\right) - \tfrac{1}{N(N-2)}\left(y_{+it}+y_{j+t}\right) \\ &+ \tfrac{1}{(N-1)(N-2)} y_{++t}\,,\end{aligned} \tag{1.24}$$

and

$$\begin{aligned}\tilde{y}_{ijt} = y_{ijt} &- \tfrac{N-3}{N(N-2)}\left(y_{i+t}+y_{+jt}\right) + \tfrac{N-3}{N(N-2)T}\left(y_{i++}+y_{+j+}\right) - \tfrac{1}{T} y_{ij+} \\ &+ \tfrac{1}{N(N-2)}\left(y_{+it}+y_{j+t}\right) - \tfrac{1}{N(N-2)T}\left(y_{+i+}+y_{j++}\right) \\ &+ \tfrac{N^2-6N+4}{N^2(N-1)(N-2)}\left(y_{++t}-y_{+++}\right)\,,\end{aligned} \tag{1.25}$$

respectively. So overall, the no self-flow data problem can be overcome by using an appropriate Within transformation. Optimality of the estimators is preserved, as the transformations are derived from the Frisch–Waugh–Lovell theorem.

1.5.2.2 General Incompleteness

Next we 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 $t \in T_{ij}$, 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 in fact they can be represented as 2D models with 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.

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 $(t = 1 \ldots T)$ 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

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

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:

$$M_{D_a}^{(1)} = I_R - D_1^a (D_1^{a'} D_1^a)^{-1} D_1^{a'} ,$$
$$M_{D_a}^{(2)} = M_{D_a}^{(1)} - M_{D_a}^{(1)} D_2^a (D_2^{a'} M_{D_a}^{(1)} D_2^a)^{-} D_2^{a'} M_{D_a}^{(1)} ,$$

and finally

$$M_{D_a} = M_{D_a}^{(3)} = M_{D_a}^{(2)} - M_{D_a}^{(2)} D_3^a (D_3^{a'} M_{D_a}^{(2)} D_3^a)^{-} D_3^{a'} M_{D_a}^{(2)} . \tag{1.26}$$

It is easy to see that in fact $M_{D_a} D_a = 0$ projects out all three dummy matrices. Note that the first inverse calculation of this repetitive process is always easy, as $(D_1^{a'} D_1^a)$ is diagonal. It is recommended then to order the fixed effects in such a way that the largest of the three comes at the beginning. With this in mind, we only have to calculate two inverses instead of three, $(D_2^{a'} M_{D_a}^{(1)} D_2^a)^{-}$, and $(D_3^{a'} M_{D_a}^{(2)} D_3^a)^{-}$, with respective sizes $(N_2 \times N_2)$ and $(T \times T)$. 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 $(I_{N_1 N_2} \otimes I_{N_1 N_2})$ 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,

$$D_1^b = (W_1', W_2', \dots, W_T')' \quad \text{of size} \quad (R \times N_1 N_2) ,$$
$$D_2^b = D_3^a \quad \text{of size} \quad (R \times T) .$$

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

$$M_{D_b} = M_{D_b}^{(2)} = M_{D_b}^{(1)} - M_{D_b}^{(1)} D_2^b (D_2^{b'} M_{D_b}^{(1)} D_2^b)^{-} D_2^{b'} M_{D_b}^{(1)} . \tag{1.27}$$

As $(D_1^{b'} D_1^b)$ is diagonal again, 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 (1.27) is in fact also valid for model (1.2).

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

$$D_1^c = \text{diag}\{U_1, U_2, \dots, U_T\} \quad \text{of size} \quad (R \times N_1 T) \quad \text{and}$$
$$D_2^c = \text{diag}\{V_1, V_2, \dots, V_T\} \quad \text{of size} \quad (R \times N_2 T) ,$$

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

$$M_{D_c} = M_{D_c}^{(2)} = M_{D_c}^{(1)} - M_{D_c}^{(1)} D_2^c (D_2^{c'} M_{D_c}^{(1)} D_2^c)^{-} D_2^{c'} M_{D_c}^{(1)} . \tag{1.28}$$

Note that for M_{D_c}, we have to invert an order $\min\{N_1T, N_2T\}$ matrix, 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:

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

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

$$\begin{aligned} M_{D_d}^{(1)} &= I_R - D_1^d (D_1^{d\prime} D_1^d)^{-1} D_1^{d\prime} \text{ and} \\ M_{D_d}^{(2)} &= M_{D_d}^{(1)} - M_{D_d}^{(1)} D_2^{d\prime} (D_2^{d\prime} M_{D_d}^{(1)} D_2^d)^{-} D_2^{d\prime} M_{D_d}^{(1)}\,, \end{aligned}$$

the appropriate transformation for model (1.7) is now

$$M_{D_d} = M_{D_d}^{(3)} = M_{D_d}^{(2)} - M_{D_d}^{(2)} D_3^{d\prime} (D_3^{d\prime} M_{D_d}^{(2)} D_3^d)^{-} D_3^{d\prime} M_{D_d}^{(2)}\,. \tag{1.29}$$

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.[7] As model (1.6) is covered by model (1.7), projection (1.29) also eliminates the fixed effects from that model. Moreover, as all three-way fixed effects models are in fact nested into model (1.7), it is intuitive that transformation (1.29) clears the fixed effects in all model formulations. Using (1.7) is not always advantageous though, as (i) the transformation involves the inversion of potentially large matrices (of order N_1T, and N_2T) and (ii) the underlying estimator is no longer BLUE. 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. Table 1.7 collects the orders of the largest matrices to be inverted for all model specifications considered. In the table, we assume that $N_1 >> T$ and $N_2 >> T$ holds, and that N_1 and N_2 are of similar magnitudes.

It is worth noting that transformations (1.26), (1.27), (1.28), and (1.29) 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 (1.29), as all subsequent transformations can be dealt with in the same way. Let

[7] A STATA program code for transformation (1.29) with a user-friendly detailed explanation is available at http://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.

Table 1.7 Orders of the largest matrix to be inverted

Model	Order
(1.2)	$\min\{N_1, N_2\}$
(1.3)	K
(1.4)	T
(1.5)	K
(1.6)	$\min\{N_1T, N_2T\}$
(1.7)	$\max\{N_1T, N_2T\}$

$$\phi = C^{-}\bar{D}'y \quad \text{and} \quad \omega = \tilde{C}^{-}(M_{D_d}^{(2)}D_3^d)'y \quad \xi = C^{-}\bar{D}'D_3^d\omega\,,$$

where

$$C = \left(D_2^d\right)'\bar{D}\,, \quad \bar{D} = \left(I_R - D_1^d(D_1^{d'}D_1^d)^{-1}D_1^{d'}\right)D_2^d\,, \text{ and } \quad \tilde{C} = D_3^{d'}M_{D_d}^{(2)}D_3^d\,.$$

Now the scalar representation of transformation (1.29) is

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

where a_{ij} and $\tilde{a}_{ij}$ are the column vectors corresponding to observations (i,j) from matrices $A = D_2^{d'}D_1^d$ and $\tilde{A} = D_3^{d'}D_1^d$, respectively; ϕ_{it} is the (i,t)-th element of the $(N_1T\times 1)$ column vector ϕ; ω_{jt} is the (j,t)-th element of the $(N_2T\times 1)$ column vector ω; and finally, ξ_{it} is the element corresponding to the (i,t)-th observation from the $(N_1T\times 1)$ column vector, ξ. 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 $(N_1T\times N_1T)$, and $(N_2T\times N_2T)$ size matrices. Simplifications related to this can dramatically reduce CPU and storage requirements. This topic, however, is well beyond the scope of this chapter.

1.6 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 in 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.

1.6.1 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 Sect. 1.2 takes the form

$$\mathrm{E}(\varepsilon\varepsilon') = \Omega = (\Upsilon \otimes I_T) + \rho_1(I_{N_1} \otimes J_{N_2} \otimes J_T) + \rho_2(J_{N_1} \otimes I_{N_2} \otimes J_T), \tag{1.30}$$

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 collecting X and D in Z and β and π in δ, gives

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

The GLS estimator then reads as

$$\hat{\delta} = \left(Z'\Omega^{-1}Z\right)^{-1} Z'\Omega^{-1}y\,. \tag{1.31}$$

As much as (1.31) is simple theoretically, it is as forbidding practically: to carry out the estimation, 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 be inverted 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 this case, Ω is simplified to

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

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

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

with

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

As now we have the exact form of Ω^{-1}, estimation (1.31) 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

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

which is proportional to the scalar representation of $\Omega^{-\frac{1}{2}} y$, then perform an OLS on the transformed model. To obtain an estimator of β, we invoke the Frisch–Waugh–Lovell theorem again, and premultiply the transformed variables with the projector

$$M_{\Omega^{-\frac{1}{2}} D} = I - \Omega^{-\frac{1}{2}} D \big(D' \Omega^{-1} D\big)^{-} D' \Omega^{-\frac{1}{2}} \,,$$

which are then estimated with OLS. 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 as long as the effects are symmetrical in i and j, as, quite intuitively, the Within transformation for those models eliminates the cross-correlations from the disturbance terms along with the fixed effects.

1.6.2 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 them. However, in the case of some models, this significantly 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 another 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 Sect. 1.3 suggests, whenever the D dummy coefficient matrix has no full column rank, the composite fixed effects parameters, π cannot be identified (and of course, estimated). However, this is not the case for $D\pi$, which is given by

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

following (1.10). The LSDV residuals are

$$\hat{\varepsilon} = y - X\hat{\beta} - D\hat{\pi} = (I - (I - M_D))\,(y - X\hat{\beta}) = M_D(y - X\hat{\beta}) = \tilde{y} - \tilde{X}\hat{\beta} \qquad (1.32)$$

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:

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

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

$$\begin{aligned} \hat{\sigma}_{ij}^2 &= \tfrac{1}{T}\textstyle\sum_t \hat{\varepsilon}_{ijt}^2 \\ \hat{\rho}_2 &= \tfrac{1}{N_1(N_1-1)}\left(\tfrac{1}{N_2 T}\textstyle\sum_{jt}\left(\sum_i \hat{\varepsilon}_{ijt}\right)^2 - \sum_i \hat{\sigma}_{ij}^2\right) \\ \hat{\rho}_1 &= \tfrac{1}{N_2(N_2-1)}\left(\tfrac{1}{N_1 T}\textstyle\sum_{it}\left(\sum_j \hat{\varepsilon}_{ijt}\right)^2 - \sum_j \hat{\sigma}_{ij}^2\right) \,. \end{aligned} \qquad (1.33)$$

Equation (1.33) 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 (1.31) 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

$$\begin{aligned} \hat{\sigma}_\varepsilon^2 &= \tfrac{1}{N_1 N_2 T}\textstyle\sum_{ijt} \hat{\varepsilon}_{ijt}^2 \\ \hat{\rho}_2 &= \tfrac{1}{N_1 - 1}\left(\tfrac{1}{N_1 N_2 T}\textstyle\sum_{jt}\left(\sum_i \hat{\varepsilon}_{ijt}\right)^2 - \hat{\sigma}_\varepsilon^2\right) \\ \hat{\rho}_1 &= \tfrac{1}{N_2 - 1}\left(\tfrac{1}{N_1 N_2 T}\textstyle\sum_{it}\left(\sum_j \hat{\varepsilon}_{ijt}\right)^2 - \hat{\sigma}_\varepsilon^2\right) \,, \end{aligned} \qquad (1.34)$$

and T-asymptotics is no longer necessary ($N_1 \to \infty$ or $N_2 \to \infty$ is enough) to make 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 no longer be represented with Kronecker products, and so the linear transformations presented 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 (as we did in Sect. 1.5.2) and rows and columns from Ω that 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 (1.33) (or (1.34)) 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 should we proceed when the data is small in the time dimension? Let us consider that disturbances are heteroscedastic only, and the cross correlations are set to null ($\rho_1 = \rho_2 = 0$). This special case can be estimated in two ways. First, we can transform the model according to the optimal Within transformation as before, then carry out an FGLS with the heteroscedastic covariance matrix

$$\Omega_h = \operatorname{diag}\left\{ \sigma_{11}^2 I_{|T_{11}|},\ \sigma_{12}^2 I_{|T_{12}|},\ \ldots,\ \sigma_{nm}^2 I_{|T_{N_1 N_2}|} \right\},$$

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, and use heteroscedasticity robust White covariance matrix to estimate $\operatorname{Var}(\hat{\beta})$. Then we get

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

where "$\sim$" indicates that the variables are transformed. Notice again that only the data X has to be transformed, but conveniently not Ω_h, due to the idempotent nature of the projection matrix. This conjecture can be easily proven, by showing that the equivalence

$$\left[(Z'Z)^{-1}Z'\Omega_h Z(Z'Z)^{-1}\right]_{1,1} = (X'M_D X)^{-1}X'M_D\Omega_h M_D X(X'M_D X)^{-1} \qquad (1.35)$$

in fact holds with $Z = (X, D)$. Applying the partitioned inverse formula for block matrices gives the upper block of the (2×1) block matrix $(Z'Z)^{-1}Z'$ as

$$\begin{aligned}\left[(Z'Z)^{-1}Z'\right]_1 &= \left((X'M_DX)^{-1}, -(X'M_DX)^{-1}X'D(D'D)^{-1}\right)\cdot(X,D)' \\ &= (X'M_DX)^{-1}X' - (X'M_DX)^{-1}X'D(D'D)^{-1}D' \\ &= (X'M_DX)^{-1}X'M_D\,,\end{aligned}$$

which is used directly to construct the right hand side of (1.35).

1.7 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 somehow restrict 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 Sects. 1.4 and 1.5. This is carried out in a way that gives indications on how to proceed beyond four dimensions.

1.7.1 Different Forms of Heterogeneity

The dependent variable is now observed along four indexes, 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 be time dependent

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

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

$$y_{ijst} = x'_{ijst}\beta + \gamma^0_{ijs} + \gamma^1_{ijt} + \gamma^2_{jst} + \gamma^3_{ist} + \varepsilon_{ijst}\,, \tag{1.36}$$

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

1.7.2 Least Squares and the Within Estimators

If we keep maintaining the standard OLS assumptions lined up in Sect. 1.2, the LSDV estimator of model (1.36), following (1.8)-(1.9), is BLUE. In addition, 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 (1.36), the composite dummy matrix D is

$$D = \left((I_{N_1 N_2 N_3} \otimes \iota_T),\ (I_{N_1 N_2} \otimes \iota_{N_3} \otimes I_T),\ (\iota_{N_1} \otimes I_{N_1 N_3 T}),\ (I_{N_1} \otimes \iota_{N_2} \otimes I_{N_3 T}) \right)$$

with size $(N_1 N_2 N_3 T \times (N_1 N_2 N_3 + N_1 N_2 T + N_2 N_3 T + N_1 N_3 T))$ and column rank $(N_1 N_2 N_3 T - (N_1 - 1)(N_2 - 1)(N_3 - 1)(T - 1))$, leading to

$$\begin{aligned} M_D = {} & I_{N_1 N_2 N_3 T} - \left(\bar{J}_{N_1} \otimes I_{N_2 N_3 T}\right) - \left(I_{N_1} \otimes \bar{J}_{N_2} \otimes I_{N_3 T}\right) \\ & - \left(I_{N_1 N_2} \otimes \bar{J}_{N_3} \otimes I_T\right) - \left(I_{N_1 N_2 N_3} \otimes \bar{J}_T\right) + \left(\bar{J}_{N_1 N_2} \otimes I_{N_3 T}\right) \\ & + \left(\bar{J}_{N_1} \otimes I_{N_2} \otimes \bar{J}_{N_3} \otimes I_T\right) + \left(\bar{J}_{N_1} \otimes I_{N_2 N_3} \otimes \bar{J}_T\right) \\ & + \left(I_{N_1} \otimes \bar{J}_{N_2 N_3} \otimes I_T\right) + \left(I_{N_1} \otimes \bar{J}_{N_2} \otimes I_{N_3} \otimes \bar{J}_T\right) + \left(I_{N_1 N_2} \otimes \bar{J}_{N_3 T}\right) \\ & - \left(\bar{J}_{N_1 N_2 N_3} \otimes I_T\right) - \left(\bar{J}_{N_1 N_2} \otimes I_{N_3} \otimes \bar{J}_T\right) - \left(\bar{J}_{N_1} \otimes I_{N_2} \otimes \bar{J}_{N_3 T}\right) \\ & - \left(I_{N_1} \otimes \bar{J}_{N_2 N_3 T}\right) + \bar{J}_{N_1 N_2 N_3 T} \ . \end{aligned}$$

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

$$\begin{aligned} \tilde{y}_{ijst} = {} & y_{ijst} - \bar{y}_{.jst} - \bar{y}_{i.st} - \bar{y}_{ij.t} - \bar{y}_{ijs.} + \bar{y}_{..st} + \bar{y}_{.j.t} + \bar{y}_{.js.} \\ & + \bar{y}_{i..t} + \bar{y}_{i.s.} + \bar{y}_{ij..} - \bar{y}_{...t} - \bar{y}_{..s.} - \bar{y}_{.j..} - \bar{y}_{i...} + \bar{y}_{....} \end{aligned} \tag{1.37}$$

(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 can 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. For example, for model (1.36), the correct degrees of freedom (coming from the rank of M_D) is $(N_1 - 1)(N_2 - 1)(N_3 - 1)(T - 1) - K$.

1.7.3 Incomplete Panels

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 cannot be reached at reasonable cost when the data is large. However, the optimal Within estimator offers a better way to tackle this problem. Just like in Sect. 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 Sect. 1.5 can be easily generalized to any higher dimensions. For model (1.36), assuming that $N_1 = N_2 = N$, the no self-flow transformation can be represented in a smart scalar form using group averages, and reads as

$$\begin{aligned}\tilde{y}_{ijst} = {} & y_{ijst} - \tfrac{1}{N-1}y_{+jst} - \tfrac{1}{N-1}y_{i+st} - \tfrac{1}{N_3}y_{ij+t} - \tfrac{1}{T}y_{ijs+} + \tfrac{1}{(N-1)^2}y_{++st} \\ & + \tfrac{1}{(N-1)N_3}y_{+j+t} + \tfrac{1}{(N-1)T}y_{+js+} + \tfrac{1}{(N-1)N_s}y_{i++t} + \tfrac{1}{(N-1)T}y_{i+s+} \\ & + \tfrac{1}{N_3T}y_{ij++} - \tfrac{1}{(N-1)^2N_3}y_{+++t} - \tfrac{1}{(N-1)^2T}y_{++s+} - \tfrac{1}{(N-1)N_3T}y_{+j++} \\ & - \tfrac{1}{(N-1)N_3T}y_{i+++} + \tfrac{1}{(N-1)^2N_3T}y_{++++} - \tfrac{1}{(N-1)N_3T}y_{ji++} \\ & + \tfrac{1}{(N-1)T}y_{jis+} + \tfrac{1}{(N-1)N_3}y_{ji+t} - \tfrac{1}{N-1}y_{jist}\ ,\end{aligned} \tag{1.38}$$

fully eliminating any computational burden.

General incomplete data can also be handled quite flexibly in the case of four-dimensional models. Remember that the key (iterative) unbalanced-robust transformation in Sect. 1.5 was (1.29), which can be generalized simply into a four dimensional setup. Let the dummy variables matrices for the four fixed effects in (1.36) 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 out the first k fixed effects; 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

$$M_{D_e}^{(k)} = M_{D_e}^{(k-1)} - \left(M_{D_e}^{(k-1)} D_k^e\right) \left[\left(M_{D_e}^{(k-1)} D_k^e\right)' \left(M_{D_e}^{(k-1)} D_k^e\right)\right]^{-} \left(M_{D_e}^{(k-1)} D_k^e\right)', \tag{1.39}$$

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 the model specification and can be done to any other multi-dimensional fixed effects model. The drawback, which cannot 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.

1.8 Varying Coefficients Models

So far we have assumed that the slope coefficients of the models considered are constant. This in fact meant that the heterogeneity was captured through the regression constant only, i.e., via the shifts of this term for different individuals and time points. One of the most important statistical features of multidimensional data sets, however, is that heterogeneity is likely to take more complicated forms, which begs for more complex econometric models. One such approach with a more sophisticated form of heterogeneity is the varying coefficients model, where, along with the fixed effects, we allow the slope coefficients to also vary.

The most general model we can imagine within this framework is

$$y_{ijt} = z'_{ijt}\delta_{ijt} + \varepsilon_{ijt} \tag{1.40}$$

where we force some structure on δ_{ijt}.[8] Note, that this is the general form of any standard multi-dimensional fixed effects model if we assume that $z'_{ijt} = (x'_{ijt},\ 1)$, and that $\delta_{ijt} = (\beta',\ \pi'_{ijt})'$, with π_{ijt} being the composite fixed effect parameters.

The benchmark model we are focusing on, however, follows the spirit of Balestra and Krishnakumar (2008) (pp. 40–43) and Hsiao (2015) (chapter 6), and takes the form

$$y_{ijt} = x'_{ijt}(\beta + \gamma_{ij} + \lambda_t) + \varepsilon_{ijt} \tag{1.41}$$

or similarly,

$$y = X_1\beta + X_2\gamma + X_3\lambda + \varepsilon$$

with

$$\begin{array}{ll} X_1 \equiv \Delta(\iota_{N_1N_2T} \otimes I_K) & (N_1N_2T \times K) \\ X_2 \equiv \Delta(I_{N_1N_2} \otimes \iota_T \otimes I_K) & (N_1N_2T \times N_1N_2K) \\ X_3 \equiv \Delta(\iota_{N_1N_2} \otimes I_T \otimes I_K) & (N_1N_2T \times TK) \end{array}$$

where

$$\Delta = \begin{pmatrix} x'_{111} & & & \\ & x'_{112} & & \\ & & \ddots & \\ & & & x'_{N_1N_2T} \end{pmatrix} \quad (N_1N_2T \times N_1N_2TK)$$

is the diagonally arranged data matrix. Intuitively, this model suggests that the explanatory variables have an effect on y through a common parameter β, but also through γ_{ij} and λ_t, which varies over individual pairs, and time periods. Note that $X = (X_1,\ X_2,\ X_3)$ has no full column rank; in fact it has a rank deficiency of $2K$. Therefore, for identification $2K$ restrictions have to be imposed on the model. We can proceed by simply leaving out for example $\gamma_{N_1N_2}$ and λ_T. A more symmetric way, suggested by Hsiao (2015), is to normalize the average of the heterogeneous parameters:

[8] In this section, we assume that δ_{ijt} is a fixed, unknown coefficient. Random coefficients models, positing distributional assumptions on δ_{ijt} are visited in Chap. 5.

$$\sum_{ij}\gamma_{ij}=0\,;\quad \sum_{t}\lambda_{t}=0\,. \tag{1.42}$$

Then $\tilde{X}=(X_1,\,\tilde{X}_2,\,\tilde{X}_3)$ has full column rank, where $\tilde{X}_1,\tilde{X}_2,\tilde{X}_3$ denote X_1,X_2,X_3 after imposing the proper restrictions. To proceed, the adjusted model can be estimated with straight Least Squares optimally, to get

$$\left(\hat{\beta}'\;\hat{\gamma}'\;\hat{\lambda}'\right)'=\left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}y$$

or alternatively, partialling out γ and λ, and so expressing for β,

$$\hat{\beta}=\left(X_1'M_{\tilde{X}_2\tilde{X}_3}X_1\right)^{-1}X_1'M_{\tilde{X}_2\tilde{X}_3}y$$

with $M_{\tilde{X}_2\tilde{X}_3}$ being the projector matrix orthogonal to $(\tilde{X}_2,\,\tilde{X}_3)$. The problem is that to get $M_{\tilde{X}_2\tilde{X}_3}$, we are faced with inverting $(KN_1N_2\times KN_1N_2)$ matrices, which becomes 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}, but the algebra soon becomes complex. Even if the above estimators can be computed for small samples, we still have the inconvenience of incorporating the restrictions first. Having said this, if we are uncertain about what the proper set of restriction would be, or simply there is scope for experimenting with different restrictions, we would have to redo the estimation each time.

There is, however, a more general, and useful approach to be used to derive estimators for β, and for the heterogeneous parameters as well. For this, we have to apply the theory of Least Squares of incomplete rank detailed in (Searle, 1971, p. 9). Searle shows that all least squares estimators are given by

$$\hat{\delta}=\begin{pmatrix}\hat{\gamma}\\ \hat{\lambda}\\ \hat{\beta}\end{pmatrix}=\left(X'X\right)^{-}X'y+H\zeta=\delta^0+H\zeta\,, \tag{1.43}$$

with δ^0 being the generalized solution, $X=(X_2,\,X_3,\,X_1)$, H being its null-space (for which $XH=0$ holds), and ζ being an arbitrary vector.[9] We want to pick a solution from the set of the infinitely many solutions, which satisfies some conditions. An attractive, natural way to do so is to assume that

$$\sum_{ij}\hat{\gamma}_{ij}=0\,;\quad \sum_{t}\hat{\lambda}_{t}=0\,. \tag{1.44}$$

This can be represented by

$$R'\hat{\delta}=0\,,$$

when

[9] 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 easily computed.

$$R = \begin{pmatrix} 0 & 0 \\ \iota_{N_1 N_2} & 0 \\ 0 & \iota_T \end{pmatrix} \otimes I_K \, .$$

As

$$R' \hat{\delta} = R' \delta^0 + R' H \zeta = 0$$

holds because of (1.44),

$$\zeta = -(R'H)^{-1} R' \delta^0$$

must also hold. As now we have a ζ vector defined explicitly, estimator (1.43) of the parameters becomes

$$\hat{\delta} = (I - H(R'H)^{-1} R') \delta^0 \, . \tag{1.45}$$

As we know that

$$H = \begin{pmatrix} 1 & 1 \\ -\iota_{N_1 N_2} & 0 \\ 0 & -\iota_T \end{pmatrix} \otimes I_K \, ,$$

the only step remaining to be taken is to find generalized solutions for the parameters. First, we set $\beta = 0$, so X_1 drops out. This leaves us in (X_2, X_3) with a rank deficiency of K, which we handle through a generalized inverse. From the Frisch–Waugh–Lovell theorem (with a minor adaptation to handle the singularity) and adding the "estimator" for β we get in the first round, the generalized solutions read as

$$\delta^0 = \begin{pmatrix} \beta^0 \\ \gamma^0 \\ \lambda^0 \end{pmatrix} = \begin{pmatrix} 0 \\ (X_2' X_2)^{-1} X_2' (y - X_3 \lambda^0) \\ (X_3' M_{X_2} X_3)^{-} X_3' M_{X_2} y \end{pmatrix} \, , \tag{1.46}$$

with M_{X_2} being the projection orthogonal to X_2. Putting (1.46) and the definitions of R and H into (1.45) gives the unique estimators

$$\begin{aligned} \hat{\beta} &= \tfrac{1}{N_1 N_2} \sum_{ij} \gamma_{ij}^0 + \frac{1}{T} \sum_t \lambda_t^0 \\ \hat{\gamma}_{ij} &= \gamma_{ij}^0 - \tfrac{1}{N_1 N_2} \sum_{ij} \gamma_{ij}^0 \qquad (i, j = 1 \ldots N_1, N_2) \\ \hat{\lambda}_t &= \lambda_t^0 - \tfrac{1}{T} \sum_t \lambda_t^0 \qquad (t = 1 \ldots T) \end{aligned} \tag{1.47}$$

Fortunately, unbalanced data does not complicate our cause substantially, as the estimators are formulation-wise equivalent to (1.47). Specifically, after we have found the general solutions β^0, γ^0 and λ^0 (in incomplete data), they can be used as in (1.47) to derive estimators.

As seen, this section only considered one specific model. Of course, there is substantial space for experimenting with other possible three-way specifications. For example, models

$$y_{ijt} = x_{ijt}' (\beta + \alpha_{it} + \alpha_{jt}^*) + \varepsilon$$

and

$$y_{ijt} = x'_{ijt}(\beta + \gamma_{ij} + \alpha_{it} + \alpha^*_{jt}) + \varepsilon$$

can also be considered, and can be estimated with the same steps and with slightly modified identifying restrictions as model (1.41). We must keep track, however, of the total number of parameters to be estimated. For the last model considered, this number is $(1 + N_1N_2 + N_1T + N_2T)K$ which can either be a classic case of over-specification, or in worse cases, can exceed the number of observations. This is the main reason why this section focused on simpler models, like (1.41).

Naturally, nothing stops us from generalizing the above models to four, or even to higher dimensions, but computational requirements frequently limit the practical use of such formulations. The estimation of model

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

has the same light computational requirement as model (1.41) (inverting a matrix of order T), but, for example, the estimation of

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

involves matrices of order N_1N_2T, N_2N_3T, and N_1N_3T, which is forbidding even for moderate sample sizes.

References

Abbott, A. and Silles, M. (2016). Determinants of international student migration. *The World Economy*, 30(5):621–635.

Adserà, A. and Pytliková, M. (2015). The role of language in shaping international migration. *The Economic Journal*, 125(586):F49–F81.

Aghion, P., Burgess, R., Redding, S. J., and Zilibotti, F. (2008). The unequal effects of liberalization: Evidence from dismantling the license raj in india. *American Economic Review*, 98(4):1397–1412.

Baier, S. and Bergstrand, J. (2002). On the endogeneity of international trade flows and free trade agreements. Unpublished Manuscript. University of Notre Dame.

Baier, S. L. and Bergstrand, J. H. (2007). Do free trade agreements actually increase members international trade? *Journal of International Economics*, 71:72–95.

Baier, S. L. and Bergstrand, J. H. (2009). Estimating the effects of free trade agreements on international trade flows using matching econometrics. *Journal of International Economics*, 77(1):63–76.

Balazsi, L., Matyas, L., and Wansbeek, T. (2015). The estimation of multidimensional fixed effects panel data models. *Econometric Reviews*, forthcoming, published online: April 2015.

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. In Matyas and Sevestre, editors, *The Econometrics of Panel Data, 3rd edition*, pages 23–48. Springer Verlag.

Baltagi, B. H. (2013). *Econometric Analysis of Panel Data*. John Wiley & Sons, 5 edition.

Baltagi, B. H., Bresson, G., and Etienne, J.-M. (2015). Hedonic housing prices in Paris: An unbalanced spatial lag pseudo-panel model with nested random effects. *Journal of Applied Econometrics*, 30:509–528.

Baltagi, B. H., Egger, P., and Pfaffermayr, M. (2003). A generalized design for bilateral trade flow models. *Economics Letters*, 80:391–397.

Barthel, F. and Neumayer, E. (2015). Spatial dependence in asylum migration. *Journal of Ethnic and Migration Studies*, 41(7):1131–1151.

Bayer, P., McMillan, R., Murphy, A., and Timmins, C. (2016). A dynamic model of demand for houses and neighborhoods. *Econometrica*, 84:893–942.

Beine, M. and Parsons, C. (2015). Climatic factors as determinants of international migration. *Scandinavian Journal of Economics*, 117(2):723–767.

Bellak, C., Leibrecht, M., and Riedl, A. (2008). Labour cost and FDI flows into central and eastern european countries: A survey of the literature and empirical evidence. *Structural Change and Economic Dynamics*, 19(1):17–37.

Belot, M. and Ederveen, S. (2012). Cultural barriers in migration between oecd countries. *Journal of Population Economics*, 25(3):1077–1105.

Bergstrand, J. H., Larch, M., and Yotov, Y. V. (2015). Economic integration agreements, border effects, and distance elasticities in the gravity equation. *European Economic Review*, 78(C):307–327.

Berthelemy, J.-C. (2006). Bilateral donor's interest vs. recipients' development motives in aid allocation: Do all donors behave the same? *Review of Developement Economics*, 10:179–194.

Bertoli, S., Brücker, H., and Fernández-Huertas Moraga, J. (2016). The european crisis and migration to Germany. *Regional Science and Urban Economics*, 60:61–72.

Bertoli, S. and Fernández-Huertas Moraga, J. (2013). Multilateral resistance to migration. *Journal of Development Economics*, 102:79–100.

Biscourp, P., Boutin, X., and Verge, T. (2013). The effect of retail regulations on prices: Evidence from the Loi Galland. *Economic Journal*, 123(573):1279–1312.

Borenstein, S. and Rose, N. L. (1994). Competition and price dispersion in the U.S. airline industry. *Journal of Political Economy*, 102(4):653683.

Bun, M. J. G. and Klaassen, F. J. G. M. (2002). The importance of dynamics in panel gravity models of trade. Unpublished Manuscript. University of Amsterdam.

Bussiere, M., Fidrmuc, J., and Schnatz, B. (2005). Trade integration of central and eastern european countries: Lessons from a gravity model. *Working Paper Series 0545, European Central Bank*.

Carneiro, A., Guimaraes, P., and Portugal, P. (2012). Real wages and the business cycle: Accounting for worker and firm heterogeneity. *American Economic Journal: Macroeconomics*, 4(2):133–152.

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.

Crozet, M., Milet, E., and Mirza, D. (2016). The impact of domestic regulations on international trade in services: Evidence from firm-level data. *Journal of Comparative Economics*, 44(3):585–607.

Czaika, M. and Hobolth, M. (2016). Do restrictive asylum and visa policies increase irregular migration into Europe? *European Union Politics*, 17(3):345–365.

Dubois, P. and Perrone, H. (2015). Price dispersion and informational frictions: Evidence from supermarket purchases. *CEPR Discussion Papers 10906*.

Echevarria, J. and Gardeazabal, J. (2016). Refugee gravitation. *Public Choice*, 169(3–4):269–292.

Egger, P. (2000). A note on the proper econometric specification of the gravity equation. *Economics Letters*, 66(1):25–31.

Egger, P. (2001). European exports and outward foreign direct investment: A dynamic panel data approach. *Review of World Economics (Weltwirtschaftliches Archiv)*, 137(3):427–449.

Egger, P., Larch, M., E., S. K., and Winkelmann, R. (2011). The trade effects of endogenous preferential trade agreements. *American Economic Journal: Economic Policy*, 3(3):113–43.

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.

Egger, P. and Pfaffermayr, M. (2011). Structural estimation of gravity models with path-dependent market entry. Technical report, C.E.P.R. Discussion Papers.

Eilat, Y. and Einav, L. (2004). Determinants of international tourism: A three-dimensional panel data analysis. *Applied Economics*, 36(12):1315–1327.

Figueiredo, E., Lima, L. R., and Orefice, G. (2016). Migration and regional trade agreements: A new gravity estimation. *Review of International Economics*, 24(1):99–125.

Fourie, J. and Santana-Gallego, M. (2011). The impact of mega-sport events on tourist arrivals. *Tourism Management*, 32(6):1364–1370.

Fu, Q., Zhu, Y., and Ren, Q. (2015). The downside of marketization: A multilevel analysis of housing tenure and types in reform-era urban China. *Social Science Research*, 49:126–140.

Gayer, T., Hamilton, J., and Viscusi, W. (2000). Private values of risk tradeoffs at superfund sites: Housing market evidence on learning about risk. *Review of Economics and Statistics*, 82:439–451.

Gerardi, K. S. and Shapiro, A. H. (2009). Does competition reduce price dispersion? new evidence from the airline industry. *Journal of Political Economy*, 117(1):1–37.

Glick, R. and Rose, A. (2002). Does a currency union affect trade? The time series evidence. *European Economic Review*, 46:1125–1151.

Gorodnichenko, Y., Sheremirov, V., and Talavera, O. (2014). Price setting in online markets: Does it click? *NBER Working Papers 20819, National Bureau of Economic Research, Inc.*

Gorodnichenko, Y. and Talavera, O. (2017). Price setting in online markets: Basic facts, international comparisons, and cross-border integration. *American Economic Review*, 107(1):249–282.

Guimaraes, P. and Portugal, P. (2010). A simple feasible procedure to fit models with high-dimensional fixed effects. *Stata Journal*, 10:628–649.

Haller, L. E. and Cotterill, R. W. (1996). Evaluating traditional share-price and residual demand measures of market power in the catsup industry. *Review of Industrial Organization*, 11(3):293–306.

Harris, M. N., Konya, L., and Matyas, L. (2000). Modelling the impact of environmental regulations on bilateral trade flows: OECD, 1990-96. *Melbourne Institute Working Paper No. 11/00*.

Head, K., Mayer, T., and Ries, J. (2010). The erosion of colonial trade linkages after independence. *Journal of International Economics*, 81(1):1–14.

Heyman, F., Sjoholm, F., and Tingvall, P. G. (2007). Is there really a foreign ownership wage premium? Evidence from matched employer-employee data. *Journal of International Economics*, 73(2):355–376.

Hirsch, B. (2013). The impact of female managers on the gender pay gap: Evidence from linked employer-employee data for Germany. *Economics Letters*, 119(3):348–350.

Horrace, W. C. and Schnier, K. E. (2010). Fixed-effect estimation of highly mobile production technologies. *American Journal of Agricultural Economics*, 92(5):1432–1445.

Hsiao, C. (2015). *Analysis of Panel Data*. Econometric Society Monographs, Cambridge University Press, 3 edition.

Hur, J., Alba, J. D., and Park, D. (2010). Effects of hub-and-spoke free trade agreements on trade: A panel data analysis. *World Development*, 38(8):1105–1113.

Koren, M. and Hornok, C. (2017). Winners and losers of globalisation: Sixteen challenges for measurement and theory. In Blundell, R., Cantillon, E., Chizzolini, B., Ivaldi, M., Leininger, W., Marimon, R., Matyas, L., Ogden, T., and Steen, F., editors, *Economics Without Borders – Economic Research for European Policy Challenges*. Cambridge University Press, forthcoming.

Llull, J. (2016). Understanding international migration: Evidence from a new dataset of bilateral stocks (1960–2000). *SERIEs*, 7(2):221–255.

Magee, C. S. (2003). Endogenous preferential trade agreements: An empirical analysis. *Contributions in Economic Analysis & Policy*, 2(1).

Martin, P., Mayer, T., and Thoenig, M. (2008). Make trade not war? *Review of Economic Studies*, 75(3):865–900.

Matyas, L. (1997). Proper econometric specification of the gravity model. *The World Economy*, 20:363–369.

Melitz, J. and Toubal, F. (2012). Native language, spoken language, translation and trade. *Journal of International Economics*, 93(2).

Nuroglu, E. and Kunst, R. (2014). Competing specifications of the gravity equation: A three-way model, bilateral interaction effects, or a dynamic gravity model with time-varying country effects? *Empirical Economics*, 46(2):733–741.

Olivero, M. P. and Yotov, Y. V. (2012). Dynamic gravity: Endogenous country size and asset accumulation. *Canadian Journal of Economics*, 45(1):64–92.

Ortega, F. and Peri, G. (2013). The effect of income and immigration policies on international migration. *Migration Studies*, 1(1):1–35.

Parsley, D. C. (2003). Exchange rate pass-through in a small open economy: Panel evidence from Hong Kong. *International Journal of Finance & Economics*, 8(2):99–107.

Parsley, D. C. and Wei, S.-J. (1999). Border, border, wide and far, how we wonder what you are. *Policy Research Working Papers No. 25*.

Perkins, R. and Neumayer, E. (2014). Geographies of educational mobilities: Exploring the uneven flows of international students. *Geographical Journal*, 180(3):246–259.

Poot, J., Alimi, O., Cameron, M. P., and Maré, D. C. (2016). The gravity model of migration: The successful comeback of an ageing superstar in regional science. *Investigaciones Regionales – Journal of Regional Research*, 36:63–86.

Romalis, J. (2007). NAFTA's and CUSFTA's impact on international trade. *The Review of Economics and Statistics*, 89(3):416–435.

Rose, A. K. and van Wincoop, E. (2001). National money as a barrier to international trade: The real case for currency union. *American Economic Review*, 91(2):386–390.

Searle, S. R. (1971). *Linear Models*. Wiley, New York.

Shin, Y. and Serlenga, L. (2007). Gravity models of intra-EU trade: Application of the CCEP-HT estimation in heterogeneous panels with unobserved common time-specific factors. *Journal of Applied Econometrics*, 22(2):361–381.

Smith, G. W. and Yetman, J. (2007). The curse of Irving Fisher (professional forecasters' version). *Queen's Economics Department Working Paper No. 1144*.

Syed, I., Hill, R. J., and Melser, D. (2008). Flexible spatial and temporal hedonic price indexes for housing in the presence of missing data. *The University of New South Wales, School of Economics, Discussion Paper: 2008/14*.

Thompson, J. M. and Pendell, D. L. (2016). A system approach for three-dimensional panel data to estimate poultry trade impacts due to animal disease outbreaks. *2016 Annual Meeting, Boston, Massachusetts*.

Turnbull, G. K. and van der Vlist, A. J. (2015). Uninformed house buyers and foreclosures. *University of Central Florida, Department of Finance, Working Paper 1503*.

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(18):2105–2112.

Wansbeek, T. and Kapteyn, A. (1989). Estimation of error-components model with incomplete panels. *Journal of Econometrics*, 41:341–361.

Chapter 2
Random Effects Models

Laszlo Balazsi, Badi H. Baltagi, Laszlo Matyas, and Daria Pus

Abstract This chapter deals with the most relevant multi-dimensional random effects panel data models, where, unlike in the case of fixed effects, the number of parameters to be estimated does not increase with the sample size. First, optimal (F)GLS estimators are presented for the textbook-style complete data case, paying special attention to asymptotics. Due to the many (semi-)asymptotic cases, special attention is given to checking under which cases the presented estimators are consistent. Interestingly, some asymptotic cases also carry a "convergence" property, that is the respective (F)GLS estimator converges to the Within estimator, carrying over some of its identification issues. The results are extended to incomplete panels and to higher dimensions as well. Lastly, mixed fixed–random effects models are visited, and some insights on testing for model specifications are considered.

2.1 Introduction

The disturbances of an econometric model in principle include all the factors influencing the behaviour of the dependent variable that cannot be explicitly specified. In a statistical sense, this means all the 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

Laszlo Balazsi
Central European University, Budapest, Hungary

Badi H. Baltagi
Syracuse University, New York, USA

Laszlo Matyas
Central European University, Budapest, Hungary

Daria Pus
The University of Texas at Austin, Texas, USA

L. Matyas (ed.), *The Econometrics of Multi-dimensional Panels*, Advanced Studies in Theoretical and Applied Econometrics 50, DOI 10.1007/978-3-319-60783-2_2

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 Chap. 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 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; Bryk and Raudenbush, 1992). Here, however, we 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 Sect. 2.2 we introduce the most frequently used models in a three-dimensional (3D) panel data setup, Sect. 2.3 deals with the Feasible GLS estimation of these models, while Sect. 2.4 analyses the behaviour of this estimator for incomplete/unbalanced data. Section 2.5 generalizes the models presented to four and higher dimensional data sets, and extends the random effects approach toward a mixed effects framework. Finally, Sect. 2.6 deals with some testing issues, and Sect. 2.7 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 mirror some fixed effects model formulations known from the literature (see Chap. 1 and also, for example, Baltagi et al., 2003; Egger and Pfaffermayr, 2003; Baldwin and Taglioni, 2006; Baier and Bergstrand, 2007). Table 2.1 collects the empirical applications of the random effects models which the empirical chapters of this volume rely upon, along with some selected applications from the literature. It is clear from Table 2.1 that three-dimensional applications are scarce, and that the models used are mostly less sophisticated (some are only 3D representations of 2D models). It is thus clear that the empirical literature can benefit from taking the econometrics of "true" three-dimensional models into account.

The general form of these random effects (or error components) models can be cast 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 is well known from the Gauss–Markov theorem, the Generalized Least Squares (GLS)

Table 2.1 Examples of empirical studies for multi-dimensional random effects models

Study	Topic	Indices (*i-j-t*)	Sample Size	Random Effects	Balanced
Chapter 11 – Trade					
Glick and Rose (2002)	Currency Union	origin country - destination country - year	220 000	μ_{ij}	No
Shin and Serlenga (2007)	Intra-EU Trade		3 800	$\mu_{ij}+\lambda_t$	No
Nuroglu and Kunst (2014)	Factors Explaining Trade		150 000	$\mu_{ij}+\upsilon_{it}+\zeta_{jt}$	No
Chapter 12 – Housing and Prices					
Jun (2016)	Rational and Ethnic Compositions	neighborhood - area - time	35 000	ζ_j	No
Baltagi et al. (2014)	English House Prices	county - district - time	8 300	$\upsilon_i+\mu_{ij}$	No
Baltagi et al. (2015)	Neighbor's Prices	year - arrondissement - quartier - block - flat	157 000	$\upsilon_{ta}+\zeta_{taq}+\lambda_{taqi}$	No
Chapter 15 – Consumer Price Heterogeneity					
Moen et al. (2014)	Retail Price Dispersion	product - store - month	2 775 000	$\alpha_i+\gamma_j$	No
Borenstein and Rose (1994)	US Airline Industry	airport - airport - carrier	1 000	$\gamma_{ij}(FE)+\lambda_t(RE)$	No
Further Examples					
Verropoulou and Joshi (2009)	Mother's Employment and Child Development	outcome - child - family	1 700	υ_{it}	Yes
Svensson and Hagquist (2010)	Adolescent Alcohol Use	individual - municipality - year	15 000	ζ_j	No
Elshandidy and Hussainey (2013)	Risk Disclosure Incentives	firm - sector - year	1 200	$\zeta_j+\lambda_t$	No
Chit et al. (2010)	Exchange Rate Volatility	exporting country - importing country - quarter	8 500	$\mu_{ij}+\lambda_t$	No
Bussiere et al. (2005)	Trade Integration	origin country - destination country - year	50 000	$\mu_{ij}+\lambda_t$	No
Fairbrother (2013)	Comparative Surveys	respondent - country - year	350 000	$\zeta_j+\zeta_{jt}$	No

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), $\mathrm{E}(uu') = \Omega$, 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 at 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 $u_{ijt} = \mu_{ij} + \upsilon_{it} + \zeta_{jt} + \varepsilon_{ijt}$ are elements of the $(N_1 N_2 T \times 1)$, $(N_1 N_2 T \times K)$, and $(N_1 N_2 T \times 1)$ size 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 pairwise uncorrelated, $\mathrm{E}(\mu_{ij}) = 0$, $\mathrm{E}(\upsilon_{it}) = 0$, $\mathrm{E}(\zeta_{jt}) = 0$, and further,

$$\begin{aligned}
\mathrm{E}(\mu_{ij}\mu_{i'j'}) &= \begin{cases} \sigma_\mu^2 & i = i' \text{ and } j = j' \\ 0 & \text{otherwise} \end{cases} \\
\mathrm{E}(\upsilon_{it}\upsilon_{i't'}) &= \begin{cases} \sigma_\upsilon^2 & i = i' \text{ and } t = t' \\ 0 & \text{otherwise} \end{cases} \\
\mathrm{E}(\zeta_{jt}\zeta_{j't'}) &= \begin{cases} \sigma_\zeta^2 & j = j' \text{ and } t = t' \\ 0 & \text{otherwise.} \end{cases}
\end{aligned} \tag{2.3}$$

The covariance matrix of such error components structure is simply

$$\Omega = \mathrm{E}(uu') = \sigma_\mu^2 (I_{N_1 N_2} \otimes J_T) + \sigma_\upsilon^2 (I_{N_1} \otimes J_{N_2} \otimes I_T) + \sigma_\zeta^2 (J_{N_1} \otimes I_{N_2 T}) + \sigma_\varepsilon^2 I_{N_1 N_2 T} \ , \tag{2.4}$$

where 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

$$y_{ijt} = x'_{ijt}\beta + \upsilon_{it} + \zeta_{jt} + \varepsilon_{ijt} \ , \tag{2.5}$$

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

$$\Omega = \sigma_\upsilon^2 (I_{N_1} \otimes J_{N_2} \otimes I_T) + \sigma_\zeta^2 (J_{N_1} \otimes I_{N_2} \otimes I_T) + \sigma_\varepsilon^2 I_{N_1 N_2 T} \ . \tag{2.6}$$

A further restriction on the above model is

$$y_{ijt} = x'_{ijt}\beta + \zeta_{jt} + \varepsilon_{ijt} \ , \tag{2.7}$$

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

$$\Omega = \sigma_\zeta^2 (J_{N_1} \otimes I_{N_2 T}) + \sigma_\varepsilon^2 I_{N_1 N_2 T} \ . \tag{2.8}$$

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

$$y_{ijt} = x'_{ijt}\beta + \mu_{ij} + \lambda_t + \varepsilon_{ijt} \tag{2.9}$$

[1] The symmetric counterpart of model (2.7), with υ_{it} random effects could also be listed here, however, as it has the exact same properties as model (2.7), we take the two models together.

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

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

Now

$$\Omega = \sigma_\mu^2 (I_{N_1 N_2} \otimes J_T) + \sigma_\lambda^2 (J_{N_1 N_2} \otimes I_T) + \sigma_\varepsilon^2 I_{N_1 N_2 T} \,. \tag{2.10}$$

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

$$y_{ijt} = x_{ijt}' \beta + \upsilon_i + \zeta_j + \lambda_t + \varepsilon_{ijt} \,, \tag{2.11}$$

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

$$\begin{aligned} \mathrm{E}(\upsilon_i \upsilon_{i'}) &= \begin{cases} \sigma_\upsilon^2 & i = i' \\ 0 & \text{otherwise} \end{cases} \\ \mathrm{E}(\zeta_j \zeta_{j'}) &= \begin{cases} \sigma_\zeta^2 & j = j' \\ 0 & \text{otherwise} \end{cases} \\ \mathrm{E}(\lambda_t \lambda_{t'}) &= \begin{cases} \sigma_\lambda^2 & t = t' \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \tag{2.12}$$

Its covariance structure is

$$\Omega = \sigma_\upsilon^2 (I_{N_1} \otimes J_{N_2 T}) + \sigma_\zeta^2 (J_{N_1} \otimes I_{N_2} \otimes J_T) + \sigma_\lambda^2 (J_{N_1 N_2} \otimes I_T) + \sigma_\varepsilon^2 I_{N_1 N_2 T} \,. \tag{2.13}$$

Lastly, the simplest model is

$$y_{ijt} = x_{ijt}' \beta + \mu_{ij} + \varepsilon_{ijt} \tag{2.14}$$

with

$$\Omega = \sigma_\mu^2 (I_{N_1 N_2} \otimes J_T) + \sigma_\varepsilon^2 I_{N_1 N_2 T} \,. \tag{2.15}$$

Note that model (2.14) can in fact be considered as a straight panel data model, where the individuals are now the (ij) pairs (so essentially it does not take into account the three-dimensional nature of the data).

2.2.2 Spectral Decomposition of the Covariance Matrices

To estimate the above models, the inverse of Ω is needed, a matrix of size $(N_1 N_2 T \times N_1 N_2 T)$. For even moderately large samples, this is not practically feasible without further elaboration. The common practice is to use the spectral decomposition of

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

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

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

with $\bar{J}_{N_1} = \frac{1}{N_1} J_{N_1}$. 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 group 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

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

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

$$\begin{aligned}\Omega = \sigma_\varepsilon^2 B_{ijt} + (\sigma_\varepsilon^2 + T\sigma_\mu^2)B_{ij.} + (\sigma_\varepsilon^2 + N_2\sigma_\upsilon^2)B_{i.t} + (\sigma_\varepsilon^2 + N_1\sigma_\zeta^2)B_{.jt} \\ +(\sigma_\varepsilon^2 + T\sigma_\mu^2 + N_2\sigma_\upsilon^2)B_{i..} + (\sigma_\varepsilon^2 + T\sigma_\mu^2 + N_1\sigma_\zeta^2)B_{.j.} \\ +(\sigma_\varepsilon^2 + N_2\sigma_\upsilon^2 + N_1\sigma_\zeta^2)B_{..t} + (\sigma_\varepsilon^2 + T\sigma_\mu^2 + N_2\sigma_\upsilon^2 + N_1\sigma_\zeta^2)B_{...} \ .\end{aligned} \tag{2.16}$$

Also note 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

$$\begin{aligned}\sigma_\varepsilon^2\Omega^{-1} = {} & I_{N_1N_2T} - (1-\theta_1)(\bar{J}_{N_1}\otimes I_{N_2T}) - (1-\theta_2)(I_{N_1}\otimes\bar{J}_{N_2}\otimes I_T)\\ & -(1-\theta_3)(I_{N_1N_2}\otimes\bar{J}_T) + (1-\theta_1-\theta_2+\theta_4)(\bar{J}_{N_1N_2}\otimes I_T)\\ & +(1-\theta_1-\theta_3+\theta_5)(\bar{J}_{N_1}\otimes I_{N_2}\otimes\bar{J}_T)\\ & +(1-\theta_2-\theta_3+\theta_6)(I_{N_1}\otimes\bar{J}_{N_2T})\\ & -(1-\theta_1-\theta_2-\theta_3+\theta_4+\theta_5+\theta_6-\theta_7)\bar{J}_{N_1N_2T}\ ,\end{aligned} \tag{2.17}$$

with

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

A nice aspect of this 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 a scalar form as well. This transformation can be represented with its typical element

$$\begin{aligned}\tilde{y}_{ijt} = {} & y_{ijt} - (1-\sqrt{\theta_1})\bar{y}_{.jt} - (1-\sqrt{\theta_2})\bar{y}_{i.t} - (1-\sqrt{\theta_3})\bar{y}_{ij.}\\ & +(1-\sqrt{\theta_1}-\sqrt{\theta_2}+\sqrt{\theta_4})\bar{y}_{..t}\\ & +(1-\sqrt{\theta_1}-\sqrt{\theta_3}+\sqrt{\theta_5})\bar{y}_{.j.} + (1-\sqrt{\theta_2}-\sqrt{\theta_3}+\sqrt{\theta_6})\bar{y}_{i..}\\ & -(1-\sqrt{\theta_1}-\sqrt{\theta}_2-\sqrt{\theta}_3+\sqrt{\theta}_4+\sqrt{\theta}_5+\sqrt{\theta}_6-\sqrt{\theta}_7)\bar{y}_{...}\ ,\end{aligned}$$

where, following the standard ANOVA notation, a bar over the variable means that the mean of the variable was taken with respect to the missing indices. 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 (2.5),

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

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

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

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 (2.7), 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 (2.9), it is

$$\begin{aligned}\sigma_\varepsilon^2 \Omega^{-1} = {} & I_{N_1 N_2 T} - (I_{N_1 N_2} \otimes \bar{J}_T) - (\bar{J}_{N_1 N_2} \otimes I_T) + \bar{J}_{N_1 N_2 T} \\ & + \tfrac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T\sigma_\mu^2}((I_{N_1 N_2} \otimes \bar{J}_T) - \bar{J}_{N_1 N_2 T}) \\ & + \tfrac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + N_1 N_2 \sigma_\lambda^2}((\bar{J}_{N_1 N_2} \otimes I_T) - \bar{J}_{N_1 N_2 T}) + \tfrac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T\sigma_\mu^2 + N_1 N_2 \sigma_\lambda^2} \bar{J}_{N_1 N_2 T} \ ,\end{aligned}$$

so $\sigma_\varepsilon^2 \Omega^{-1/2} y$ in a 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

$$\theta_{12} = \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T\sigma_\mu^2} \ , \quad \theta_{13} = \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + N_1 N_2 \sigma_\lambda^2} \ , \quad \theta_{14} = \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T\sigma_\mu^2 + N_1 N_2 \sigma_\lambda^2} \ .$$

The spectral decomposition of model (2.11), which was in fact proposed by Baltagi (1987), is

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

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

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

where

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

For model (2.14), the inversion gives

$$\sigma_\varepsilon^2 \Omega^{-1} = I_{N_1 N_2 T} - (I_{N_1 N_2} \otimes \bar{J}_T) + \frac{\sigma_\varepsilon^2}{T\sigma_\mu^2 + \sigma_\varepsilon^2}(I_{N_1 N_2} \otimes \bar{J}_T) \,,$$

and so $\sigma_\varepsilon^2 \Omega^{-1/2} y$ can be written up in a 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 2.2 summarizes the key elements in each model's inverse covariance matrix in the finite case.

Table 2.2 Structure of the Ω^{-1} matrices

Model	(2.2)	(2.5)	(2.7)	(2.9)	(2.11)	(2.14)
$I_{N_1 N_2 T}$	+[a]	+	+	+	+	+
$(I_{N_1 N_2} \otimes \bar{J}_T)$	+			+		+
$(I_{N_1} \otimes \bar{J}_{N_2} \otimes I_T)$	+	+				
$(\bar{J}_{N_1} \otimes I_{N_2 T})$	+	+	+			
$(I_{N_1} \otimes \bar{J}_{N_2 T})$	+				+	
$(\bar{J}_{N_1} \otimes I_{N_2} \otimes \bar{J}_T)$	+				+	
$(\bar{J}_{N_1 N_2} \otimes I_T)$	+	+		+	+	
$\bar{J}_{N_1 N_2 T}$	+			+	+	

[a] A "+" sign in a column says which building element is part of the given model's Ω^{-1}.

When the "+" signs in the column of a given model A overlap with that of another model B means that model B is nested in this model A. It can be seen, for example, that all models are in fact nested in (2.2), or that model (2.14) is nested in model (2.9).

When the number of observations grows in one or more dimensions, it may 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, $(k = 1, \ldots, 19)$ 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

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

which is the covariance matrix of the Within estimator. Table 2.3 collects the asymptotic conditions when the models' (F)GLS estimator is converging to a Within estimator.

Table 2.3 Asymptotic conditions when the model's FGLS converges to a Within estimator

Model	Condition
(2.2)	$N_1 \to \infty, N_2 \to \infty, T \to \infty$
(2.5)	$N_1 \to \infty, N_2 \to \infty$
(2.7)	$N_1 \to \infty$
(2.9)	$(N_1 \to \infty,\ T \to \infty)$ or $(N_2 \to \infty,\ T \to \infty)$
(2.11)	$(N_1 \to \infty,\ N_2 \to \infty)$ or $(N_1 \to \infty,\ T \to \infty)$ or $(N_2 \to \infty,\ T \to \infty)$
(2.14)	$T \to \infty$

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), while for the other models, the job is essentially the same. Using the assumptions that the error components are pairwise uncorrelated,

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

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

$$\tilde{u}_{ijt} = u_{ijt} - \bar{u}_{.jt} - \bar{u}_{i.t} - \bar{u}_{ij.} + \bar{u}_{..t} + \bar{u}_{.j.} + \bar{u}_{i..} - \bar{u}_{...} \,. \tag{2.18}$$

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 $i = 1 \ldots N_1$, and $j = 1 \ldots N_2$,

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

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

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

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

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

Transformation according to

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

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

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

Finally, using

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

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

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

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 variance components in model (2.2) can only be consistently estimated 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. Table 2.4 collects the conditions needed for consistency of the estimators of the variance components 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 the 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 are estimated consis-

tently, but σ_ζ^2 is not. 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 at the so-called "mixed effects models", something explored in Sect. 2.5.

Prucha (1984) showed that although in a two-way error components model the consistent estimation of all variance components is a sufficient condition for the GLS and FGLS estimators to be asymptotically equivalent, this is not necessary in all instances. In some cases it is sufficient to have consistent estimation of the variance of the idiosyncratic disturbance terms, but not the other variance components. Generalizing his results to the models we consider, it turns out that Prucha's results hold exactly in the cases cited in Table 2.4. Then, only the variance of σ_ε^2 need to be estimated consistently for this equivalence to hold.

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 (2.5),

$$\begin{array}{ll} \mathrm{E}(\tilde{u}_{ijt}^2) & = \frac{(N_1-1)(N_2-1)}{N_1N_2}\sigma_\varepsilon^2\,, \quad \mathrm{E}((\tilde{u}_{ijt}^a)^2) = \frac{N_1-1}{N_1}(\sigma_\upsilon^2+\sigma_\varepsilon^2) \quad \text{and} \\ \mathrm{E}((\tilde{u}_{ijt}^b)^2) & = \frac{N_2-1}{N_2}(\sigma_\zeta^2+\sigma_\varepsilon^2)\,, \end{array}$$

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

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

respectively. The estimators for the variance components then are

$$\begin{array}{l} \hat{\sigma}_\varepsilon^2 = \frac{1}{(N_1-1)(N_2-1)T}\sum_{ijt}\tilde{\hat{u}}_{ijt}^2 \quad , \quad \hat{\sigma}_\upsilon^2 = \frac{1}{(N_1-1)N_2T}\sum_{ijt}(\tilde{\hat{u}}_{ijt}^a)^2 - \hat{\sigma}_\varepsilon^2\,, \quad \text{and} \\ \hat{\sigma}_\zeta^2 = \frac{1}{N_1(N_2-1)T}\sum_{ijt}(\tilde{\hat{u}}_{ijt}^b)^2 - \hat{\sigma}_\varepsilon^2\,, \end{array}$$

where again $\tilde{\hat{u}}_{ijt}$, $\tilde{\hat{u}}_{ijt}^a$ and $\tilde{\hat{u}}_{ijt}^b$ are obtained by transforming the residual $\hat{u}_{ijt}$ according to M, M^a, and M^b respectively. For model (2.7), as

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

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

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

For model (2.9),

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

with $\tilde{u}_{ijt} = u_{ijt} - \bar{u}_{..t} - \bar{u}_{ij.} + \bar{u}_{...}$, and $\tilde{u}_{ijt}^a = u_{ijt} - \bar{u}_{..t}$, and $\tilde{u}_{ijt}^b = u_{ijt} - \bar{u}_{ij.}$, 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

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

For model (2.11), as

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

with $\tilde{u}_{ijt} = u_{ijt} - \bar{u}_{..t} - \bar{u}_{.j.} - \bar{u}_{i..} + 2\bar{u}_{...}$, $\tilde{u}_{ijt}^a = u_{ijt} - \bar{u}_{..t} - \bar{u}_{.j.} + \bar{u}_{...}$, $\tilde{u}_{ijt}^b = u_{ijt} - \bar{u}_{..t} - \bar{u}_{i..} + \bar{u}_{...}$, and $\tilde{u}_{ijt}^c = u_{ijt} - \bar{u}_{i..} - \bar{u}_{.j.} + \bar{u}_{...}$, 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

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

Lastly, for model (2.14) we get

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

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

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{N_1 N_2 (T-1)} \sum_{ijt} \tilde{\hat{u}}_{ijt}^2 , \quad \text{and} \quad \hat{\sigma}_\mu^2 = \frac{1}{N_1 N_2 T} \sum_{ijt} \hat{u}_{ijt}^2 - \hat{\sigma}_\varepsilon^2 .$$

Standard errors are computed accordingly using $\mathrm{Var}(\hat{\beta}_{FGLS}) = (X'\hat{\Omega}^{-1}X)^{-1}$. In the limiting cases, the usual normalization factors are needed to obtain finite variances. When, for example, N_1 and T are growing, $\sqrt{N_1 T}(\hat{\beta}_{FGLS} - \beta)$ has a normal distribution with zero mean, and $Q_{X\Omega X}^{-1}$ variance, where $Q_{X\Omega X}^{-1} = \mathrm{plim}_{N_1,T\to\infty} \frac{X'\hat{\Omega}^{-1}X}{N_1 T}$ 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 (2.14). It can be shown, just as with the usual 2D panel models, $\mathrm{Var}(\hat{\beta}_{OLS}) = (X'X)^{-1}X'\hat{\Omega}X(X'X)^{-1}$ (with $\hat{\Omega}$ being model-specific, but let us assume for now that it corresponds to (2.15)).

Table 2.4 Conditions for the consistency of the variance components estimation

Model	Consistency requirements
(2.2)	$(N_1 \to \infty, N_2 \to \infty)$ or $(N_1 \to \infty, T \to \infty)$ or $(N_2 \to \infty, T \to \infty)$
(2.5)	$(T \to \infty)$ or $(N_1 \to \infty, N_2 \to \infty)$
(2.7)	$(N_2 \to \infty)$ or $(T \to \infty)$
(2.9)	$(N_1 \to \infty, T \to \infty)$ or $(N_2 \to \infty, T \to \infty)$
(2.11)	$(N_1 \to \infty, N_2 \to \infty, T \to \infty)$
(2.14)	$(N_1 \to \infty)$ or $(N_2 \to \infty)$

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 2.5 collects normalization factors needed for a finite $\mathrm{Var}(\hat{\beta}_{OLS})$ for the different models considered. As it is uncommon to normalize with 1, or with an expression like $\frac{\sqrt{N_1 N_2}}{\sqrt{A}}$, some insights into the normalizations are given in Appendix 1.

Table 2.5 Normalization factors for the finiteness of $\hat{\beta}_{OLS}$

Model	(2.2)	(2.5)	(2.7)	(2.9)	(2.11)	(2.14)
$N_1 \to \infty$	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 at the highest rates.

Another interesting aspect is revealed by comparing Tables 2.3 and 2.4, that is the consistency requirements for the estimation of the variance components (Table 2.4) and the asymptotic results, when the FGLS converges to the Within estimator (Table 2.3).

Table 2.6 Asymptotic results when the OLS should be used

Model	(2.2)	(2.5)	(2.7)	(2.9)	(2.11)	(2.14)
$N_1 \to \infty$			$+^a$			$-$
$N_2 \to \infty$			$-$			$-$
$T \to \infty$		$-$	$-$			$+$
$N_1, N_2 \to \infty$	$-$	$+$	$+$		$+$	$-$
$N_1, T \to \infty$	$-$	$-$	$+$	$+$	$+$	$+$
$N_2, T \to \infty$	$-$	$-$	$-$	$+$	$+$	$+$
$N_1, N_2, T \to \infty$	$+$	$+$	$+$	$+$	$+$	$+$

[a] 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 cannot be estimated consistently (under the respective asymptotics).

As can be seen in Table 2.6, 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 previously estimable parameters suddenly become 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 2.6, most significant for models (2.9) and (2.11). 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 enough for the consistency of the FGLS, but do 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 Unbalanced Data

2.4.1 Structure of the Covariance Matrices

Our analysis has so far concentrated on balanced panels. We know, however, that real life data sets usually have some kind of incompleteness embedded. This may 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 Chap. 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 is quite complex now, we need to introduce a few new notations 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, $\tilde{n}_{ij} = N_1 N_2$, $\tilde{n}_{it} = N_1 T$, and $\tilde{n}_{jt} = N_2 T$. 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

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

with μ, υ, ζ, ε being the stacked vectors of μ_{ij}, υ_{it}, ζ_{jt}, and ε_{ijt}, of respective lengths $N_1 N_2$, $N_1 T$, $N_2 T$, $N_1 N_2 T$, and ι is the column of ones with the size in the index. The covariance matrix can then be represented by

$$\mathrm{E}(uu') = \Omega = D_1 D_1' \sigma_\mu^2 + D_2 D_2' \sigma_\upsilon^2 + D_3 D_3' \sigma_\zeta^2 + I\sigma_\varepsilon^2\,,$$

which is identical to (2.4). 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

[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}.

$$
\begin{aligned}
D_1 &= \text{diag}\{V_{11}\iota_T, V_{12}\iota_T, \ldots, V_{N_1N_2}\iota_T\} \quad \text{of size} \quad (\textstyle\sum_{ij}|T_{ij}| \times \tilde{n}_{ij}) \,, \\
D_2 &= \text{diag}\left\{(V'_{11}, V'_{12}, \ldots, V'_{1N_2})', \ldots, (V'_{N_1 1}, V'_{N_1 2}, \ldots, V'_{N_1N_2})'\right\} \\
&\qquad\qquad \text{of size} \quad (\textstyle\sum_{ij}|T_{ij}| \times \tilde{n}_{it}) \\
D_3 &= \left(\text{diag}\{V'_{11}, V'_{12}, \ldots, V'_{1N_2}\}', \ldots, \text{diag}\{V'_{N_1 1}, V'_{N_1 2}, \ldots, V'_{N_1N_2}\}'\right)' \\
&\qquad\qquad \text{of size} \quad (\textstyle\sum_{ij}|T_{ij}| \times \tilde{n}_{jt}) \,.
\end{aligned}
$$

These then can be used to construct the covariance matrix as

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

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 (2.4). The job is the same for other models. For models (2.5) and (2.7),

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

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 (2.5), and

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

for model (2.7). Again, if the panel were in fact complete, we would get back (2.6) and (2.8). The incomplete data covariance matrix of model (2.9) is

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

with

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

The covariance matrix for model (2.11) is

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

where

$$
\begin{aligned}
D_5 &= \text{diag}\left\{(V'_{11}\iota_T, V'_{12}\iota_T, \ldots, V'_{1N_2}\iota_T)', \ldots, (V'_{N_1 1}\iota_T, V'_{N_1 2}\iota_T, \ldots, V'_{N_1N_2}\iota_T)'\right\} \\
D_6 &= \Big(\text{diag}\{V'_{11}\iota_T, V'_{12}\iota_T, \ldots, V'_{1N_2}\iota_T\}', \ldots \times \\
&\qquad \times \ldots, \text{diag}\{V'_{N_1 1}\iota_T, V'_{N_1 2}\iota_T, \ldots, V'_{N_1N_2}\iota_T\}'\Big)' \,.
\end{aligned}
$$

of sizes $(\sum_{ij}|T_{ij}| \times N_1)$, and $(\sum_{ij}|T_{ij}| \times N_2)$. Lastly, for model (2.14) we simply get

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

An important practical difficulty is that the spectral decompositions of the covariance matrices introduced in Sect. 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 (2.16) 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 a 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 it's quasi-spectral decomposition, Ω^{*-1}, approximates 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 observations 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 trades 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.

2.4.2 The Inverse of the Covariance Matrices

The solution proposed above, however, has two potential drawbacks. First, the inverse, though reached at a 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 so, 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 components 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

[4] This can be demonstrated to be valid by simulation.

$$\sigma_\varepsilon^2 \Omega^{-1} = P^b - P^b D_3 (R^c)^{-1} D_3' P^b \tag{2.19}$$

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

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

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, and then construct R^b to get P^b. Proof of (2.19) can be found in Appendix 2. Note that to get the inverse, we have to invert $\min\{N_1 T; N_2 T; N_1 N_2\}$ matrices. The quasi-scalar form of (2.19) (which corresponds to the incomplete data version of transformation (2.17)) 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 (2.5), it is with appropriately redefining P and R,

$$\sigma_\varepsilon^2 \Omega^{-1} = P^a - P^a D_3 (R^b)^{-1} D_3' P^a\,, \tag{2.20}$$

where now

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

with the largest matrix to be inverted now of size $\min\{N_1 T; N_2 T\}$. For model (2.7), it is even simpler,

$$\sigma_\varepsilon^2 \Omega^{-1} = I - D_3 (R^a)^{-1} D_3' \quad \text{with} \quad R^a = D_3' D_3 + \frac{\sigma_\varepsilon^2}{\sigma_\zeta^2} I\,, \tag{2.21}$$

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 (2.9), the inverse is

$$\sigma_\varepsilon^2 \Omega^{-1} = P^a - P^a D_4 (R^b)^{-1} D_4' P^a \tag{2.22}$$

where

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

and we have to invert a $\min\{N_1 N_2; T\}$ sized matrix. For model (2.11), the inverse is again the result of a three-step procedure:

$$\sigma_\varepsilon^2 \Omega^{-1} = P^b - P^b D_4 (R^c)^{-1} D_4' P^b \ , \tag{2.23}$$

where

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

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

$$\sigma_\varepsilon^2 \Omega^{-1} = I - D_1 (R^a)^{-1} D_1' \quad \text{with} \quad R^a = D_1' D_1 + \frac{\sigma_\varepsilon^2}{\sigma_\mu^2} I \ , \tag{2.24}$$

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 (2.18) 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 an incompleteness-robust alternative to (2.18), i.e., a transformation which clears the non-idiosyncratic random effects from u_{ijt} in the case of incomplete

data (see Chap. 1). The problem is that most of these transformations involve the manipulation of large matrices resulting in a 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

$$\mathrm{E}(u_{ijt}^2) = \sigma_\mu^2 + \sigma_\upsilon^2 + \sigma_\zeta^2 + \sigma_\varepsilon^2 \,, \tag{2.25}$$

but now let us define

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

It can be seen that

$$\begin{aligned} \mathrm{E}((\tilde{u}_{ijt}^a)^2) = \tfrac{|T_{ij}|-1}{|T_{ij}|}(\sigma_\upsilon^2 + \sigma_\zeta^2 + \sigma_\varepsilon^2) \,, \quad \mathrm{E}((\tilde{u}_{ijt}^b)^2) = \tfrac{n_{it}-1}{n_{it}}(\sigma_\mu^2 + \sigma_\zeta^2 + \sigma_\varepsilon^2) \,, \\ \text{and} \quad \mathrm{E}((\tilde{u}_{ijt}^c)^2) = \tfrac{n_{jt}-1}{n_{jt}}(\sigma_\mu^2 + \sigma_\upsilon^2 + \sigma_\varepsilon^2) \,. \end{aligned} \tag{2.26}$$

Combining (2.25) with (2.26) identifies all four variance components. The appropriate estimators are then

$$\begin{aligned} \hat{\sigma}_\mu^2 &= \tfrac{1}{\sum_{ij}|T_{ij}|}\textstyle\sum_{ijt}\hat{u}_{ijt}^2 - \tfrac{1}{\tilde{n}_{ij}}\sum_{ij}\tfrac{1}{|T_{ij}|-1}\sum_t(\tilde{\hat{u}}_{ijt}^a)^2 \\ \hat{\sigma}_\upsilon^2 &= \tfrac{1}{\sum_{ij}|T_{ij}|}\textstyle\sum_{ijt}\hat{u}_{ijt}^2 - \tfrac{1}{\tilde{n}_{it}}\sum_{it}\tfrac{1}{n_{it}-1}\sum_j(\tilde{\hat{u}}_{ijt}^b)^2 \\ \hat{\sigma}_\zeta^2 &= \tfrac{1}{\sum_{ij}|T_{ij}|}\textstyle\sum_{ijt}\hat{u}_{ijt}^2 - \tfrac{1}{\tilde{n}_{jt}}\sum_{jt}\tfrac{1}{n_{jt}-1}\sum_i(\tilde{\hat{u}}_{ijt}^c)^2 \\ \hat{\sigma}_\varepsilon^2 &= \tfrac{1}{\sum_{ij}|T_{ij}|}\textstyle\sum_{ijt}\hat{u}_{ijt}^2 - \hat{\sigma}_\mu^2 - \hat{\sigma}_\upsilon^2 - \hat{\sigma}_\zeta^2 \,, \end{aligned} \tag{2.27}$$

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

$$\begin{aligned} \hat{\sigma}_\upsilon^2 &= \tfrac{1}{\sum_{ij}|T_{ij}|}\textstyle\sum_{ijt}\hat{u}_{ijt}^2 - \tfrac{1}{\tilde{n}_{it}}\sum_{it}\tfrac{1}{n_{it}-1}\sum_j(\tilde{\hat{u}}_{ijt}^b)^2 \\ \hat{\sigma}_\zeta^2 &= \tfrac{1}{\sum_{ij}|T_{ij}|}\textstyle\sum_{ijt}\hat{u}_{ijt}^2 - \tfrac{1}{\tilde{n}_{jt}}\sum_{jt}\tfrac{1}{n_{jt}-1}\sum_i(\tilde{\hat{u}}_{ijt}^c)^2 \\ \hat{\sigma}_\varepsilon^2 &= \tfrac{1}{\sum_{ij}|T_{ij}|}\textstyle\sum_{ijt}\hat{u}_{ijt}^2 - \hat{\sigma}_\upsilon^2 - \hat{\sigma}_\zeta^2 \,, \end{aligned} \tag{2.28}$$

whereas for model (2.7), with $u_{ijt} = \zeta_{jt} + \varepsilon_{ijt}$, they are

$$\begin{aligned} \hat{\sigma}_\zeta^2 &= \tfrac{1}{\sum_{ij}|T_{ij}|}\textstyle\sum_{ijt}\hat{u}_{ijt}^2 - \tfrac{1}{\tilde{n}_{jt}}\sum_{jt}\tfrac{1}{n_{jt}-1}\sum_i(\tilde{\hat{u}}_{ijt}^c)^2 \\ \hat{\sigma}_\varepsilon^2 &= \tfrac{1}{\sum_{ij}|T_{ij}|}\textstyle\sum_{ijt}\hat{u}_{ijt}^2 - \hat{\sigma}_\zeta^2 \,, \end{aligned} \tag{2.29}$$

Note that the latter two estimators can be obtained from (2.27) by assuming $\hat{\sigma}_\mu^2 = 0$ for model (2.5), and $\hat{\sigma}_\mu^2 = \hat{\sigma}_\upsilon^2 = 0$ for model (2.7).

For model (2.9), let us redefine the $\tilde{u}_{ijt}^k$-s as

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

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

$$\begin{gathered}\mathrm{E}((\tilde{u}_{ijt}^a)^2) = \tfrac{|T_{ij}|-1}{|T_{ij}|}\ (\sigma_\lambda^2 + \sigma_\varepsilon^2)\ , \quad \mathrm{E}((\tilde{u}_{ijt}^b)^2) = \tfrac{n_t-1}{n_t}(\sigma_\mu^2 + \sigma_\varepsilon^2)\ , \\ \text{and} \quad \mathrm{E}(u_{ijt}^2) = \sigma_\mu^2 + \sigma_\lambda^2 + \sigma_\varepsilon^2\ .\end{gathered}$$

From this set of identifying equations, the estimators are simply

$$\begin{aligned}\hat{\sigma}_\mu^2 &= \tfrac{1}{\sum_{ij}|T_{ij}|}\textstyle\sum_{ijt}\hat{u}_{ijt}^2 - \tfrac{1}{\bar{n}_{ij}}\sum_{ij}\tfrac{1}{|T_{ij}|-1}\sum_t(\tilde{\hat{u}}_{ijt}^a)^2 \\ \hat{\sigma}_\lambda^2 &= \tfrac{1}{\sum_{ij}|T_{ij}|}\textstyle\sum_{ijt}\hat{u}_{ijt}^2 - \tfrac{1}{T}\sum_t\tfrac{1}{n_t-1}\sum_{ij}(\tilde{\hat{u}}_{ijt}^b)^2 \\ \hat{\sigma}_\varepsilon^2 &= \tfrac{1}{\sum_{ij}|T_{ij}|}\textstyle\sum_{ijt}\hat{u}_{ijt}^2 - \hat{\sigma}_\mu^2 - \hat{\sigma}_\lambda^2\ .\end{aligned} \tag{2.30}$$

For model (2.14), with $u_{ijt} = \mu_{ij} + \varepsilon_{ijt}$, keeping the definition of $\tilde{u}_{ijt}^a$, we get

$$\begin{aligned}\hat{\sigma}_\mu^2 &= \tfrac{1}{\sum_{ij}|T_{ij}|}\textstyle\sum_{ijt}\hat{u}_{ijt}^2 - \tfrac{1}{\bar{n}_{ij}}\sum_{ij}\tfrac{1}{|T_{ij}|-1}\sum_t(\tilde{\hat{u}}_{ijt}^a)^2 \\ \hat{\sigma}_\varepsilon^2 &= \tfrac{1}{\sum_{ij}|T_{ij}|}\textstyle\sum_{ijt}\hat{u}_{ijt}^2 - \hat{\sigma}_\mu^2\ .\end{aligned} \tag{2.31}$$

Finally, for model (2.11), as now $u_{ijt} = \upsilon_i + \zeta_j + \lambda_t + \varepsilon_{ijt}$, using

$$\tilde{u}_{ijt}^a = u_{ijt} - \frac{1}{n_i}\sum_{jt} u_{ijt}\ , \quad \tilde{u}_{ijt}^b = u_{ijt} - \frac{1}{n_j}\sum_{it} u_{ijt}\ , \quad \tilde{u}_{ijt}^c = u_{ijt} - \frac{1}{n_t}\sum_{ij} u_{ijt}\ ,$$

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

$$\begin{gathered}\mathrm{E}((\tilde{u}_{ijt}^a)^2) = \tfrac{n_i-1}{n_i}(\sigma_\zeta^2 + \sigma_\lambda^2 + \sigma_\varepsilon^2)\ , \quad \mathrm{E}((\tilde{u}_{ijt}^b)^2) = \tfrac{n_j-1}{n_j}(\sigma_\upsilon^2 + \sigma_\lambda^2 + \sigma_\varepsilon^2)\ , \\ \mathrm{E}((\tilde{u}_{ijt}^c)^2) = \tfrac{n_t-1}{n_t}(\sigma_\upsilon^2 + \sigma_\zeta^2 + \sigma_\varepsilon^2)\ , \quad \text{and} \quad \mathrm{E}(u_{ijt}^2) = \sigma_\upsilon^2 + \sigma_\zeta^2 + \sigma_\lambda^2 + \sigma_\varepsilon^2\ ,\end{gathered}$$

in turn leading to

$$\begin{aligned}\hat{\sigma}_\upsilon^2 &= \tfrac{1}{\sum_{ij}|T_{ij}|}\textstyle\sum_{ijt}\hat{u}_{ijt}^2 - \tfrac{1}{\bar{n}_i}\sum_{ij}\tfrac{1}{n_i-1}\sum_{jt}(\tilde{\hat{u}}_{ijt}^a)^2 \\ \hat{\sigma}_\upsilon^2 &= \tfrac{1}{\sum_{ij}|T_{ij}|}\textstyle\sum_{ijt}\hat{u}_{ijt}^2 - \tfrac{1}{\bar{n}_j}\sum_{it}\tfrac{1}{n_j-1}\sum_{it}(\tilde{\hat{u}}_{ijt}^b)^2 \\ \hat{\sigma}_\zeta^2 &= \tfrac{1}{\sum_{ij}|T_{ij}|}\textstyle\sum_{ijt}\hat{u}_{ijt}^2 - \tfrac{1}{T}\sum_{jt}\tfrac{1}{n_t-1}\sum_{ij}(\tilde{\hat{u}}_{ijt}^c)^2 \\ \hat{\sigma}_\varepsilon^2 &= \tfrac{1}{\sum_{ij}|T_{ij}|}\textstyle\sum_{ijt}\hat{u}_{ijt}^2 - \hat{\sigma}_\upsilon^2 - \hat{\sigma}_\zeta^2 - \hat{\sigma}_\lambda^2\ .\end{aligned} \tag{2.32}$$

2.5 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 grows exponentially along with the dimensions, there is no point in attempting 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.

2.5.1 4D and Beyond

The baseline 4D model we use reads, with $i = 1 \dots N_1$, $j = 1 \dots N_2$, $s = 1 \dots N_3$, and $t = 1 \dots T$, as

$$y_{ijst} = x'_{ijst}\beta + \mu_{ijs} + \upsilon_{ist} + \zeta_{jst} + \lambda_{ijt} + \varepsilon_{ijst} = x'_{ijst}\beta + u_{ijst} \ , \tag{2.33}$$

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

$$\begin{aligned}
\Omega = \mathrm{E}(uu') = \ & \sigma_\mu^2(I_{N_1N_2N_3} \otimes J_T) + \sigma_\upsilon^2(I_{N_1} \otimes J_{N_2} \otimes I_{N_3T}) \\
& + \sigma_\zeta^2(J_{N_1} \otimes I_{N_2N_3T}) + \sigma_\lambda^2(I_{N_1N_2} \otimes J_{N_3} \otimes I_T) + \sigma_\varepsilon^2 I_{N_1N_2N_3T} \ .
\end{aligned} \tag{2.34}$$

Its inverse can be simply calculated following the method developed in Sect. 2.3, and the estimation of the variance components can also be derived as in Sect. 2.4 (see for details Appendix 3).

The estimation procedure is not too difficult in the incomplete case either, at least 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 $|T_{ijs}|$ identical observa-

tions for each (ijs) pair. First, let us write up the covariance matrix of (2.33) as

$$\Omega = \mathrm{E}(uu') = \sigma_\varepsilon^2 I + \sigma_\mu^2 D_1 D_1' + \sigma_\upsilon^2 D_2 D_2' + \sigma_\zeta^2 D_3 D_3' + \sigma_\lambda^2 D_4 D_4' \,, \tag{2.35}$$

where, in the complete case,

$$D_1 = (I_{N_1 N_2 N_3} \otimes \iota_T) \,, \quad D_2 = (I_{N_1} \otimes \iota_{N_2} \otimes I_{N_3 T}) \,, \quad D_3 = (\iota_{N_1} \otimes I_{N_2 N_3 T}) \,,$$
$$D_4 = (I_{N_1 N_2} \otimes \iota_{N_3} \otimes I_T) \,,$$

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

$$\Omega^{-1}\sigma_\varepsilon^2 = P^c - P^c D_4 (R^d)^{-1} D_4' P^c \,, \tag{2.36}$$

where P^c and R^d are obtained in the following steps:

$$R^d = D_4' P^c D_4 + \frac{\sigma_\varepsilon^2}{\sigma_\lambda^2} \,, \quad P^c = P^b - P^b D_3 (R^c)^{-1} D_3' P^b \,,$$
$$R^c = D_3' P^b D_3 + \frac{\sigma_\varepsilon^2}{\sigma_\zeta^2} \,, \quad P^b = P^a - P^a D_2 (R^b)^{-1} D_2' P^a \,,$$
$$R^b = D_2' P^a D_2 + \frac{\sigma_\varepsilon^2}{\sigma_\upsilon^2} \,, \quad P^a = I - D_1 (R^a)^{-1} D_1' \,, \quad \text{and } R^a = D_1' D_1 + \frac{\sigma_\varepsilon^2}{\sigma_\mu^2} \,.$$

Even though the calculation above alleviates some of the "dimensionality curse",[5] 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, notwithstanding the size of the formulas. The results are presented in Appendix 3.

2.5.2 Mixed FE-RE Models

As briefly noted earlier, when one of the indices is small, it makes more sense to treat the effects depending on that index as fixed. As an illustration, consider an *employee i–employer j–time t*-type data set, where we usually have a very large set of i, but relatively low j and t. All this means, that the all-encompassing model (2.2) can now be rewritten as

$$y_{ijt} = x_{ijt}'\beta + \alpha_{jt} + \mu_{ij} + \upsilon_{it} + \varepsilon_{ijt} \,, \tag{2.37}$$

or, similarly,

[5] The higher the dimension of the panel, the larger the size of the matrices we have to work with.

$$y = X\beta + D_1\alpha + D_2\mu + D_3\upsilon + \varepsilon = X\beta + D_1\alpha + u\ ,$$

with $D_1 = (\iota_{N_1} \otimes I_{N_2 T})$, $D_2 = (I_{N_1 N_2} \otimes \iota_T)$, and $D_3 = (I_{N_1} \otimes \iota_{N_2} \otimes I_T)$. We assume that α_{jt} enters the model as a fixed effect and through dummy variables, and that $u_{ijt} = \mu_{ij} + \upsilon_{it} + \varepsilon_{ijt}$ remains the random component. To estimate such model specification, keeping an eye on optimality, we have to follow a two-step procedure. First, to get rid of the fixed effects, we define a projection orthogonal to α_{jt}. Then, on this transformed model, we perform FGLS. The resulting estimator is analytically not too complicated, and although as seen in Chap. 1, restricted x_{jt} regressors cannot be estimated from (2.37), $\hat{\beta}_{Mixed}$ is identified and consistent for the rest of the variables. This is a substantial improvement over the FGLS estimation of model (2.2), when N_2 and T are both small, as in such cases, as shown in Sect. 2.4, the inconsistency of the variance components estimators carries over to the model parameters. The projection needed to eliminate α_{jt} is

$$M_{D_1} = I - D_1(D_1'D_1)^{-1}D_1' \quad \text{or in a scalar form,} \quad \tilde{y}_{ijt} = y_{ijt} - \bar{y}_{.jt}\ . \tag{2.38}$$

Note that the resulting transformed (2.37),

$$\tilde{y}_{ijt} = \tilde{x}_{ijt}\beta + \tilde{u}_{ijt}\ , \tag{2.39}$$

is now a simple error components model with a slightly less trivial random effects structure embedded in $\tilde{u}_{ijt}$. In fact,

$$\begin{aligned}
\Omega &= \mathrm{E}(\tilde{u}\tilde{u}') \\
&= \mathrm{E}(M_{D_1} uu' M_{D_1}) = M_{D_1} D_2 D_2' M_{D_1}\sigma_\mu^2 + M_{D_1} D_3 D_3' M_{D_1}\sigma_\upsilon^2 + M_{D_1}\sigma_\varepsilon^2 \\
&= ((I_{N_1} - \bar{J}_{N_1}) \otimes I_{N_2} \otimes \bar{J}_T) T\sigma_\mu^2 + ((I_{N_1} - \bar{J}_{N_1}) \otimes \bar{J}_{N_2} \otimes I_T) N_2\sigma_\upsilon^2 \\
&\quad + ((I_{N_1} - \bar{J}_{N_1}) \otimes I_{N_2 T})\sigma_\varepsilon^2\ ,
\end{aligned}$$

while its inverse can be derived using the trick introduced in Sect. 2.3 (using the substitution $I_{N_1} = Q_{N_1} + \bar{J}_{N_1}$), giving

$$\begin{aligned}
\Omega^{-1}\sigma_\varepsilon^2 = {} & [I_{N_1 N_2 T} - (\bar{J}_{N_1} \otimes I_{N_2 T})] \\
& -(1-\theta_1)\,[(I_{N_1} \otimes \bar{J}_{N_2} \otimes I_T) - (\bar{J}_{N_1 N_2} \otimes I_T)] \\
& -(1-\theta_2)\,[(I_{N_1 N_2} \otimes \bar{J}_T) - (\bar{J}_{N_1} \otimes I_{N_2} \otimes \bar{J}_T)] \\
& +(1-\theta_1-\theta_2+\theta_3)\,[(I_{N_1} \otimes \bar{J}_{N_2 T}) - \bar{J}_{N_1 N_2 T}]
\end{aligned}$$

with

$$\theta_1 = \frac{\sigma_\varepsilon^2}{N_2\sigma_\upsilon^2 + \sigma_\varepsilon^2}\ , \quad \theta_2 = \frac{\sigma_\varepsilon^2}{T\sigma_\mu^2 + \sigma_\varepsilon^2}\ , \quad \text{and} \quad \theta_3 = \frac{\sigma_\varepsilon^2}{N_2\sigma_\upsilon^2 + T\sigma_\mu^2 + \sigma_\varepsilon^2}\ .$$

After all, the mixed effects estimation of (2.37) is identical to the FGLS estimation of (2.39). The only step remaining is to estimate the variance components. In particular,

$$\begin{aligned}\hat{\sigma}_\varepsilon^2 &= \tfrac{1}{(N_1-1)(N_2-1)(T-1)} \textstyle\sum_{ijt} (\tilde{\hat{u}}^a_{ijt})^2 \\ \hat{\sigma}_\mu^2 &= \tfrac{1}{(N_1-1)(N_2-1)T} \textstyle\sum_{ijt} (\tilde{\hat{u}}^b_{ijt})^2 - \hat{\sigma}_\varepsilon^2 \\ \hat{\sigma}_\upsilon^2 &= \tfrac{1}{(N_1-1)N_2(T-1)} \textstyle\sum_{ijt} (\tilde{\hat{u}}^c_{ijt})^2 - \hat{\sigma}_\varepsilon^2 \,,\end{aligned} \tag{2.40}$$

where $\hat{u}_{ijt}$ is the OLS residual, and now

$$\begin{aligned}\tilde{u}^a_{ijt} &= u_{ijt} - \bar{u}_{.jt} - \bar{u}_{i.t} - \bar{u}_{ij.} + \bar{u}_{..t} + \bar{u}_{.j.} + \bar{u}_{i..} - \bar{u}_{...} \,, \\ \tilde{u}^b_{ijt} &= u_{ijt} - \bar{u}_{.jt} - \bar{u}_{i.t} + \bar{u}_{..t} \,, \quad \text{and} \quad \tilde{u}^c_{ijt} = u_{ijt} - \bar{u}_{.jt} - \bar{u}_{ij.} + \bar{u}_{.j.} \,.\end{aligned} \tag{2.41}$$

The next question is to what extent the above algorithm has to be modified for unbalanced data. First, transformation (2.38) is also successful in eliminating α_{jt} from model (2.37) in this case. Second, the resulting transformed covariance matrix now cannot be represented by kronecker products; instead, to invert it, we have to rely on tricks derived in Sect. 2.4. The estimation of the variance components is done by first adjusting the transformations $\tilde{u}^a_{ijt}$, $\tilde{u}^b_{ijt}$ in (2.41) to incomplete data, that is, using their semi-scalar representatives (1.29), (1.28), and for $\tilde{u}^c_{ijt}$

$$\tilde{u}^c = M^{(2)}u = M^{(1)}u - M^{(1)}\tilde{D}_3(\tilde{D}_3'M^{(1)}\tilde{D}_3)^-\tilde{D}_3'M^{(1)}u \,,$$

where u contains the stacked disturbances (with elements u_{ijt}), $\tilde{u}^c$ is its transformed counterpart, $M^{(1)} = I - \tilde{D}_1(\tilde{D}_1'\tilde{D}_1)^-\tilde{D}_1'$, and $\tilde{D}_1$ and $\tilde{D}_3$ are obtained from $D_1 = (I_{N_1N_2} \otimes \iota_T)$ and $D_3 = (\iota_{N_1} \otimes I_{N_2T})$ by leaving out the rows corresponding to missing observations. Finally, we have to set the proper sample sizes in (2.40).

2.6 Testing

In this section, we show for the all-encompassing model (2.2) how to test for the different components of the unobserved heterogeneity. More specifically, we test the nullity of the variance of some random components against the alternative that the given variance is positive. We have to be careful, however, regarding what we assume about the rest of the variances. Testing $H_0: \ \sigma_\mu^2 = 0$ against $H_A: \ \sigma_\mu^2 > 0$ implicitly assumes that $\sigma_\upsilon^2 = \sigma_\zeta^2 = 0$, and so on. In what follows, we collect some null, and alternative hypotheses, and present the mechanism to test them:

$$\begin{aligned}
&\mathrm{H}_0^a: \ \sigma_\mu^2 = 0 \mid \sigma_\upsilon^2 > 0 \,, \ \sigma_\zeta^2 > 0; && \mathrm{H}_\mathrm{A}^a: \ \sigma_\mu^2 > 0 \mid \sigma_\upsilon^2 > 0 \,, \ \sigma_\zeta^2 > 0 \\
&\mathrm{H}_0^b: \ \sigma_\mu^2 = 0 \mid \sigma_\upsilon^2 = 0 \,, \ \sigma_\zeta^2 > 0; && \mathrm{H}_\mathrm{A}^b: \ \sigma_\mu^2 > 0 \mid \sigma_\upsilon^2 = 0 \,, \ \sigma_\zeta^2 > 0 \\
&\mathrm{H}_0^c: \ \sigma_\mu^2 = 0 \mid \sigma_\upsilon^2 = 0 \,, \ \sigma_\zeta^2 = 0; && \mathrm{H}_\mathrm{A}^c: \ \sigma_\mu^2 > 0 \mid \sigma_\upsilon^2 = 0 \,, \ \sigma_\zeta^2 = 0 \\
&\mathrm{H}_0^d: \ \sigma_\mu^2 = 0 \mid \sigma_\upsilon^2 > 0 \,, \ \sigma_\zeta^2 > 0; && \mathrm{H}_\mathrm{A}^d: \ \sigma_\mu^2 > 0 \mid \sigma_\upsilon^2 = 0 \,, \ \sigma_\zeta^2 > 0 \\
&\mathrm{H}_0^e: \ \sigma_\mu^2 = 0 \mid \sigma_\upsilon^2 > 0 \,, \ \sigma_\zeta^2 = 0; && \mathrm{H}_\mathrm{A}^e: \ \sigma_\mu^2 > 0 \mid \sigma_\upsilon^2 = 0 \,, \ \sigma_\zeta^2 = 0 \,.
\end{aligned}$$

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),

$$F = \frac{y' M_{Z_1} D (D' M_{Z_1} D)^{-} D' M_{Z_1} y / (p-r)}{y' M_{Z_2} y / (N_1 N_2 T - \tilde{k} - p + r)} , \tag{2.42}$$

where both M_1 and M_2 are orthogonal projectors, and the degrees of freedom is calculated from p, r, and $\tilde{k}$. Table 2.7 captures each specific matrix and constant for all hypotheses listed above.

Table 2.7 Specific functional forms of the ANOVA F-test

Hypothesis	Z_1	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

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

Although (2.42) 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 (2.42), and find out what the respective projection matrices do to the data:

$$F = \frac{F_1/(p-r)}{F_2/(N_1N_2T - \tilde{k} - p + r)} ,$$

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 (see below). For H_0^a and H_A^a, for example, these are

$$\tilde{y}_{ijt} = y_{ijt} - \bar{y}_{.jt} - \bar{y}_{i.t} - \bar{y}_{ij.} + \bar{y}_{..t} + \bar{y}_{.j.} + \bar{y}_{i..} - \bar{y}_{...} \tag{2.43}$$

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

$$\tilde{\tilde{y}}_{ijt} = y_{ijt} - \bar{y}_{.jt} - \bar{y}_{i.t} + \bar{y}_{..t} . \tag{2.44}$$

To get an insight into the specific formula, note that we actually compare two models, the one where the sources of all variations are cleared (the denominator of (2.42)) with the one where all variation is cleared, except for the one coming from μ_{ij} (the numerator of (2.42)). This is because both under the null and the alternative

we assume, that $\sigma_\upsilon^2 > 0$ and $\sigma_\zeta^2 > 0$, 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, $\sigma_\mu^2 > 0$ also holds, so we eliminate μ_{ij} as well, but save it under the null. The numerator and the denominator of (2.42) are then compared, and if this is sufficiently close to 1, we cannot reject the nullity of σ_μ^2.

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 Sect. 1.5.2. For example, (2.43) corresponds to (1.28), while (2.44) corresponds to (1.28) in the 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 (2.42), with the properly computed degrees of freedom.

2.7 Conclusion

For large data sets, when observations can be considered as samples from an underlying population, random effects specifications seem to be suited to deal with multidimensional data sets. FGLS estimators for three-way error components models are almost as easily obtained as for the traditional 2D panel models, however the resulting asymptotic requirements for their consistency are more peculiar. In fact, now the data may 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 one. This is most 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 may be sufficient to get the full set of parameter estimates (of course, bearing the price of inefficiency), wherever this identification problem persists. The main results of the chapter are extended to treat incomplete data and towards higher dimensions as well.

References

Baier, S. L. and Bergstrand, J. H. (2007). Do free trade agreements actually increase members international trade? *Journal of International Economics*, 71:72–95.

Balazsi, L., Matyas, L., and Wansbeek, T. (2015). The estimation of multidimensional fixed effects panel data models. *Econometric Reviews*, forthcoming, published online: April 2015.

Baldwin, R. and Taglioni, D. (2006). Gravity for dummies and dummies for the gravity equations. *NBER Working Paper 12516*.

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., Bresson, G., and Etienne, J.-M. (2015). Hedonic housing prices in Paris: An unbalanced spatial lag pseudo-panel model with nested random effects. *Journal of Applied Econometrics*, 30:509–528.

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., Egger, P., and Pfaffermayr, M. (2003). A generalized design for bilateral trade flow models. *Economic Letters*, 80:391–397.

Baltagi, B. H., Fingleton, B., and Pirotte, A. (2014). Spatial lag models with nested random effects: An IV procedure with application to English house prices. *Journal of Urban Economics*, 80:76–86.

Baltagi, B. H., Matyas, L., and Sevestre, P. (2008). Error components models. In Matyas, L. and Sevetre, P., editors, *The Econometrics of Panel Data*, pages 49–88. Springer Verlag.

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.

Borenstein, S. and Rose, N. L. (1994). Competition and price dispersion in the U.S. airline industry. *Journal of Political Economy*, 102(4):653683.

Bryk, A. S. and Raudenbush, S. W. (1992). *Hierarchical Linear Models: Applications and Data Analysis Methods*. Newbury Park, CA: Sage Publications, 1 edition.

Bussiere, M., Fidrmuc, J., and Schnatz, B. (2005). Trade integration of central and eastern european countries: Lessons from a gravity model. *Working Paper Series 0545, European Central Bank*.

Chit, M. M., Rizov, M., and Willenbockel, D. (2010). Exchange rate volatility and exports: New empirical evidence from the emerging east asian economies. *The World Economy*, 33(2):239–263.

Davis, P. (2002). Estimating multi-way error components models with unbalanced data structures using instrumental variables. *Journal of Econometrics*, 106:67–95.

Ebbes, P., Bockenholt, U., and Wedel, M. (2004). Regressor and random-effects dependencies in multilevel models. *Statistica Neerlandica*, 58:161–178.

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.

Elshandidy, T. and Hussainey, K. (2013). Aggregated, voluntary, and mandatory risk disclosure incentives: Evidence from uk ftse all-share companies. *International Review of Financial Analysis*, 30:320–333.

Fairbrother, M. (2013). Two multilevel modeling techniques for analyzing comparative longitudinal survey datasets. *Political Science Research and Methods*, 2(1):119–140.

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.

Glick, R. and Rose, A. (2002). Does a currency union affect trade? The time series evidence. *European Economic Review*, 46:1125–1151.

Goldstein, H. (1995). *Multilevel Statistical Models*. John Wiley & Sons, Ltd, Chichester, UK, 2 edition.

Hornok, C. (2011). Gravity or dummies? The limits of identification in gravity estimations. *CeFiG Working Papers*, No. 15.

Hubler, O. (2006). Multilevel and nonlinear panel data models. *Allgemeines Statistisches Archiv*, 90:121–136.

Jun, H.-J. (2016). The effect of racial and ethnic composition on neighborhood economic change: A multilevel and longitudinal look. *Housing Studies*, 31:102–125.

Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, 38:963–974.

Leeuw, J. D. and Kreft, I. (1986). Random coefficient models for multilevel analysis. *Journal of Educational and Behavioral Statistics*, 11:57–85.

Luke, D. A. (2004). *Multilevel Modeling*. Thousand Oaks, CA: SAGE Publications.

Matyas, L. (1998). The gravity model: Some econometric considerations. *The World Economy*, 21:397–401.

Moen, E., Wulfsberg, F., and Aas, O. (2014). Dispersion in retail prices. *Working paper*.

Nuroglu, E. and Kunst, R. (2014). Competing specifications of the gravity equation: A three-way model, bilateral interaction effects, or a dynamic gravity model with time-varying country effects? *Empirical Economics*, 46(2):733–741.

Prucha, I. R. (1984). On the asymptotic efficiency of feasible Aitken estimators for seemingly unrelated regression models with error components. *Econometrica*, 52:203–208.

Rao, R. and Kleffe, J. (1980). Handbook of statistics 1. In Krishnaiah, P., editor, *Estimation of Variance Components*, pages 1–40. North-Holland Publisher.

Scott, M. A., Simonoff, J. S., and Marx, B. D. (2013). *The SAGE handbook of multilevel modeling*. London: SAGE Publications Ltd.

Shin, Y. and Serlenga, L. (2007). Gravity models of intra-eu trade: Application of the CCEP–HT estimation in heterogeneous panels with unobserved common time-specific factors. *Journal of Applied Econometrics*, 22(2):361–381.

Svensson, M. and Hagquist, C. (2010). Adolescents alcohol-use and economic conditions: A multilevel analysis of data from a period with big economic changes. *The European Journal of Health Economics*, 11(6):533–541.

Verropoulou, G. and Joshi, H. (2009). Does mother's employment conflict with child development? Multilevel analysis of british mothers born in 1958. *Journal of Population Economics*, 22(3):665–692.

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.

Appendix 1

Example for normalizing with 1: Model (2.14), $T \to \infty$

$$\begin{aligned}\text{plim}_{T\to\infty}\text{Var}(\hat{\beta}_{OLS}) &= \text{plim}_{T\to\infty}(X'X)^{-1}X'\Omega X(X'X)^{-1} \\ &= \text{plim}_{T\to\infty}\left(\frac{X'X}{T}\right)^{-1}\frac{X'\Omega X}{T^2}\left(\frac{X'X}{T}\right)^{-1} .\end{aligned}$$

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,

$$\text{plim}_{T\to\infty}\text{Var}(\hat{\beta}_{OLS}) = Q_{XX}^{-1}\cdot\text{plim}_{T\to\infty}\frac{\sigma_\varepsilon^2 X'X + \sigma_\mu^2 X'(I_{N_1N_2}\otimes J_T)X}{T^2}\cdot Q_{XX}^{-1} ,$$

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

$$\text{plim}_{T\to\infty}\text{Var}(\hat{\beta}_{OLS}) = Q_{XX}^{-1}\cdot Q_{XBX}\cdot Q_{XX}^{-1} .$$

Note that we can arrive at the same result by first normalizing with the usual $\sqrt{T}$ term, and then adjusting it with $1/\sqrt{T}$ to arrive at a non-zero but bounded variance:

$$\begin{aligned}\text{plim}_{T\to\infty}\text{Var}(\sqrt{T}\hat{\beta}_{OLS}) &= \text{plim}_{T\to\infty}T(X'X)^{-1}X'\Omega X(X'X)^{-1} \\ &= \text{plim}_{T\to\infty}\left(\frac{X'X}{T}\right)^{-1}\frac{X'\Omega X}{T}\left(\frac{X'X}{T}\right)^{-1} ,\end{aligned}$$

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 similar for all other cases and other models.

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

$$\begin{aligned}\text{plim}_{N_1,N_2\to\infty}\text{Var}(\sqrt{N_1N_2}\hat{\beta}_{OLS}) &= \text{plim}_{N_1,N_2\to\infty} N_1N_2 \cdot (X'X)^{-1}X'\Omega X(X'X)^{-1} \\ &= \text{plim}_{N_1,N_2\to\infty}\left(\frac{X'X}{N_1N_2}\right)^{-1}\frac{X'\Omega X}{N_1N_2}\left(\frac{X'X}{N_1N_2}\right)^{-1} \\ &= Q_{XX}^{-1}\cdot\text{plim}_{N_1,N_2\to\infty}\frac{X'\Omega X}{N_1N_2}\cdot Q_{XX}^{-1}\,,\end{aligned}$$

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

$$\begin{aligned}\text{plim}_{N_1,N_2\to\infty}\frac{X'\Omega X}{N_1N_2} &= \text{plim}_{N_1,N_2\to\infty}\frac{\sigma_\varepsilon^2X'X}{N_1N_2} + \text{plim}_{N_1,N_2\to\infty}\frac{\sigma_\mu X'(I_{N_1N_2}\otimes J_T)X}{N_1N_2} \\ &+\text{plim}_{N_1,N_2\to\infty}\frac{\sigma_\upsilon^2X'(I_{N_1}\otimes J_{N_2}\otimes I_T)X}{N_1N_2} + \text{plim}_{N_1,N_2\to\infty}\frac{\sigma_\zeta^2X'(J_{N_1}\otimes I_{N_2T})X}{N_1N_2}\end{aligned} \tag{2.45}$$

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 turn 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 (2.45) are zero, but the fourth is finite:

$$\text{plim}_{N_1,N_2\to\infty}\frac{\sigma_\zeta^2X'(J_{N_1}\otimes I_{N_2T})X}{N_1^2N_2} = Q_{XBX}\,,$$

with some Q_{XBX} finite, positive definite matrix. The same reasoning holds for other models and other asymptotics as well.

Appendix 2: Proof of formula (2.19)

Let us make the proof only for model (2.2) (so formula (2.19)), the rest is just direct application of the derivation below. The outline of the proof is based on Wansbeek and Kapteyn (1989).

First, note 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

$$
D_2' P^a D_2 = D_2' D_2 - D_2' D_1 (R^a)^{-1} D_1' D_2 = R^b - \frac{\sigma_\varepsilon^2}{\sigma_\upsilon^2} I
$$

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 3: Inverse of (2.34), and the estimation of the variance components

$$\begin{aligned}\sigma_\varepsilon^2\Omega^{-1} = {} & I_{N_1N_2N_3T} - (1-\theta_{20})(J_{N_1}\otimes I_{N_2N_3T}) - (1-\theta_{21})(I_{N_1}\otimes J_{N_2}\otimes I_{N_3T}) \\ & -(1-\theta_{22})(I_{N_1N_2}\otimes J_{N_3}\otimes I_T) - (1-\theta_{23})(I_{N_1N_2N_3}\otimes J_T) \\ & +(1-\theta_{24})(J_{N_1N_2}\otimes I_{N_3T}) + (1-\theta_{25})(J_{N_1}\otimes I_{N_2}\otimes J_{N_3}\otimes I_T) \\ & +(1-\theta_{26})(J_{N_1}\otimes I_{N_2N_3}\otimes J_T) + (1-\theta_{27})(I_{N_1}\otimes J_{N_2N_3}\otimes I_T) \\ & +(1-\theta_{28})(I_{N_1}\otimes J_{N_2}\otimes I_{N_3}\otimes J_T) + (1-\theta_{29})(I_{N_1N_2}\otimes J_{N_3T}) \\ & -(1-\theta_{30})(J_{N_1N_2N_3}\otimes I_T) - (1-\theta_{31})(J_{N_1N_2}\otimes I_{N_3}\otimes J_T) \\ & -(1-\theta_{32})(J_{N_1}\otimes I_{N_2}\otimes J_{N_3T}) - (1-\theta_{33})(I_{N_1}\otimes J_{N_2N_3T}) \\ & +(1-\theta_{34})J_{N_1N_2N_3T}\end{aligned}$$

with

$$\begin{aligned}\theta_{20} &= \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+N_1\sigma_\zeta} \quad \theta_{21} = \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+N_2\sigma_\upsilon} \quad \theta_{22} = \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+N_3\sigma_\lambda} \quad \theta_{23} = \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+T\sigma_\mu} \\ \theta_{24} &= \theta_{20}+\theta_{21} - \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+N_1\sigma_\zeta^2+N_2\sigma_\upsilon^2} \quad \theta_{25} = \theta_{20}+\theta_{22} - \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+N_1\sigma_\zeta^2+N_3\sigma_\lambda^2} \\ \theta_{26} &= \theta_{20}+\theta_{23} - \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+N_1\sigma_\zeta^2+T\sigma_\mu^2} \quad \theta_{27} = \theta_{21}+\theta_{22} - \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+N_2\sigma_\upsilon^2+N_3\sigma_\lambda^2} \\ \theta_{28} &= \theta_{21}+\theta_{23} - \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+N_2\sigma_\upsilon^2+T\sigma_\mu^2} \quad \theta_{29} = \theta_{22}+\theta_{23} - \frac{\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} + \frac{\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} + \frac{\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} + \frac{\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} + \frac{\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} - \frac{\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 the case of complete data is as follows:

$$\begin{aligned}\hat\sigma_\varepsilon^2 &= \tfrac{1}{(N_1-1)(N_2-1)(N_3-1)(T-1)}\textstyle\sum_{ijst}\tilde{\tilde u}_{ijst}^2 \\ \hat\sigma_\mu^2 &= \tfrac{1}{(N_1-1)(N_2-1)(N_3-1)T}\textstyle\sum_{ijst}(\tilde{\tilde u}_{ijst}^a)^2 - \hat\sigma_\varepsilon^2 \\ \hat\sigma_\upsilon^2 &= \tfrac{1}{(N_1-1)N_2(N_3-1)(T-1)}\textstyle\sum_{ijst}(\tilde{\tilde u}_{ijst}^b)^2 - \hat\sigma_\varepsilon^2 \\ \hat\sigma_\zeta^2 &= \tfrac{1}{N_1(N_2-1)(N_3-1)(T-1)}\textstyle\sum_{ijst}(\tilde{\tilde u}_{ijst}^c)^2 - \hat\sigma_\varepsilon^2 \\ \hat\sigma_\lambda^2 &= \tfrac{1}{(N_1-1)(N_2-1)N_3(T-1)}\textstyle\sum_{ijst}(\tilde{\tilde u}_{ijst}^d)^2 - \hat\sigma_\varepsilon^2\,,\end{aligned}$$

where, as before, $\hat u_{ijst}$ is the OLS residual, and

$$\begin{aligned}\tilde{\tilde u}_{ijst} &= u_{ijst} - \bar u_{ijs.} - \bar u_{ij.t} - \bar u_{i.st} - \bar u_{.jst} + \bar u_{ij..} + \bar u_{i.s.} + \bar u_{.js.} \\ &\quad +\bar u_{i..t} + \bar u_{.j.t} + \bar u_{..st} - \bar u_{i...} - \bar u_{.j..} - \bar u_{..s.} - \bar u_{...t} + \bar u_{....} \\ \tilde{\tilde u}_{ijst}^a &= u_{ijst} - \bar u_{ij.t} - \bar u_{i.st} - \bar u_{.jst} + \bar u_{i..t} + \bar u_{.j.t} + \bar u_{..st} - \bar u_{...t} \\ \tilde{\tilde u}_{ijst}^b &= u_{ijst} - \bar u_{ijs.} - \bar u_{ij.t} - \bar u_{.jst} + \bar u_{ij..} + \bar u_{.js.} + \bar u_{.j.t} - \bar u_{.j..} \\ \tilde{\tilde u}_{ijst}^c &= u_{ijst} - \bar u_{ijs.} - \bar u_{ij.t} - \bar u_{i.st} + \bar u_{ij..} + \bar u_{i.s.} + \bar u_{i..t} - \bar u_{i...} \\ \tilde{\tilde u}_{ijst}^d &= u_{ijst} - \bar u_{ijs.} - \bar u_{i.st} - \bar u_{.jst} + \bar u_{i.s.} + \bar u_{.js.} + \bar u_{..st} - \bar u_{..s.}\,.\end{aligned}$$

Estimation of the variance components in the case of incomplete data yields

$$\begin{aligned}
\hat{\sigma}_{\mu}^2 &= \frac{1}{\sum_{ijs}|T_{ijs}|}\sum_{ijst}\hat{u}_{ijst}^2 - \frac{1}{\tilde{n}_{ijs}}\sum_{ijs}\frac{1}{|T_{ijs}|-1}\sum_{t}(\tilde{\hat{u}}_{ijst}^a)^2 \\
\hat{\sigma}_{\upsilon}^2 &= \frac{1}{\sum_{ijs}|T_{ijs}|}\sum_{ijst}\hat{u}_{ijst}^2 - \frac{1}{\tilde{n}_{ist}}\sum_{ist}\frac{1}{n_{ist}-1}\sum_{j}(\tilde{\hat{u}}_{ijst}^b)^2 \\
\hat{\sigma}_{\zeta}^2 &= \frac{1}{\sum_{ijs}|T_{ijs}|}\sum_{ijst}\hat{u}_{ijst}^2 - \frac{1}{\tilde{n}_{jst}}\sum_{jt}\frac{1}{n_{jst}-1}\sum_{i}(\tilde{\hat{u}}_{ijst}^c)^2 \\
\hat{\sigma}_{\lambda}^2 &= \frac{1}{\sum_{ijs}|T_{ijs}|}\sum_{ijst}\hat{u}_{ijst}^2 - \frac{1}{\tilde{n}_{ijt}}\sum_{ijt}\frac{1}{n_{ijt}-1}\sum_{s}(\tilde{\hat{u}}_{ijst}^d)^2 \\
\hat{\sigma}_{\varepsilon}^2 &= \frac{1}{\sum_{ijs}|T_{ijs}|}\sum_{ijst}\hat{u}_{ijst}^2 - \hat{\sigma}_{\mu}^2 - \hat{\sigma}_{\upsilon}^2 - \hat{\sigma}_{\zeta}^2 - \hat{\sigma}_{\lambda}^2 \,,
\end{aligned} \tag{2.46}$$

where $\hat{u}_{ijst}$ are the OLS residuals, and $\tilde{\hat{u}}_{ijst}^k$ are its transformations ($k = a,b,c,d$) according to

$$\begin{aligned}
\tilde{u}_{ijst}^a &= u_{ijst} - \frac{1}{|T_{ijs}|}\sum_t u_{ijst} \,, \quad \tilde{u}_{ijst}^b = u_{ijst} - \frac{1}{n_{ist}}\sum_j u_{ijst} \,, \\
\tilde{u}_{ijst}^c &= u_{ijst} - \frac{1}{n_{jst}}\sum_i u_{ijst} \,, \quad \tilde{u}_{ijst}^d = u_{ijst} - \frac{1}{n_{ijt}}\sum_s u_{ijst} \,.
\end{aligned}$$

Further, $|T_{ijs}|$, n_{ist}, n_{jst}, and 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.

Chapter 3
Models with Endogenous Regressors

Laszlo Balazsi, Maurice J.G. Bun, Felix Chan, and Mark N. Harris

Abstract This chapter examines various estimation and testing issues concerning models with endogenous regressors. The complexity of these issues increases as the number of potential unobserved heterogeneities increases with the dimension of the data. The chapter examines the properties of least squares type estimators, including the Within estimator, under different specifications of the error components and different correlation assumptions with the regressors. The latter induces different types of endogeneity not studied previously. In terms of estimation, the chapter includes an extension to the well-known Hausman-Taylor estimator for models with multiple dimensions. It also proposes a set of valid orthogonality conditions for purposes of implementing Generalised Method of Moments (GMM) estimators under these different specifications and endogeneity assumptions. The theoretical results in this chapter identify consistent and efficient estimators for different specifications. These results allow an extension of the Hausman specification test to detect endogeneity in multi-dimensional panel data models. Other issues, such as mixed effects models, self-flow, incomplete data and higher dimensional models, will also be discussed.

Laszlo Balazsi
Central European University, Budapest, Hungary

Maurice J.G. Bun
University of Amsterdam, Amsterdam, the Netherlands

Felix Chan
Curtin University, Perth, Western Australia

Mark N. Harris
Curtin University, Perth, Western Australia

L. Matyas (ed.), *The Econometrics of Multi-dimensional Panels*, Advanced Studies in Theoretical and Applied Econometrics 50, DOI 10.1007/978-3-319-60783-2_3

3.1 Introduction

For the ease of exposition, this chapter focuses on three dimensional panel data models. Extensions to the higher dimension will be discussed in some detail at the end of the chapter. Following Model (2.10) of Chap. 2,

$$y_{ijt} = \underline{x}'_{ijt}\beta + \pi_{ijt} + \varepsilon_{ijt} = \underline{x}'_{ijt}\beta + u_{ijt}, \tag{3.1}$$

with $i = 1 \ldots N_1$, $j = 1 \ldots N_2$ and $t = 1 \ldots T$, where y_{ijt} and $\underline{x}'_{ijt}$ are the dependent variables and the $(1 \times K)$ vector of regressors,[1] respectively. π_{ijt} is the unobserved heterogeneity which can correspond to any of the random effects formulations in (2.2), (2.5), (2.7), (2.9), (2.11), and (2.14) in Chap. 2, while ε_{ijt} is the idiosyncratic shock with $u_{ijt} = \pi_{ijt} + \varepsilon_{ijt}$. Model (3.1) can also be presented using matrix notation,

$$\underset{(N_1N_2T \times 1)}{y} = \underset{(N_1N_2T \times k)}{X} \underset{(k \times 1)}{\beta} + \underset{(N_1N_2T \times 1)}{\pi} + \underset{(N_1N_2T \times 1)}{\varepsilon}, \tag{3.2}$$

where π is the vector of all unobserved heterogeneous effects. The index t will uniquely define the time period; indexes i and j will typically define the units of observations, measured over time. In some instances, i and j may represent different roles of the same economic unit. For example, in a model of trade flows, i may represent the export nation with j denoting the import nation, *e.g.* export flows from the Netherlands to Hungary. In other instances they may be nested; for example, individual i in family unit j; or finally, no direct relationship: earnings of individual i in region j.

As usual, when the unobserved effects are correlated with the regressors, standard ordinary least squares (OLS) and generalised least squares (GLS) estimators are biased and inconsistent.[2] An alternative approach is to seek an appropriate transformation to eliminate the unobserved heterogeneity. As is shown in Sect. 1.5 of Chap. 1, with an M_D $(N_1N_2T \times N_1N_2T)$ projection matrix such that $M_D\pi = 0$, equation (3.2) can be transformed and yields

$$M_D y = M_D X\beta + M_D\varepsilon\,. \tag{3.3}$$

The absence of π means that the parameter vector β can be estimated consistently, if not efficiently, by OLS or GLS, in equation (3.3). This is the basis of Within and *Between* estimators, as well as other variants that combine these two estimators for

[1] Note that $\underline{x}_{ijt}$ denotes the usual vector of covariates, defined by x_{ijt} in Chaps. 1 and 2. As we form various partitions on this vector in the subsequent sections, we need to distinguish between regressors based on their fixedness. For example, x_{ijt} denotes the partition of $\underline{x}_{ijt}$, whose elements vary over all three indexes. Likewise, x_{ij} denotes the partition of $\underline{x}_{ijt}$ whose elements vary over indexes i and j but not t.

[2] In such cases $E(X'u) \neq 0$ and $E(X'\Omega^{-1}u) \neq 0$, with $\Omega = E(uu')$, and so both OLS and GLS estimators are biased. Similarly, depending on the source of endogeneity, $\text{plim}_{N_1\to\infty} X'u/N_1 \neq 0$ and $\text{plim}_{N_1\to\infty} X'\Omega^{-1}u/N_1 \neq 0$ circumvent the consistency of the OLS and GLS estimators (in this case endogeneity enters the model through i).

the "two-way" fixed effects specifications. While these 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, π, as well as some of the explanatory variables in $\underline{x}_{ijt}$. As a result, parameters associated with the time-invariant or individual-invariant variables cannot be estimated using these estimators. This may potentially be a very important issue, especially when one wishes to evaluate the impact of certain policies. The conventional approach of including dummy variables to represent policies in the model will no longer be effective as these variables are likely to be eliminated by the transformations. 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 loss 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. This is partially due to the fact that most multidimensional datasets have an overwhelmingly large number of observations over most, if not all, 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 is often the more serious issue. For example, standard Gravity models of trade, such as those considered in Harris et al. (2002) and Bun and Klaassen (2007), employ distance, the GDP of the export and the import countries as key regressors. Distance is clearly 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.

Therefore, the aim of this chapter is to examine the appropriate methods by which to estimate the parameters in a linear multi-dimensional panel data model under different assumptions of endogeneity, where the unobserved heterogeneity terms are treated as random, not fixed (as in Chap. 2), for the reasons outlined above. We first revisit the seminal results of Hausman and Taylor (1981) (hereafter HT) in Sect. 3.2 and extend their approach to multi-dimensional panel data models. The results indicate that the HT approach can be applied to any three-dimensional model in principle but the transformations required to generate the set of internal instruments are not trivial. In addition, the order conditions to ensure parameter identification are sometimes too restrictive to be practical. As a viable alternative, this chapter also presents a non-linear GMM approach in Sect. 3.3. This approach allows a more flexible correlation structure between the regressors and the unobserved heterogeneity, but requires slightly stronger assumptions on the correlation structures between the unobserved heterogeneities. Sect. 3.4 briefly discusses estimation issues with mixed effects models, that is, models with both fixed and random effects. In order

to assess the validity of the proposed instruments as well as the estimators, Sect. 3.5 constructs various parameter tests, while Sect. 3.6 extends the results intuitively to incomplete data and to higher-dimensional models.

Throughout the chapter, we adopt the standard ANOVA notation similarly to the rest of the book. Specifically, I_N and ι_N denote the $(N \times N)$ identity matrix and the $(N \times 1)$ vector of ones, respectively. The subscript may be omitted if the dimension is clear from the context. $J_N = \iota_N \iota_N'$, $\bar{J}_N = J_N/N$ and $Q_N = I_N - \bar{J}_N$, so that $Q_N y$ yields the "de-meaned" version of y. M_A projects A into its null-space, namely, $M_A = I - A(A'A)^{-1}A'$, while P_A projects A into its column space, specifically, $P_A = I - M_A = A(A'A)^{-1}A'$, and finally, $\bar{x}_{.jt}$ denotes the average of x over the index i while the definition extends naturally to other quantities such as $\bar{x}_{i.t}$, $\bar{x}_{ij.}$, $\bar{x}_{..t}$ and $\bar{x}_{...}$.

3.2 The Hausman-Taylor-like Instrument Variable Estimator

Clearly the parameters in model (3.1)-(3.2) cannot be identified in their current form without making some further assumptions regarding π_{ijt}. Consider firstly model (2.11) in Chap. 2,

$$y_{ijt} = \underline{x}_{ijt}'\beta + \upsilon_i + \zeta_j + \lambda_t + \varepsilon_{ijt} \,, \tag{3.4}$$

which is clearly a special case of model (3.1) with $\pi_{ijt} = \upsilon_i + \zeta_j + \lambda_t$ and this will serve as our starting point. Indeed, this specification spawned much of the subsequent research in the empirical literature, see for example, Ghosh (1976) and Matyas (1997).

3.2.1 A Simple Approach

Note that when $\pi_{ijt} = \upsilon_i + \zeta_j + \lambda_t$, it is possible to rewrite this as a standard two-dimensional panel data model by grouping all the individual indexes. Specifically, let $\xi_{ij} = \upsilon_i + \zeta_j$ then $\pi_{st} = \xi_s + \lambda_t$ with each s corresponding to each (ij) pair. Thus, Model (3.4) can be rewritten as

$$y_{st} = \underline{x}_{st}'\beta + \xi_s + \lambda_t + \varepsilon_{st} \,, \tag{3.5}$$

which is a standard two-dimensional panel data model with two-way error components. Wyhowski (1994) proposed consistent estimators for β in equation (3.5). While this provides a simple approach to estimating the parameter for a multidimensional panel data model, it does not utilise the additional information provided by such data. From a practical viewpoint, the number of instruments is also limited under such an approach. This has practical implications as the relatively small number of instruments may potentially violate the order condition which is necessary for the approach proposed in Wyhowski (1994). In the event that the order condition is not satisfied under this simple approach, we need to derive a larger instrument set

by utilising the additional dimension(s) available and the specification of the error components. We focus on this issue in the next subsection.

3.2.2 Sources of Endogeneity

In addition to the specification as stated in equation (3.4), we consider the random effects in π_{ijt} to be well-behaved, that is assumption (2.12) holds, and that all have zero means, finite variances and are pairwise uncorrelated. However, we do not impose any correlation restriction between the regressors and the unobserved heterogeneity. Therefore, the following results hold under a wide range of correlation structures between the regressors and the 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. Variables such as x'_i, x'_j, or x'_t are eliminated by the Within transformations, therefore their associated parameters cannot be identified with the Within estimators.

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 summarised in Table 3.1 and the subsequent analysis does not explicitly impose any further assumptions on the correlations between the regressors and the unobserved heterogeneities. Multicollinearity does not generally violate the feasibility and consistency of the estimators with the exception of perfect collinearity, a case which will be excluded from the rank condition.

Following Hausman and Taylor (1981) 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 by the usual 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 following subsections discuss various implementations of this general idea.

Table 3.1 Sources of endogeneity on the level of partitions of the regressors for model (3.4)

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}			
$\upsilon_i, \zeta_j, \lambda_t$	x'_{8ijt}			

3.2.3 The Hausman-Taylor Estimator

Using matrix notation for the partitions of the regressors, we define $X^{(1)}$ as the stacked matrix version of $\vec{x}'_{ijt}$, 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 an instrument for $X^{(1)}$ in order to obtain a 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 is applied to equation (3.4). 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 find instruments for 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

$$H_2 = (Q_{N_1} \otimes \bar{J}_{N_2 T}) \cdot \left(X_1^{(1)}, X_3^{(1)}, X_4^{(1)}, X_7^{(1)}, X_1^{(2)}\right),$$

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 variables included. That is, it only contains variation over i and therefore, this transformation 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 variable set uncorrelated with λ_t is

$$H_4 = (\bar{J}_{N_1 N_2} \otimes Q_T) \cdot \left(X_1^{(1)}, X_2^{(1)}, X_3^{(1)}, X_4^{(1)}, X_1^{(4)} \right).$$

In other words, we have, by construction

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

While the construction of transformation matrices, such as $(Q_{N_1} \otimes \bar{J}_{N_2} \otimes \bar{J}_T)$, may seem memory demanding even on powerful personal computers when the dataset is large, these transformations represent relatively simple operations on the data which can be carried out sequentially. This makes the transformations computationally feasible in practice. Specifically, these transformations define operations on simple group means and the deviations from these group means. Table 3.2 provides a list of matrix transformations with their equivalent scalar operations on each observation.

Table 3.2 Translation of matrix operations into scalar

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

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

3.2.3.1 Extending the Hausman-Taylor Two-Stage Least Squares Estimator

Performing two-stage least squares (2SLS) on model (3.4) with instrument set H is identical to estimating

$$P_H y = P_H X\beta + P_H u \tag{3.6}$$

with Least Squares where P_H is the projection matrix, $H(H'H)^{-1}H'$. Define

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

and note that P_H is symmetric idempotent given the orthogonality nature of H_p. The estimator is then defined as

$$\hat{\beta}_{HT1} = (X'P_H X)^{-1}X'P_H y. \tag{3.7}$$

Note that the H_p matrices can be constructed by calculating simple group means and deviations from group means from the original data matrix. The elements of P_{H_p} are therefore straightforward to calculate. Hence, the estimator as defined in equation (3.7) is computationally simple without excessive burden on the memory or computation requirement.

3.2.3.2 The More Efficient Hausman-Taylor Estimator

The asymptotic efficiency of the estimator as defined in equation (3.7) can be improved if we exploit the error component structure. As shown in Fuller and Battese (1973) and following the same arguments as those in Chap. 2 (pp. 6-7), pre-multiply model (3.2) by $\Omega^{-1/2}$ gives

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

where $\Omega = E(uu')$, then the same estimator as defined in equation (3.7) is asymptotically efficient. Alternatively, following the arguments in Maddala (1971), this 2SLS can be interpreted 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:[3]

$$\sigma_1^2 = \sigma_\varepsilon^2\,, \quad \sigma_2^2 = \sigma_\varepsilon^2 + N_2 T \sigma_\upsilon^2\,, \quad \sigma_3^2 = \sigma_\varepsilon^2 + N_1 T \sigma_\zeta^2\,, \quad \text{and} \quad \sigma_4^2 = \sigma_\varepsilon^2 + N_1 N_2 \sigma_\lambda^2\,.$$

The more efficient estimator can then be computed as

$$\hat{\beta}_{HT2} = (X' P_H^* X)^{-1} X' P_H^* y\,. \tag{3.8}$$

In practice, these variances are unknown. Specifically, the components of the variance-covariance matrix Ω are usually unknown and must be estimated in order to implement (3.8). 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 estimate 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

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

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 the residuals are given by

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

which can be used to estimate the variance components as

$$\begin{aligned} \hat{\sigma}_\zeta^2 &= \hat{u}_2' \hat{u}_2 / (N_1 N_2 T) - \tfrac{1}{N_1 T} \hat{\sigma}_\varepsilon^2 \\ \hat{\sigma}_\lambda^2 &= \hat{u}_3' \hat{u}_3 / (N_1 N_2 T) - \tfrac{1}{N_1 N_2} \hat{\sigma}_\varepsilon^2\,. \end{aligned}$$

[3] The result $P_H \Omega^{-1/2} = P_H^*$ implied by the definition of the Q and $\bar{J}$ matrices.

Note that for the consistent estimation of σ_υ^2, 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. Similarly to Hausman and Taylor (1981), there is also a set of necessary order conditions that is easier to verify. Specifically,

$$\begin{aligned} k_1^{(1)} + k_3^{(1)} + k_4^{(1)} + k_7^{(1)} &\geq k_2^{(2)} \\ k_1^{(1)} + k_2^{(1)} + k_4^{(1)} + k_6^{(1)} &\geq k_2^{(3)} \\ k_1^{(1)} + k_2^{(1)} + k_3^{(1)} + k_5^{(1)} &\geq k_2^{(4)} \end{aligned} \tag{3.9}$$

have to be satisfied jointly. Although it seems 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 Hausman and Taylor (1981) in the case of standard two-dimensional panels. 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. Let s denotes the joined index of j and t, 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 are uncorrelated with υ_i. If we denote the total number of their columns k_1, the first order condition in (3.9) simplifies to

$$k_1 \geq k_2^{(5)}\,. \tag{3.10}$$

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 (3.10), 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 Hausman and Taylor (1981).

We can also relate our order conditions to those of the two-way panel models studied 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).

3.2.4 Time Varying Individual Specific Effects

Model (3.4) assumes the individual specific effects are time invariant. Given the multi-dimensional nature of the data, it is possible to incorporate time varying individual effects. Specifically, consider model (2.2) from Chap. 2, which was referred

to as the "all-encompassing model",

$$y_{ijt} = \underline{x}'_{ijt}\beta + \mu_{ij} + \upsilon_{it} + \zeta_{jt} + \varepsilon_{ijt}\,, \tag{3.11}$$

where the unobserved heterogeneities, μ_{ij}, υ_{it}, ζ_{jt}, and the idiosyncratic disturbance term, ε_{ijt}, are assumed to be pairwise uncorrelated, with zero means and finite variances. Recall in the case of model (3.4), 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 the case with model (3.11). 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 also be unidentified.

Following the same approach as in the previous subsection, Table 3.3 shows the partitions of the regressor vector based on the sources of endogeneity:

Table 3.3 Sources of endogeneity on the level of partitions of the regressors for model (3.11)

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, group 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 3.1.

Allowing individual specific effects to be time varying leads to some additional complications in implementing HT-type estimators. 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} signifies 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 equation (3.11) and is summarised in Table 3.4.

While the general HT approach is still theoretically sound in this setting, the time varying nature of individual specific effects imposes additional restrictions on the

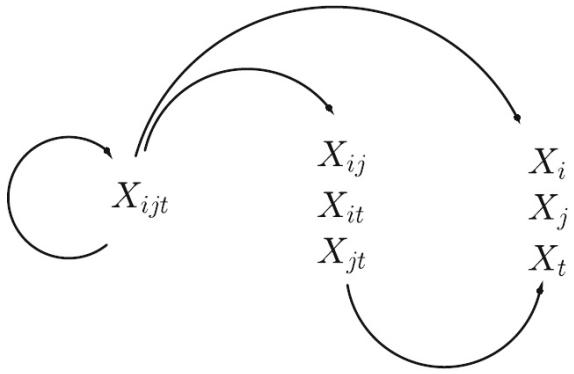

Fig. 3.1 Possible instrumental variables

Table 3.4 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}	$x'_{4t})$

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 3.3 where $k_l^{(m)}$ 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 3.5 presents the instruments, H_p, the associated transforms, R_p, and the original variable set, X_p.

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.

Table 3.5 Proposed instruments H_p for each endogenous variable

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

Note: 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, and 16.

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

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}:

$$\hat{\beta}_{HT2} = (X' P_H^* X)^{-1} X' P_H^* y, \tag{3.12}$$

with

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

where

$$
\begin{aligned}
\sigma_1^2 &= \sigma_\varepsilon^2 \\
\sigma_p^2 &= \sigma_\varepsilon^2 + T\sigma_\mu^2 \qquad p = 2,\ldots,5 \\
\sigma_p^2 &= \sigma_\varepsilon^2 + N_2\sigma_\upsilon^2 \qquad p = 6,\ldots,9 \\
\sigma_p^2 &= \sigma_\varepsilon^2 + N_1\sigma_\zeta^2 \qquad p = 10,\ldots,13 \\
\sigma_{14}^2 &= \sigma_\varepsilon^2 + N_2\sigma_\upsilon^2 + T\sigma_\mu^2 \\
\sigma_{15}^2 &= \sigma_\varepsilon^2 + N_1\sigma_\zeta^2 + T\sigma_\mu^2 \\
\sigma_{16}^2 &= \sigma_\varepsilon^2 + N_1\sigma_\zeta^2 + N_2\sigma_\upsilon^2.
\end{aligned}
\tag{3.13}
$$

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 of 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 also *jointly*. Table 3.6 organizes these conditions, which all have to be satisfied in order to have as many instruments as endogenous variables.

Table 3.6 Order conditions for model (3.11)

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

The complexity of these necessary order conditions means that the general HT approach may not be practical in higher dimensions. This provides a motivation to consider other estimation strategies, such as the non-linear GMM estimator, which

will be discussed in Sect. 3.3. Nevertheless, we present several statistical properties of the HT estimator for Model (3.11) in the following subsection.

3.2.5 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 to ensure the validity of all the instrument sets. For both model (3.4) and (3.11), the $\hat{\beta}_{HT1}$ and $\hat{\beta}_{HT2}$ are consistent only when all $N_1, N_2, T \to \infty$ jointly. This is perhaps unsurprising, 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. The main asymptotic result is presented in the following proposition:

Proposition 1. *Consider model (3.11) with $\hat{\beta}_{HT2}$ defined as (3.12), then under assumptions (1) - (15) in the Appendix*

$$S(\hat{\beta}_{HT2} - \beta) \to^d N(0, \Gamma) \text{ as } N_1, N_2, T \to \infty,$$

where

$$\begin{aligned} S = diag\big\{ &\sqrt{N_1 N_2 T} \cdot I_{k^{(1)}},\ \sqrt{N_1 N_2} \cdot I_{k^{(2)}},\ \sqrt{N_1 T} \cdot I_{k^{(3)}},\ \sqrt{N_2 T} \cdot I_{k^{(4)}},\ \times \\ &\times \sqrt{N_1 N_2} \cdot I_{k^{(5)}},\ \sqrt{N_1 T} \cdot I_{k^{(6)}},\ \sqrt{N_2 T} \cdot I_{k^{(7)}} \big\} \end{aligned}$$

and

$$\Gamma = diag\left\{ 1/\sigma_\varepsilon^2 \Gamma_1,\ 1/\sigma_\mu^2 \Gamma_2,\ 1/\sigma_\upsilon^2 \Gamma_3,\ 1/\sigma_\xi^2 \Gamma_4,\ 1/\sigma_\mu^2 \Gamma_5,\ 1/\sigma_\upsilon^2 \Gamma_6,\ 1/\sigma_\xi^2 \Gamma_7 \right\}$$

with

$$\begin{aligned} \Gamma_1 &= diag\{V_1, 0_{k^{(2)}+\ldots+k^{(7)}}\} \\ \Gamma_2 &= diag\{0_{k^{(1)}}, V_{2,22}, 0_{k^{(3)}+\ldots+k^{(7)}}\} \\ \Gamma_3 &= diag\{0_{k^{(1)}+k^{(2)}}, V_{3,22}, 0_{k^{(4)}+\ldots+k^{(7)}}\} \\ \Gamma_4 &= diag\{0_{k^{(1)}+\ldots+k^{(3)}}, V_{4,22}, 0_{k^{(5)}+\ldots+k^{(7)}}\} \\ \Gamma_5 &= diag\{0_{k^{(1)}}, V_{5,22}, 0_{k^{(3)}+k^{(4)}}, V_{5,44}, 0_{k^{(6)}+k^{(7)}}\} \\ \Gamma_6 &= diag\{0_{k^{(1)}+k^{(2)}}, V_{6,33}, 0_{k^{(4)}+k^{(5)}}, V_{6,44}, 0_{k^{(7)}}\} \\ \Gamma_7 &= diag\{0_{k^{(1)}+k^{(2)}+k^{(3)}}, V_{7,33}, 0_{k^{(5)}+k^{(6)}}, V_{7,44}\}. \end{aligned}$$

$V_{p,rr}$ denotes the sub-matrix rr of V_p.

Proof. See the Appendix.

Corollary 1. *Consider model (3.4) with $\hat{\beta}_{HT2}$ defined as equation (3.8), under assumptions (A) - (I) in the Appendix,*

$$S(\hat{\beta}_{HT2} - \beta) \to^d N(0, \Gamma^{-1}) \; as \; N_1, N_2, T \to \infty,$$

where

$$S = diag\left\{\sqrt{N_1 N_2 T} \cdot I_{k^{(1)}},\ \sqrt{N_1} \cdot I_{k^{(2)}},\ \sqrt{N_2} \cdot I_{k^{(3)}},\ \sqrt{T} \cdot I_{k^{(4)}}\right\}$$

and

$$\Gamma = diag\left\{1/\sigma_\varepsilon^2 \Gamma_1,\ 1/\sigma_\upsilon^2 \Gamma_2,\ 1/\sigma_\xi^2 \Gamma_3,\ 1/\sigma_\lambda^2 \Gamma_4\right\}.$$

with

$$\begin{aligned}
\Gamma_1 &= diag\{V_1, 0_{k^{(2)}+k^{(3)}+k^{(4)}}\} \\
\Gamma_2 &= diag\{0_{k^{(1)}}, V_{2,22}, 0_{k^{(3)}+k^{(4)}}\} \\
\Gamma_3 &= diag\{0_{k^{(1)}+k^{(2)}}, V_{3,22}, 0_{k^{(4)}}\} \\
\Gamma_4 &= diag\{0_{k^{(1)}+k^{(2)}+k^{(3)}}, V_{4,22}\}.
\end{aligned}$$

Proof. See the Appendix.

3.2.6 Using External Instruments

The discussion has 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 (3.4) with the *a priori* knowledge that x'_{ijt}, or any of its transformed counterparts, are not correlated with x'_i. Clearly, we can still use the Within transform of $\bar{x}'_{ijt}$ 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 can be fixed over j and/or t, such that

$$\text{Corr}\left(z'_{ijt}, x'_i\right) \neq 0 \quad \text{and} \quad \text{Corr}\left(z'_{i\cdot\cdot} - z'_{\cdot\cdot\cdot}, u_{ijt}\right) = 0\,. \tag{3.14}$$

Note that for the second condition in (3.14) to hold, we only require $\text{Corr}(z'_{ijt}, \upsilon_i) = 0$. Once we obtain z'_{ijt}, the instrument is constructed as in Sect. 3.2.3, where z'_{ijt} is used as instruments instead of $\bar{x}'_{ijt}$ 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 of endogenous regressors in x'_i, which coincides with the standard result of identifiability in the instrumental variable literature.

3.3 The Non-linear Generalized Method of Moments Estimator

Section 3.2 discussed how the HT-type estimators can be generalised to consistently estimate three-way panel data models with endogenous regressors. It showed that the complexity of the order conditions increase as we allow for more flexible error component structures. This imposes practical difficulties in implementing such estimators. An alternative is to consider a non-linear GMM estimator, which exploits the structure of the composite disturbance term and the assumptions regarding the sources of endogeneity. It seeks to utilise a set of orthogonality conditions to construct a consistent estimator.

Chan et al. (2016) proposed a GMM estimator, similar to Ahn and Schmidt (1999), for model (3.11) on incomplete data. Although they only assumed that $\mathrm{E}(x'_{it}\upsilon_{it}) \neq 0$ and $\mathrm{E}(x'_{jt}\zeta_{jt}) \neq 0$, with the rest of the variables assumed to be exogenous, their methodology can be readily extended to more general cases.

In particular, from the error component structure of model (3.11), $u_{ijt} = \mu_{ij} + \upsilon_{it} + \zeta_{jt}$, and from the underlying assumptions about the random effects, it is straightforward to show the following. Define $u_{jt} = (u_{1jt}, \ldots, u_{N_1 jt})'$ and $u_{it} = (u_{i1t}, \ldots, u_{iN_2 t})'$ as $(N_1 \times 1)$ and $(N_2 \times 1)$ vectors, respectively

$$
\begin{array}{lll}
\mathrm{E}\left(u_{it}u'_{it} - u_{kt}u'_{kt}\right) & = 0 & \text{for all } i,k = 1\ldots N_1,\ k \neq i,\ t = 1\ldots T \\
\mathrm{E}\left(u_{jt}u'_{jt} - u_{lt}u'_{lt}\right) & = 0 & \text{for all } j,l = 1\ldots N_2,\ k \neq i,\ t = 1\ldots T \\
\mathrm{E}\left(u_{it}u'_{it-1} - u_{kt}u'_{kt-1}\right) & = 0 & \text{for all } i,k = 1\ldots N_1,\ k \neq i,\ t = 2\ldots T \\
\mathrm{E}\left(u_{jt}u'_{jt-1} - u_{lt}u'_{lt-1}\right) & = 0 & \text{for all } j,l = 1\ldots N_2,\ k \neq i,\ t = 2\ldots T \\
\mathrm{E}\left(u_{it}u'_{kt-1}\right) & = 0 & \text{for all } i,k = 1\ldots N_1,\ k \neq i,\ t = 1\ldots T \\
\mathrm{E}\left(u_{jt}u'_{lt-1}\right) & = 0 & \text{for all } j,l = 1\ldots N_2,\ l \neq j,\ t = 1\ldots T
\end{array}
\tag{3.15}
$$

Note that $k \neq i$ and $l \neq j$ are required to avoid the degenerated case, as for $k = i$ and $l = j$ the orthogonality condition becomes an identity. Equations (3.15) lead to a maximum number of $(N_1 - 1)(2N_1 - 1) + (N_2 - 1)(2N_2 - 1)$ moment conditions. Define $g(\beta, X, y)$ as the vector of orthogonality conditions in (3.15), the GMM estimator is constructed as

$$
\hat{\beta}_{GMM} = \arg\min_{\beta} g'(\beta, X, y)\Sigma^{-1} g(\beta, X, y)\,, \tag{3.16}
$$

where Σ denotes the optimal weight matrix. The common practice to obtain the efficient GMM estimator is to first minimise (3.16) with $\Sigma = I$ to β^*, then estimate Σ with

$$
\hat{\Sigma}(\beta^*) = \frac{1}{T}\sum_{t=1}^{T} g(\beta^*, X_t, y_t) g'(\beta^*, X_t, y_t)\,, \tag{3.17}
$$

with $X_t = (x_{11t}, \ldots, x_{N_1 N_2 t})'$ which definition extends naturally to y_t. Using the estimated weight matrix, the efficient GMM is calculated from the functional form

$$
\hat{\beta}_{GMM,eff} = \arg\min_{\beta} g'(\beta, X, y)\hat{\Sigma}^{-1}(\beta^*) g(\beta, X, y)\,. \tag{3.18}
$$

Chan et al. (2016) also showed that under the assumptions of specific sources of endogeneity and the error component structure, $\hat{\beta}_{GMM}$ and $\hat{\beta}_{GMM,eff}$ are consistent as $T \to \infty$, and (from the central limit theorem), $\sqrt{T}(\hat{\beta}_{GMM,eff} - \beta) \to^d N(0,V)$, where

$$V = \mathbb{E}\left(\frac{\partial g'(\beta,X,y)}{\partial \beta} \left(g(\beta,X,y) g'(\beta,X,y) \right)^{-1} \frac{\partial g(\beta,X,y)}{\partial \beta} \right)^{-1} \Big|_{\beta=\hat{\beta}_{GMM,eff}} .$$

As model (3.4) is nested in model (3.11), the GMM estimator (3.18) with the same set of underlying moment conditions can also be applied.

3.4 Mixed Effects Models

In Sects. 3.2 and 3.3, various estimators of pure random effects models were proposed in the presence of endogeneity. Clearly such linear dependencies between the effects and the observables do not matter in the case of fixed effects, as the orthogonal transformations remove the endogeneity bias by eliminating the fixed effects. Sometimes, however, the application is such that it requires "mixed effects models", models of both fixed and random effects. A brief discussion on the estimation of such models can be found in Sect. 2.5.2. but the discussion did not consider any endogeneity bias.

The solution here is to first eliminate the fixed effects from any mixed effects model by applying the appropriate transformation. This is followed by finding viable instruments for the endogenous (transformed) variables. As we will see, the task is fundamentally similar to the original endogeneity problem discussed in Sect. 3.2.

Reconsider model (3.4),

$$y_{ijt} = \underline{x}'_{ijt}\beta + \upsilon_i + \zeta_j + \lambda_t + \varepsilon_{ijt} , \tag{3.19}$$

with the sole difference that λ_t is assumed to be a fixed parameter, rather than a random variable. To eliminate λ_t apply the transformation $\tilde{y}_{ijt} = y_{ijt} - \bar{y}_{..t}$, represented by the matrix $M = I - (\bar{J}_{N_1 N_2} \otimes I_T)$. This gives

$$\tilde{y}_{ijt} = \underline{\tilde{x}}'_{ijt}\beta + \tilde{\upsilon}_i + \tilde{\zeta}_j + \tilde{\varepsilon}_{ijt} ,$$

which is a pure random effects model with de-meaned random effects and transformed variables. To identify the sources of endogeneity in this model, we follow the same approach as before by forming partitions of the regressors as presented in Table 3.7. Let $\underline{\tilde{x}}_{ijt} = (\tilde{\tilde{x}}_{ijt}, \tilde{x}_{ij}, \tilde{x}_{it}, \tilde{x}_{jt})$ with $\tilde{\tilde{x}}_{ijt} = (\tilde{x}_{1ijt}, \tilde{x}_{2ijt}, \tilde{x}_{3ijt}, \tilde{x}_{4ijt})$.

All regressors that vary over the time index, t, but fixed over i and j have been removed along with the fixed effect, λ_t. Therefore, we can only identify the partitions that are correlated with $\tilde{\upsilon}_i$ and $\tilde{\zeta}_j$.

From here, the approach follows the same arguments in Sect. 3.2.3: endogenous partitions in $\tilde{x}'_{ijt}$ are instrumented from the Within estimator, while $\tilde{x}'_{2i}$ is instru-

Table 3.7 Sources of endogeneity at the level of partitions of the regressors for model (3.19)

Correlated with	Partition		
None	$\tilde{x}'_{1ijt}$	$\tilde{x}'_{1i}$	$\tilde{x}'_{1j}$
$\tilde{\upsilon}_i$	$\tilde{x}'_{2ijt}$	$\tilde{x}'_{2i}$	
$\tilde{\zeta}_j$	$\tilde{x}'_{3ijt}$		$\tilde{x}'_{2j}$
$\tilde{\upsilon}_i$, $\tilde{\zeta}_j$	$\tilde{x}'_{4ijt}$		

mented with partitions of $\tilde{x}'_{ijt}$ that are uncorrelated with υ_i, namely, $\tilde{x}'_{1ijt}$ and $\tilde{x}'_{3ijt}$. $\tilde{x}'_{2j}$ is instrumented with partitions of $\tilde{x}'_{ijt}$ that are uncorrelated with ζ_j, namely, $\tilde{x}'_{1ijt}$ and $\tilde{x}'_{2ijt}$. The instruments have to be transformed to eliminate the remaining correlation with $\tilde{\zeta}_j$ in the case of $(\tilde{x}'_{1ijt}, \tilde{x}'_{3ijt})$. The same is also true for $\tilde{\upsilon}_i$ in the case of $(\tilde{x}'_{1ijt}, \tilde{x}'_{2ijt})$.

Let $\tilde{X}_l^{(1)}$ be the matrix stacked version of $\tilde{x}'_{lijt}$, with $\tilde{X}_l^{(2)}$ and $\tilde{X}_l^{(3)}$ defined similarly for $\tilde{x}'_{li}$ and $\tilde{x}'_{lj}$. Define $\tilde{X} = MX$, the instruments in matrix form can be written as

$$\begin{aligned} H_1 &= (I - I_{N_1} \otimes \bar{J}_{N_2T} - \bar{J}_{N_1} \otimes I_{N_2} \otimes \bar{J}_T + \bar{J}_{N_1N_2T})MX^{(1)} \\ &= (I - I_{N_1} \otimes \bar{J}_{N_2T} - \bar{J}_{N_1} \otimes I_{N_2} \otimes \bar{J}_T - \bar{J}_{N_1N_2} \otimes I_T + 2\bar{J}_{N_1N_2T})X^{(1)} \end{aligned}$$

for $X^{(1)}$,

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

for $X^{(2)}$, and finally,

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

for $X^{(3)}$. Interestingly, the transformed mixed model (3.19) and the non-transformed random effects model (3.4) have the same instrument set.[4] Combining the instruments in the usual way, two estimators emerge, the latter being more efficient than the former:

$$\hat{\beta}_{HT1} = (\tilde{X}'P_H\tilde{X})^{-1}\tilde{X}'P_H\tilde{y}; \qquad \hat{\beta}_{HT2} = (\tilde{X}'P_H^*\tilde{X})^{-1}\tilde{X}'P_H^*\tilde{y}\,,$$

where

$$P_H = \sum_{p=1}^{3} P_{H_p}\,, \quad \text{and} \quad P_H^* = \sum_{p=1}^{3} \frac{1}{\sigma_p} P_{H_p}\,,$$

with

[4] Naturally, with the difference that x_t'-type regressors are missing from the former and so do not need to be instrumented.

$$\sigma_1^2 = \sigma_\varepsilon^2 , \quad \sigma_2^2 = \sigma_\varepsilon^2 + N_2 T \sigma_\upsilon^2 , \quad \text{and} \quad \sigma_3^2 = \sigma_\varepsilon^2 + N_1 T \sigma_\zeta^2 .$$

Different mixed effects models can be treated similarly. First, the fixed effects should be eliminated with some M orthogonal projection, then the endogenous regressors should be instrumented with the same instruments as in the pure random effects case. In our previous example, endogenous regressors in the time dimension are removed, which leads to fewer order conditions that need to be satisfied for the instruments. This, however, comes at a price. That is, we are unable to reach full parameter identification under the mixed effects specification because some of the variables may be eliminated by the orthogonal project in the first step. In this sense, mixed effects models stand somewhere between fixed and random effects models. Fewer instruments are required yet some parameters remain unidentified.

3.5 Exogeneity Tests

The discussion thus far focuses on estimation issues in the presence of endogeneity. In practice, the presence of endogeneity is often unclear. This section presents a test of endogeneity under model (3.4), as well as a test for instrument validity.

3.5.1 Testing for Endogeneity

While the presence of endogeneity can often be argued on 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 a standard two-dimensional panel data model, a simple Hausman test (see Hausman, 1978 and further Sect. 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

$$H_0 : \mathrm{E}(\upsilon_i + \zeta_j + \lambda_t | \underline{\mathrm{x}}'_{ijt}) = 0 \quad \text{against} \quad H_1 : \mathrm{E}(\upsilon_i + \zeta_j + \lambda_t | \underline{\mathrm{x}}'_{ijt}) \neq 0 . \tag{3.20}$$

The GLS estimator under model (3.4) 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 $\hat{q}_1 = \hat{\beta}^{(1)}_{GLS} - \hat{\beta}^{(1)}_{Within}$,

$$
\begin{aligned}
m &= \hat{q}_1' \mathrm{Var}(\hat{q}_1)^{-1} \hat{q}_1 \\
&= \left(\hat{\beta}^{(1)}_{GLS} - \hat{\beta}^{(1)}_{Within}\right)' \left(\mathrm{Var}(\hat{\beta}^{(1)}_{GLS}) - \mathrm{Var}(\hat{\beta}^{(1)}_{Within})\right)^{-1} \left(\hat{\beta}^{(1)}_{GLS} - \hat{\beta}^{(1)}_{Within}\right) .
\end{aligned} \tag{3.21}
$$

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 $\hat{q}_1 = \text{plim}\, \hat{\beta}^{(1)}_{GLS} - \beta \neq 0$ under the alternative. Therefore, if $\hat{q}_1$ is sufficiently far from 0, then there is 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 (3.20) 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

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

we can test

$$
H_0 : \mathrm{E}(\tilde{\upsilon}_i|\tilde{\underline{\mathrm{x}}}'_{ijt}) = 0 \quad \text{against} \quad H_1 : \mathrm{E}(\tilde{\upsilon}_i|\tilde{\underline{\mathrm{x}}}'_{ijt}) \neq 0 , \tag{3.22}
$$

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

$$
\begin{aligned}
\tilde{y}_{ijt} &= \tilde{\underline{\mathrm{x}}}'_{ijt}\beta + \tilde{\zeta}_j + \tilde{\varepsilon}_{ijt} \quad \text{with} \quad \tilde{\underline{\mathrm{x}}}'_{ijt} = (\underline{\mathrm{x}}'_{ijt} - \underline{\bar{\mathrm{x}}}'_{i..} - \underline{\bar{\mathrm{x}}}'_{..t} + \underline{\bar{\mathrm{x}}}'_{...}) \\
\tilde{y}_{ijt} &= \tilde{\underline{\mathrm{x}}}'_{ijt}\beta + \tilde{\lambda}_t + \tilde{\varepsilon}_{ijt} \quad \text{with} \quad \tilde{\underline{\mathrm{x}}}'_{ijt} = (\underline{\mathrm{x}}'_{ijt} - \underline{\bar{\mathrm{x}}}'_{i..} - \underline{\bar{\mathrm{x}}}'_{.j.} + \underline{\bar{\mathrm{x}}}'_{...})
\end{aligned}
$$

by testing

$$
\begin{aligned}
H_0 &: \mathrm{E}(\tilde{\zeta}_j|\tilde{\underline{\mathrm{x}}}'_{ijt}) = 0 \quad \text{against} \quad H_1 : \mathrm{E}(\tilde{\zeta}_j|\tilde{\underline{\mathrm{x}}}'_{ijt}) \neq 0 \\
H_0 &: \mathrm{E}(\tilde{\lambda}_t|\tilde{\underline{\mathrm{x}}}'_{ijt}) = 0 \quad \text{against} \quad H_1 : \mathrm{E}(\tilde{\lambda}_t|\tilde{\underline{\mathrm{x}}}'_{ijt}) \neq 0 ,
\end{aligned} \tag{3.23}
$$

respectively. All three test statistics can be constructed similarly to equation (3.21) 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 more efficiently obtain a set of valid instruments.

3.5.2 Testing for Instrument Validity

The discussion so far implicitly assumed the existence of valid instruments, specifically,

$$\operatorname*{plim}_{N_1,N_2,T\to\infty} \frac{H'u}{N_1N_2T} = 0.$$

Typically economic rationale is 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 estimator. This is essentially another form of the Hausman test as proposed in Hausman (1978), where we form our null and alternative as

$$H_0 : \operatorname*{plim}_{N_1,N_2,T\to\infty} \frac{H'u}{N_1N_2T} = 0 \quad \text{against} \quad H_1 : \operatorname*{plim}_{N_1,N_2,T\to\infty} \frac{H'u}{N_1N_2T} \neq 0 \tag{3.24}$$

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 $\hat{q}_2 = \hat{\beta}_{HT2} - \hat{\beta}_{Within}$,

$$m = \hat{q}_2' \mathrm{Var}(\hat{q}_2)^{-1} \hat{q}_2$$

can be constructed, and be shown to have a χ^2_d distribution with $d = \mathrm{rank}(\mathrm{Var}(\hat{q}_2))$. We can elaborate on $\mathrm{Var}(\hat{q}^2)$ and find that

$$\begin{aligned} \mathrm{Var}(\hat{q}_2) = \mathrm{Var}(\hat{\beta}_{HT2} - \hat{\beta}_{Within}) &= \mathrm{Var}(\hat{\beta}_{HT2}) - \mathrm{Var}(\hat{\beta}_{Within}) \\ &= (X'P_H\hat{\Omega}^{-1}P_HX)^{-1} - (\tilde{X}'\tilde{X})^{-1}\hat{\sigma}^2_\varepsilon . \end{aligned}$$

Intuitively, if $\hat{q}_2$ deviates from zero it raises concerns about the validity of the instruments. If the null is rejected, it not only suggests that some variables failed as instruments, but also implies that these variables required instruments from other variables. In other words, these variables are themselves endogenous.

3.5.3 Testing in the Case of Fixed Effects

Consider the case when the individual effects υ_i, ζ_j and λ_t are represented by fixed estimable parameters, and after projectiong the effects out with some orthogonal transformation, the endogeneity issue still persists. Two scenarios are considered here: endogeneity due to not using the proper model specification, and endogeneity arising from omitted variables.

3.5.3.1 Improper Model Specifications

Assume that the hypothetical, correct specification is the basic model (3.4), with individual and time fixed effects υ_i, ζ_j and λ_t. We, however, use a different model for estimation:

$$y_{ijt} = \underline{\mathrm{x}}'_{ijt}\beta + \upsilon_i + \zeta_j + u_{ijt} \ , \tag{3.25}$$

that is, time fixed effects are left for the disturbance term. The Within estimator of the "misspecified" model reads as

$$\hat{\beta}_1 = (X'M_1X)^{-1}X'M_1y \quad \text{with} \quad M_1 = I - (I_{N_1} \otimes \bar{J}_{N_2T}) - (\bar{J}_{N_1} \otimes I_{N_2} \otimes \bar{J}_T) + \bar{J}_{N_1N_2T} \ ,$$

where M_1 is the matrix representation of the Within transformation $\underline{\mathrm{x}}'_{ijt} - \underline{\bar{\mathrm{x}}}'_{i..} - \underline{\bar{\mathrm{x}}}'_{.j.} + \underline{\bar{\mathrm{x}}}'_{...}$ removing $\upsilon_i + \zeta_j$. If the time effects are correlated with the regressors, $\hat{\beta}_1$ is inconsistent.

Another estimator for the misspecified model, which in fact "over-clears" the effects, is

$$\hat{\beta}_2 = (X'M_2X)^{-1}X'M_2y \quad \text{with} \quad M_2 = (I_{N_1} - \bar{J}_{N_1}) \otimes (I_{N_2} - \bar{J}_{N_2}) \otimes (I_T - \bar{J}_T) \ ,$$

where M_2 is the matrix form of the scalar transformation (see equation (1.16) of Chap. 1)

$$\underline{\mathrm{x}}'_{ijt} - \underline{\bar{\mathrm{x}}}'_{ij.} - \underline{\bar{\mathrm{x}}}'_{i.t} - \underline{\bar{\mathrm{x}}}'_{.jt} + \underline{\bar{\mathrm{x}}}'_{i..} + \underline{\bar{\mathrm{x}}}'_{.j.} + \underline{\bar{\mathrm{x}}}'_{..t} - \underline{\bar{\mathrm{x}}}'_{...} \ .$$

As M_2 clears all fixed effects in all three-dimensional models, $\hat{\beta}_2$ is consistent regardless of the correlation of λ_t with the regressors. To decide then if $\hat{\beta}_1$ (in particular, the M_1 transformation to be imposed) is proper, we should compare $\hat{\beta}_1$ and $\hat{\beta}_2$. We form our null and alternative as

$$H_0 : \ \mathrm{Corr}(\underline{\mathrm{x}}'_{ijt}, \lambda_t) = 0 \quad \text{against} \quad H_1 : \ \mathrm{Corr}(\underline{\mathrm{x}}'_{ijt}, \lambda_t) \neq 0 \ .$$

Again, the first estimator $\hat{\beta}_1$ is consistent under the null, but inconsistent under the alternative, while $\hat{\beta}_2$ is consistent both under the null and the alternative. This leads to another Hausman type test, that is, if the difference of the two estimators $q_3 = \hat{\beta}_1 - \hat{\beta}_2$ is sufficiently close to zero, we conclude that the null cannot be rejected and the transformation M_1 which led to estimator $\hat{\beta}_1$ is consistent. The test statistic $m = \hat{q}'_3 \mathrm{Var}(\hat{q}_3)^{-1}\hat{q}_3$ can be shown to have a centered chi-squared distribution with $k^{(1)}$ degrees of freedom, where $k^{(1)}$ is the number of regressors in x'_{ijt}.

3.5.3.2 Conventional Endogeneity

The second case corresponds to the scenario where correlation remains between the regressors and the disturbance term after the elimination of the fixed effects. That is, we wish to test the hypothesis that $\mathrm{Corr}(\underline{\tilde{\mathrm{x}}}'_{ijt}, \tilde{\varepsilon}_{ijt}) = 0$ in the transformed model

$$\tilde{y}_{ijt} = \underline{\tilde{x}}'_{ijt}\beta + \tilde{\varepsilon}_{ijt} \ ,$$

where "$\sim$" denotes transformed variables. To test this, consider the null and alternative hypotheses as

$$H_0 : \ \mathrm{Corr}(\underline{\tilde{x}}'_{ijt}, \tilde{\varepsilon}_{ijt}) = 0 \quad \text{against} \quad H_1 : \ \mathrm{Corr}(\underline{\tilde{x}}'_{ijt}, \tilde{\varepsilon}_{ijt}) \neq 0 \ .$$

In order to construct the usual Hausman-Taylor test, we need to find an estimator consistent under both the null and the alternative. An instrumental variable estimator suffices if for some z_{ijt}, the following two conditions hold:

$$\mathrm{Corr}(z_{ijt}, \underline{\tilde{x}}'_{ijt}) \neq 0 \quad \text{and} \quad \mathrm{Corr}(z_{ijt}, \tilde{\varepsilon}_{ijt}) = 0 \ .$$

Comparing the Within estimator $\hat{\beta}_1 = (X'MX)^{-1}X'My$ with the IV estimator $\hat{\beta}_2 = (X'MP_ZMX)^{-1}X'MP_ZMy$, with M being the projector orthogonal to the fixed effects and Z the matrix stacked instruments, gives all the required elements of the test. The usual test statistic can be constructed and can be shown to follow a chi-squared distribution.

3.6 Further Considerations

This section discusses some potential directions for further research.

3.6.1 Incomplete Data

Multi-dimensional panels are almost always incomplete. This can arise because of data unavailability or individuals dropping from the sample, which can occur by construction. In the case of flow-type data, such as Gravity models of international trade, even if all the between-country flows are observed, within country movements may not be well defined. That is, y_{ijt} may not exist when $i = j$. Similarly, a complete linked employer-employee data set would require all individuals to work at all firms at all points in time. This clearly cannot happen.

We can think of incompleteness, as is introduced in Sects. 1.4 and 2.4, as that (ij)-pairs in the data are not observed at all T time periods, but $t \in T_{ij}$, where $T_{ij} \subset \{1, \ldots, T\}$ is the index set specific to (ij). Typically, $T_{ij} \neq T_{i'j'}$ for different pairs, and we can define $|T_{ij}|$ as being the cardinality of the set T_{ij}.

Incompleteness, in general, does not violate the feasibility and the validity of HT-type estimators, but may lead to computational difficulties. The transformations required to construct the instruments cannot be represented as simple functions of the sample means of the variables. Instead, we have to rely on the results from Chap. 1 which derive incompleteness-robust data transformations.

Clearly,this complexity is directly related to the complexity of the error component structure. For model (3.4), $(Q_{N_1} \otimes \bar{J}_{N_2 T})$ 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:

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

for $i = 1 \ldots N_1$.

A similar logic holds for $(\bar{J}_{N_1} \otimes Q_{N_2} \otimes \bar{J}_T)$ and $(\bar{J}_{N_1 N_2} \otimes Q_T)$. 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 involving matrix operations is inevitable.

3.6.2 Notes on Higher-dimensional Panels

A full and comprehensive treatment of extensions to four- and higher-dimensional analyses would be lengthy and daunting, and beyond the scope of this chapter. Instead, we describe intuitively the characteristics of higher-dimensional endogeneity modelling.

In the case of three-dimensional models, the total number of random effects model specifications is $64 - 1$ (removing the one with no effects), however this number reaches $2^{14} - 1 = 16,383$ in the case of four-way panels. Although most of these model specifications might not make sense or do not have empirical relevance, it is clear that the number of possible specifications grows rapidly as the number of dimensions increases.

In the case of four-dimensional data, the number of partitions required to implement Hausman-Taylor type estimator grows rapidly. This is accompanied by a large set of highly interdependent order conditions, which might or might not be hard to satisfy.

We also need to consider computational costs when dealing with higher-dimensional data. To carry out the HT-type estimators, data transformations and the estimation of the variances of the random effects have to be conducted. In the case of complete data, all these transformations and estimators are represented by simple scalar operations, which are straightforward to implement computationally. The situation is fundamentally different when the underlying data is unbalanced or incomplete, as data transformations cannot be carried out without using matrices of potentially very large orders.[5]

[5] This argument critically hinges on the number of "large" random effects. If there is only one, transformation can easily be done as the underlying projection is block-diagonal. If there are more than one, however, the projection includes a direct inverse calculation as large as the second largest random effect.

Keeping this computational concern in mind, GMM methods outlined in Sect. 3.3 are more suitable for higher-dimensional models, where only the orthogonality conditions, exploiting the error component structure of the model, are to be constructed.

References

Ahn, S. C. and Schmidt, P. (1999). Estimation of linear panel data models using GMM. In Matyas, L., editor, *Generalized Method of Moments Estimation*. Cambridge University Press.

Baltagi, B. H. (2013). *Econometric Analysis of Panel Data*. Chichester, UK: John Wiley & Sons, 5 edition.

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.

Chan, F., Greene, W., Harris, M., and Konya, L. (2016). Gravity models of trade: Unobserved heterogeneity and endogeneity. *Unpublished Manuscript*, pages 1–26.

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.

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.

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.

Maddala, G. (1971). The use of variance components models in pooling cross section and time series data. *Econometrica*, 39:341–358.

Matyas, L. (1997). Proper specification of the gravity model. *The World Economy*, 20:363–368.

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.

Appendix: Proofs

Proof of Proposition 1

The proof follows the same arguments as Wyhowski (1994). We modified the assumptions in Wyhowski (1994) for a higher dimensional panel. Specifically,

Assumption 1. $\underset{N_1,N_2,T\to\infty}{\text{plim}} \frac{1}{N_1N_2T} X^{(1)'} P_{H_1} X^{(1)} = V_1$

Assumption 2. $\underset{N_1,N_2\to\infty}{\text{plim}} \frac{1}{N_1N_2T} \left(X^{(1)}, X^{(2)}\right)' (P_{H_2} + P_{H_3} + P_{H_4} + P_{H_5}) \left(X^{(1)}, X^{(2)}\right) = V_2(T)$
for fixed T, and $\underset{T\to\infty}{\text{plim}} V_2(T) = V_2$

Assumption 3. $\underset{N_1,T\to\infty}{\text{plim}} \frac{1}{N_1N_2T} \left(X^{(1)}, X^{(3)}\right)' (P_{H_6} + P_{H_7} + P_{H_8} + P_{H_9}) \left(X^{(1)}, X^{(3)}\right) = V_3(N_2)$
for fixed N_2, and $\underset{N_2\to\infty}{\text{plim}} V_3(N_2) = V_3$

Assumption 4. $\underset{N_2,T\to\infty}{\text{plim}} \frac{1}{N_1N_2T} \left(X^{(1)}, X^{(4)}\right)' (P_{H_{10}} + P_{H_{11}} + P_{H_{12}} + P_{H_{13}}) \left(X^{(1)}, X^{(4)}\right) = V_4(T)$
for fixed N_1, and $\underset{N_1\to\infty}{\text{plim}} V_4(N_1) = V_4$

Assumption 5. $\underset{N_1\to\infty}{\text{plim}} \frac{1}{N_1N_2T} \left(X^{(1)}, X^{(2)}, X^{(3)}, X^{(5)}\right)' P_{H_{14}} \left(X^{(1)}, X^{(2)}, X^{(3)}, X^{(5)}\right) = V_5(N_2, T)$
for fixed N_2, and T, and $\underset{N_2,T\to\infty}{\text{plim}} V_5(N_2, T) = V_5$

Assumption 6. $\underset{N_2\to\infty}{\text{plim}} \frac{1}{N_1N_2T} \left(X^{(1)}, X^{(2)}, X^{(4)}, X^{(6)}\right)' P_{H_{15}} \left(X^{(1)}, X^{(2)}, X^{(4)}, X^{(6)}\right) = V_6(N_1, T)$
for fixed N_1, and T, and $\underset{N_1,T\to\infty}{\text{plim}} V_6(N_1, T) = V_6$

Assumption 7. $\underset{T\to\infty}{\text{plim}} \frac{1}{N_1N_2T} \left(X^{(1)}, X^{(3)}, X^{(4)}, X^{(7)}\right)' P_{H_{16}} \left(X^{(1)}, X^{(3)}, X^{(4)}, X^{(7)}\right) = V_7(N_1, N_2)$
for fixed N_1, and N_2, and $\underset{N_1,N_2\to\infty}{\text{plim}} V_7(N_1, N_2) = V_7$

Assumption 8. All V_p, $p = 1 \ldots 7$ are finite, their lower right blocks are all non-singular.

Assumption 9. $\frac{1}{\sqrt{N_1N_2T}} X^{(1)'} P_{H_1} u \to^d N(0, \sigma_1^2 V_1)$ as $N_1, N_2, T \to \infty$

Assumption 10. $\frac{1}{\sqrt{N_1N_2T}} \left(X^{(1)}, X^{(2)}\right)' (P_{H_2} + P_{H_3} + P_{H_4} + P_{H_5}) u \to^d N(0, \sigma_2^2 V_2(T))$ as $N_1, N_2 \to \infty$

Assumption 11. $\frac{1}{\sqrt{N_1N_2T}} \left(X^{(1)}, X^{(3)}\right)' (P_{H_6} + P_{H_7} + P_{H_8} + P_{H_9}) u \to^d N(0, \sigma_3^2 V_3(N_2))$ as $N_1, T \to \infty$

Assumption 12. $\frac{1}{\sqrt{N_1N_2T}}\left(X^{(1)},X^{(4)}\right)'(P_{H_{10}}+P_{H_{11}}+P_{H_{12}}+P_{H_{13}})u \to^d N(0,\sigma_4^2V_4(N_1))$ as $N_2,T\to\infty$

Assumption 13. $\frac{1}{\sqrt{N_1N_2T}}\left(X^{(1)},X^{(2)},X^{(3)},X^{(5)}\right)'P_{H_{14}}u \to^d N(0,\sigma_5^2V_5(N_2,T))$ as $N_1\to\infty$

Assumption 14. $\frac{1}{\sqrt{N_1N_2T}}\left(X^{(1)},X^{(2)},X^{(4)},X^{(6)}\right)'P_{H_{15}}u \to^d N(0,\sigma_6^2V_6(N_1,T))$ as $N_2\to\infty$

Assumption 15. $\frac{1}{\sqrt{N_1N_2T}}\left(X^{(1)},X^{(3)},X^{(4)},X^{(7)}\right)'P_{H_{16}}u \to^d N(0,\sigma_7^2V_7(N_1,N_2))$ as $T\to\infty$

We begin the proof by first verifying the following limits based on the definition of σ_p as defined in equation (3.13).

$$
\begin{aligned}
\lim_{N_1,N_2,T\to\infty} A_1 &= \lim_{N_1,N_2,T\to\infty} \operatorname{diag}\left\{\sigma_1^{-1}I_{k^{(1)}},0_{k^{(2)}+\ldots+k^{(7)}}\right\}\\
&= \operatorname{diag}\left\{\sigma_\varepsilon^{-1}I_{k^{(1)}},0_{k^{(2)}+\ldots+k^{(7)}}\right\}\\
\lim_{T\to\infty} A_p &= \lim_{T\to\infty} \operatorname{diag}\left\{\sigma_p^{-1}I_{k^{(1)}},\sqrt{T}\sigma_p^{-1}I_{k^{(2)}},0_{k^{(3)}+\ldots+k^{(7)}}\right\}\\
&= \operatorname{diag}\left\{0_{k^{(1)}},\sigma_\mu^{-1}I_{k^{(2)}},0_{k^{(3)}+\ldots+k^{(7)}}\right\} \qquad p=1,\ldots 5.\\
\lim_{N_2\to\infty} A_p &= \lim_{N_2\to\infty} \operatorname{diag}\left\{\sigma_p^{-1}I_{k^{(1)}},0_{k^{(2)}},\sqrt{N_2}\sigma_p^{-1}I_{k^{(3)}},O_{k^{(4)}+\ldots+k^{(7)}}\right\}\\
&= \operatorname{diag}\left\{0_{k^{(1)}+k^{(2)}},\sigma_\upsilon^{-1}I_{k^{(3)}},0_{k^{(4)}+\ldots+k^{(7)}}\right\} \qquad p=6,\ldots,9.\\
\lim_{N_1\to\infty} A_p &= \lim_{N_1\to\infty} \operatorname{diag}\left\{\sigma_i^{-1}I_{k^{(1)}},0_{k^{(2)}+k^{(3)}},\sqrt{N_1}\sigma_i^{-1}I_{k^{(4)}},0_{k^{(5)}+\ldots+k^{(7)}}\right\}\\
&= \operatorname{diag}\left\{0_{k^{(1)}+k^{(2)}+k^{(3)}},\sigma_\zeta^{-1}I_{k^{(4)}},0_{k^{(5)}+\ldots+k^{(7)}}\right\} \qquad p=10,\ldots,13.\\
\lim_{T\to\infty} A_{14} &= \lim_{T\to\infty} \operatorname{diag}\left\{\sigma_{14}^{-1}I_{k^{(1)}},\sqrt{T}\sigma_{14}^{-1}I_{k^{(2)}},\sqrt{N_2}\sigma_{14}^{-1}k^{(3)},0_{k^{(4)}},\times\right.\\
&\quad \left.\times\sqrt{T}\sigma_{14}^{-1}I_{k^{(5)}},0_{k^{(6)}+k^{(7)}}\right\}\\
&= \operatorname{diag}\left\{0_{k^{(1)}},\sigma_\mu^{-1}I_{k^{(2)}},0_{k^{(3)}+k^{(4)}},\sigma_\mu^{-1}I_{k^{(5)}},0_{k^{(6)}+k^{(7)}}\right\}\\
\lim_{N_1\to\infty} A_{15} &= \lim_{N_2\to\infty} \operatorname{diag}\left\{\sigma_{15}^{-1}I_{k^{(1)}},\sqrt{T}\sigma_{15}^{-1}I_{k^{(2)}},0_{k^{(3)}},\sqrt{N_1}\sigma_{15}^{-1}k^{(4)},0_{k^{(5)}},\times\right.\\
&\quad \left.\times\sqrt{N_1}\sigma_{15}^{-1}I_{k^{(6)}},0_{k^{(7)}}\right\}\\
&= \operatorname{diag}\left\{0_{k^{(1)}+k^{(2)}+k^{(3)}},\sigma_\upsilon^{-1}I_{k^{(4)}},0_{k^{(5)}},\sigma_\upsilon^{-1}I_{k^{(6)}},0_{k^{(7)}}\right\}\\
\lim_{N_2\to\infty} A_{16} &= \lim_{N_2\to\infty} \operatorname{diag}\left\{\sigma_{16}^{-1}I_{k^{(1)}},0_{k^{(2)}},\sqrt{N_2}\sigma_{16}^{-1}I_{k^{(3)}},\sqrt{N_1}\sigma_7^{-1}I_{k^{(4)}},0_{k^{(5)}+k^{(6)}},\times\right.\\
&\quad \left.\times\sqrt{N_2}\sigma_{16}^{-1}I_{k^{(7)}}\right\}\\
&= \operatorname{diag}\left\{0_{k^{(1)}+k^{(2)}+k^{(3)}},\sigma_\zeta^{-1}I_{k^{(4)}},0_{k^{(5)}+k^{(6)}},\sigma_\zeta^{-1}I_{k^{(7)}}\right\}
\end{aligned}
$$

Defining

$$S^{-1}X'P_H^*XS^{-1}=\sum_{p=1}^{16}A_p\left(\frac{1}{N_1N_2T}X'P_{H_p}X\right)A_p$$

and evaluating the limit of each term in the sum on the right hand side gives

$$
\begin{aligned}
&\operatorname*{plim}_{N_1,N_2,T\to\infty} A_2\left(\tfrac{1}{N_1N_2T}X'(P_{H_2}+P_{H_3}+P_{H_4}+P_{H_5})X\right)A_2 \\
&\quad=\lim A_2\cdot\operatorname*{plim}_{N_1,N_2,T\to\infty}\left(\tfrac{1}{N_1N_2T}X'(P_{H_2}+P_{H_3}+P_{H_4}+P_{H_5})X\right)\cdot\lim A_2 \\
&\quad=\lim A_2\cdot V_2\cdot\lim A_2 \\
&\quad=\tfrac{1}{\sigma_\mu^2}\Gamma_2 \\
&\operatorname*{plim}_{N_1,N_2,T\to\infty} A_6\left(\tfrac{1}{N_1N_2T}X'(P_{H_6}+P_{H_7}+P_{H_8}+P_{H_9})X\right)A_6 \\
&\quad=\lim A_6\cdot\operatorname*{plim}_{N_1,N_2,T\to\infty}\left(\tfrac{1}{N_1N_2T}X'(P_{H_6}+P_{H_7}+P_{H_8}+P_{H_9})X\right)\cdot\lim A_6 \\
&\quad=\lim A_6\cdot V_3\cdot\lim A_6 \\
&\quad=\tfrac{1}{\sigma_\upsilon^2}\Gamma_3 \\
&\operatorname*{plim}_{N_1,N_2,T\to\infty} A_{10}\left(\tfrac{1}{N_1N_2T}X'(P_{H_{10}}+P_{H_{11}}+P_{H_{12}}+P_{H_{13}})X\right)A_{10} \\
&\quad=\lim A_{10}\cdot\operatorname*{plim}_{N_1,N_2,T\to\infty}\left(\tfrac{1}{N_1N_2T}X'(P_{H_{10}}+P_{H_{11}}+P_{H_{12}}+P_{H_{13}})X\right)\cdot\lim A_{10} \\
&\quad=\lim A_{10}\cdot V_4\cdot\lim A_{10} \\
&\quad=\tfrac{1}{\sigma_\zeta^2}\Gamma_4 \\
&\operatorname*{plim}_{N_1,N_2,T\to\infty} A_{14}\left(\tfrac{1}{N_1N_2T}X'P_{H_{14}}X\right)A_{14}=\lim A_{14}\cdot\operatorname*{plim}_{N_1,N_2,T\to\infty}\left(\tfrac{1}{N_1N_2T}X'P_{H_{14}}X\right)\cdot\lim A_{14} \\
&\quad=\lim A_{14}\cdot V_{14}\cdot\lim A_{14} \\
&\quad=\tfrac{1}{\sigma_\mu^2}\Gamma_5 \\
&\operatorname*{plim}_{N_1,N_2,T\to\infty} A_{15}\left(\tfrac{1}{N_1N_2T}X'P_{H_{15}}X\right)A_{15}=\lim A_{15}\cdot\operatorname*{plim}_{N_1,N_2,T\to\infty}\left(\tfrac{1}{N_1N_2T}X'P_{H_{15}}X\right)\cdot\lim A_{15} \\
&\quad=\lim A_{15}\cdot V_{15}\cdot\lim A_{15} \\
&\quad=\tfrac{1}{\sigma_\upsilon^2}\Gamma_6 \\
&\operatorname*{plim}_{N_1,N_2,T\to\infty} A_{16}\left(\tfrac{1}{N_1N_2T}X'P_{H_{16}}X\right)A_{16}=\lim A_{16}\cdot\operatorname*{plim}_{N_1,N_2,T\to\infty}\left(\tfrac{1}{N_1N_2T}X'P_{H_{16}}X\right)\cdot\lim A_{16} \\
&\quad=\lim A_{16}\cdot V_{16}\cdot\lim A_{16} \\
&\quad=\tfrac{1}{\sigma_\zeta^2}\Gamma_7
\end{aligned}
$$

Hence,

$$
\begin{aligned}
\operatorname*{plim}_{N_1,N_2,T\to\infty} S^{-1}X'P_H^*XS^{-1} &= \tfrac{1}{\sigma_\varepsilon^2}\Gamma_1+\tfrac{1}{\sigma_\mu^2}\Gamma_2+\tfrac{1}{\sigma_\upsilon^2}\Gamma_3+\tfrac{1}{\sigma_\zeta^2}\Gamma_4+\tfrac{1}{\sigma_\mu^2}\Gamma_5+\tfrac{1}{\sigma_\upsilon^2}\Gamma_6+\tfrac{1}{\sigma_\zeta^2}\Gamma_7 \\
&= \Gamma
\end{aligned}
$$

Under assumptions (1) - (15),

$$
S^{-1}X'P_H^*u=\frac{1}{N_1N_2T}\sum_{p=1}^{16}A_pX'P_{H_p}u\to^d N(0,\Gamma^{-1}).
$$

It then follows that

$$\begin{aligned} S(\hat{\beta}_{HT2}-\beta) &= S(X'P_H^*X)^{-1}X'P_H^*u \\ &= S(X'P_H^*X)^{-1}SS^{-1}X'P_H^*u \\ &= (S^{-1}X'P_H^*XS^{-1})^{-1}S^{-1}X'P_H^*u \xrightarrow{d} \Gamma^{-1}\cdot N(0,\Gamma) = N(0,\Gamma^{-1}). \end{aligned}$$

Proof of Corollary 1

Let's redefine V_p, A_p matrices, and use σ_p and H_p proper for model (3.4).

Assumption A. $\underset{N_1,N_2,T\to\infty}{\text{plim}} \frac{1}{N_1N_2T}X^{(1)\prime}P_{H_1}X^{(1)} = V_1$

Assumption B. $\underset{N_1\to\infty}{\text{plim}} \frac{1}{N_1N_2T}\left(X^{(1)},X^{(2)}\right)' P_{H_2}\left(X^{(1)},X^{(2)}\right) = V_2(N_2,T)$
for fixed N_2, and T, and $\underset{N_2,T\to\infty}{\text{plim}}\ V_2(N_2,T) = V_2$

Assumption C. $\underset{N_2\to\infty}{\text{plim}} \frac{1}{N_1N_2T}\left(X^{(1)}X^{(3)}\right)' P_{H_3}\left(X^{(1)},X^{(3)}\right) = V_3(N_1,T)$
for fixed N_1, and T, and $\underset{N_1,T\to\infty}{\text{plim}}\ V_3(N_1,T) = V_3$

Assumption D. $\underset{T\to\infty}{\text{plim}} \frac{1}{N_1N_2T}\left(X^{(1)},X^{(4)}\right)' P_{H_4}\left(X^{(1)},X^{(4)}\right) = V_4(N_1,N_2)$
for fixed N_1, and N_2, and $\underset{N_1,N_2\to\infty}{\text{plim}}\ V_4(N_1,N_2) = V_4$

Assumption E. All V_p, $p = 1\ldots4$ are finite, their lower right blocks are all non-singular.

Assumption F. $\frac{1}{\sqrt{N_1N_2T}}X^{(1)\prime}P_{H_1}u \to^d N(0,\sigma_1^2V_1)$ as $N_1,N_2,T\to\infty$

Assumption G. $\frac{1}{\sqrt{N_1N_2T}}\left(X^{(1)},X^{(2)}\right)' P_{H_2}u \to^d N(0,\sigma_2^2V_2(N_2,T))$ as $N_1\to\infty$

Assumption H. $\frac{1}{\sqrt{N_1N_2T}}\left(X^{(1)},X^{(3)}\right)' P_{H_3}u \to^d N(0,\sigma_3^2V_3(N_1,T))$ as $N_2\to\infty$

Assumption I. $\frac{1}{\sqrt{N_1N_2T}}\left(X^{(1)},X^{(4)}\right)' P_{H_4}u \to^d N(0,\sigma_4^2V_4(N_1,N_2))$ as $T\to\infty$

The proof follows the same arguments as the proof for proposition (1).

Chapter 4
Dynamic Models and Reciprocity

Maurice J.G. Bun, Felix Chan, and Mark N. Harris

Abstract This chapter discusses the specification, estimation and testing of dynamic models with multi-dimensional data. The difficulties in estimating dynamic models in standard two-dimensional panel data are well known and these challenges are exacerbated by the more complicated endogeneity problems associated with using multi-dimensional data. Furthermore, the availability of multi-dimensional data allows proper modelling of reciprocity. This chapter analyzes a general model containing both reciprocity and short-run dynamics. It is straightforward to show that endogeneity is an inherent feature of the general model and least squares type estimators will be inconsistent. A set of valid orthogonal conditions is proposed, which is then used in Generalized Method of Moments (GMM) estimation.

4.1 Introduction

Unbiased and consistent fixed and random effects estimation of multi-dimensional panel data models with strictly exogenous regressors have been considered in Chaps. 1 and 2. This chapter considers the consistent estimation of the three-dimensional panel data model when the specification contains autoregressive dynamics and reciprocal relations. Consider the three-dimensional panel data model with outcome variable y_{ijt}. Autoregressive dynamics capture the partial adjustment in the outcome variable due to recent changes in the explanatory factors. Reciprocal relations occur when there is a tendency for a relation between two entities to flow both ways.

Maurice J.G. Bun
University of Amsterdam, Amsterdam, the Netherlands

Felix Chan
Curtin University, Perth, Western Australia

Mark N. Harris
Curtin University, Perth, Western Australia

L. Matyas (ed.), *The Econometrics of Multi-dimensional Panels*, Advanced Studies in Theoretical and Applied Econometrics 50, DOI 10.1007/978-3-319-60783-2_4

The inclusion of lags of the dependent variable (y_{ijt-1} and further lagged values) or reciprocity effects (y_{jit}) leads to different endogeneity problems. Least squares based inference methods, i.e., the fixed effects and random effects estimators of Chaps. 1 and 2, are therefore biased and inconsistent. In two-dimensional panel data models with endogenous regressors, a commonly employed procedure is to transform the model into first differences in order to remove the unobserved individual specific heterogeneity. Next, sequential moment conditions are exploited where lagged levels of the variables are instruments for the endogenous differences and the parameters estimated by the Generalized Method of Moments (GMM). GMM produces consistent parameter estimates for a finite number of time periods (T) and a large cross-sectional dimension (N) (see, e.g., Arellano and Bond, 1991; Arellano and Bover, 1995; Blundell and Bond, 1998). A main reason for its popularity in empirical research is that the GMM estimation approach provides asymptotically efficient inference employing a relatively minimal set of statistical assumptions. This chapter investigates the extent in which we can apply the sequential moment conditions developed for two-dimensional panel data models to multi-dimensional panel data specifications. This includes short-run dynamics and reciprocity effects. Throughout we consider a variety of models for the individual specific heterogeneity. As such, this chapter focuses on fixed T and large N asymptotics. Other cases with large T and fixed N and large T and N are left for further research.[1] The overall approach proposed in this chapter is to apply the appropriate transformation to eliminate the unobserved heterogeneities before estimating the relevant parameters with GMM, as this approach applies to both fixed and random effects models.

Three-dimensional dynamic panel data specifications have been used in the empirical analysis of international trade flows (Egger, 2001; Bun and Klaassen, 2007; Bun et al., 2009), migration flows (Ruyssen et al., 2014) and tourism demand (Maloney and Montes Rojas, 2005). In these applications, a single observation is typically origin country i, destination country j and year t. Typically the standard two-way error components structure including pair (i, j) and time (t) effects has been used. Only in the empirical trade models have time-varying fixed effects been estimated, leading to markedly different estimates compared with traditional panel data specifications. Reciprocal relationships may also be important in these cases, as it is unlikely that import flow from country i to country j would be independent of that in the opposite direction. It would also seem important to examine if migration from state i to state j depends on the population outflow from state j to other states. An area where reciprocity receives a lot of attention is social network analysis, where the bidirectional relation between y_{ijt} and y_{jit} results in a set of simultaneous equations. Applications are the firm-level network analyses of Lincoln et al. (1992) and Keister (2001). One reason why reciprocity has not received the attention that it deserves is perhaps the lack of appropriate econometric techniques that will allow a proper modelling of reciprocity. This chapter demonstrates that the inclusion of the reciprocity term induces a different type of endogeneity problem to the presence of

[1] The number of sequential moment conditions grows rapidly in the time dimension, hence the analysis of large T panels is more complicated due to the additional issue of many moments, for further details see Alvarez and Arellano (2003) and Hayakawa (2015).

a lagged dependent variable, and that the problem is independent of the specification of the unobserved heterogeneity.

This chapter is organised as follows: Sects. 4.2 and 4.3 discuss the GMM estimation of dynamic models and reciprocity, respectively. In Sect. 4.4, we merge dynamics and reciprocity in one model, while in Sect. 4.5 we discuss extensions.

4.2 Dynamics

In this section, we analyze three-dimensional dynamic panel data models of the form:

$$y_{ijt} = \beta y_{ijt-1} + \delta' w_{ijt} + \pi_{ijt} + \varepsilon_{ijt}, \quad t = 2, \ldots, T \,, \tag{4.1}$$

where the data are observed along three indices $i = 1, \ldots, N_1$, $j = 1, \ldots, N_2$ and $t = 1, \ldots, T$ with w_{ijt} denoting a $(K \times 1)$ vector of covariates. Typically, i and j refer to entities (individuals, firms, sectors, regions, countries), while t denotes time (day, week, month, year). We will denote the total number of cross-sectional observations by $N = N_1 N_2$.

Without further structure, the two error components π_{ijt} and ε_{ijt} are indistinguishable from each other. We therefore assume that π_{ijt} is any of the cases described in Chap. 2:

$$\begin{aligned}
&\text{case 1 (equation 2.11):} \quad \pi_{ijt} = \nu_i + \zeta_j + \lambda_t \\
&\text{case 2 (equation 2.14):} \quad \pi_{ijt} = \mu_{ij} \\
&\text{case 3 (equation 2.9):} \quad \pi_{ijt} = \mu_{ij} + \lambda_t \\
&\text{case 4 (equation 2.5):} \quad \pi_{ijt} = \nu_{it} + \zeta_{jt} \\
&\text{case 5 (equation 2.2):} \quad \pi_{ijt} = \mu_{ij} + \nu_{it} + \zeta_{jt} \,,
\end{aligned}$$

where $\varepsilon_{ijt} \sim iid\left(0, \sigma_\varepsilon^2\right)$, while λ_t, ν_{it} and ζ_{jt} are assumed to be weakly stationary random effects and independent of ε_{ijt}. In all specifications in this chapter, we model the unobserved heterogeneity π_{ijt} as purely random effects. In dynamic panel data models, this is by far the most common specification and it implies that the lagged dependent variable regressor y_{ijt-1} is automatically correlated with the random effect π_{ijt}. We therefore do not consider random effects estimators as these assume a zero correlation between regressors and the random effect(s). We instead consider fixed effects estimation, that is, we first apply the appropriate transformation to eliminate the random effect π_{ijt}, and then apply GMM estimation exploiting sequential moment conditions. Following the terminology of Lee (2016), this approach can be classified as fixed effects (or related effects) as opposed to random effects (unrelated effects). As such, the estimation methods proposed in this chapter can also be applied to the fixed effects specifications, namely, equations (1.2) - (1.7), as discussed in Chap. 1.

First, we analyze the consequences of ignoring the unobserved heterogeneity for the GMM estimation using sequential moment conditions from the two-dimensional panel data literature. Second, we consider different moment conditions to deal with

the more elaborate models for the unobserved heterogeneity. We will use first differences and Within transformations to wipe out this unobserved heterogeneity from model (4.1). As such, the same method applies to both fixed and random effects applications of π_{ijt}. Third, in a number of Monte Carlo experiments, we analyze the finite sample properties of the resulting GMM estimators.

4.2.1 Estimation

For ease of exposition, consider the first-order autoregressive model without additional covariates

$$y_{ijt} = \beta y_{ijt-1} + \pi_{ijt} + \varepsilon_{ijt}, \quad t = 2, \ldots, T \,, \tag{4.2}$$

where y_{ij0} is also observed. This model has been analyzed by Balazsi et al. (2015), who show that the standard fixed effects estimator is consistent for cases 1 and 4, but it is inconsistent for cases 2, 3, and 5. The intuition is that whenever bilateral fixed effects (μ_{ij}) are present, some type of Nickell (1981) bias will emerge. In the other cases, the number of incidental parameters is of a lower order than the cross-sectional dimension, and the incidental parameter problem does not appear.

Cases 2, 3 and 5 are straightforward extensions of two-dimensional dynamic panel data models to three-dimensional specifications. As long as T is small and N is large, standard GMM methods for dynamic panel data models can be applied. We consider dynamic models where the idiosyncratic errors obey the following conditional moment restriction:

$$\mathrm{E}\left(\varepsilon_{ijt} | \mathbf{y}_{ij}^{t-1}, \pi_{ijt}\right) = 0, \quad t = 1, \ldots, T \,, \tag{4.3}$$

where $\mathbf{y}_{ij}^{t-1} = \left(y_{ij0}, y_{ij1}, \ldots, y_{ijt-1}\right)'$. Assumption (4.3) rules out serial correlation in ε_{ijt}, which is a base for constructing unconditional moments. As an example, consider case 2, for which $\pi_{ijt} = \mu_{ij}$. Taking first differences, we remove the pair-specific effects (μ_{ij}) resulting in

$$\Delta y_{ijt} = \beta \Delta y_{ijt-1} + \Delta \varepsilon_{ijt} \,, \tag{4.4}$$

for which the following unconditional moment conditions are available:

$$\mathrm{E}\left(\mathbf{y}_{ij}^{t-2} \Delta \varepsilon_{ijt}\right) = 0, \quad t = 2, \ldots, T \,. \tag{4.5}$$

Lagged levels of the endogenous variable can be used as instruments for current changes. Simple IV estimators of this type were first proposed by Anderson and Hsiao (1981, 1982) for the first order autoregressive AR(1) model and in a multivariate setting and GMM framework by Holtz-Eakin et al. (1988) and Arellano and Bond (1991).

It is well known (see, e.g., Blundell and Bond, 1998) that the GMM estimator of the first-differenced model can have poor finite sample properties in terms of bias and precision in this case. One reason for this is that lagged levels are weak predictors of the first differences when data are persistent. Blundell and Bond (1998) advocated the use of extra moment conditions that rely on certain stationarity restrictions on the time series properties of the data, as suggested by Arellano and Bover (1995). These amount to assuming

$$\mathrm{E}\left(\Delta y_{ijt}|\mu_{ij}\right)=0\,, \tag{4.6}$$

which implies that the original series in levels have constant correlation over time with the individual-specific effects. Assumption (4.5) leads to the following additional moment conditions for the model in levels

$$\mathrm{E}\left(\Delta y_{ijt-1}(\mu_{ij}+\varepsilon_{ijt})\right)=0,\quad t=2,\ldots,T\,. \tag{4.7}$$

In other words, lagged changes can be used as instruments for current levels.

With time-varying unobserved heterogeneity, as in cases 4 and 5, however, the standard moment conditions for two-dimensional panel data as outlined in (4.5) are violated. Considering for example, $\pi_{ijt}=\nu_{it}+\mu_{ij}$, we have for the equation in first differences

$$\Delta y_{ijt}=\beta\Delta y_{ijt-1}+\Delta\nu_{it}+\Delta\varepsilon_{ijt},\quad t=2,\ldots,T\,. \tag{4.8}$$

Although we removed the time-invariant incidental parameters μ_{ij}, we cannot apply OLS because the transformed regressor Δy_{ijt-1} is correlated with the transformed idiosyncratic error term $\Delta\varepsilon_{ijt}$ and possibly also with the transformed error component $\Delta\nu_{it}$. Moreover, Arellano and Bond's moments are not valid, as we can show in the following:

$$\mathrm{E}\left(\mathbf{y}_{ij}^{t-2}\Delta\varepsilon_{ijt}|\nu_{it},\nu_{it-1},\ldots\right)\neq 0\,. \tag{4.9}$$

The consequences are therefore similar to introducing an error factor structure (see, for example, Sarafidis and Robertson (2009)). The main difference here is that we can let either N_1 or N_2 go to infinity. This is of some importance for the asymptotic properties of GMM estimators, as demonstrated by the proposition below.

Proposition 1. *Let $\pi_{ijt}=\mu_{ij}+\nu_{it}$ and consider using y_{ijt-2} as a single instrument resulting in the exactly identified IV estimator (Anderson and Hsiao, 1981, 1982). Under fixed T with the assumptions (i) $\varepsilon_{ijt}\sim iid\left(0,\sigma_\varepsilon^2\right)$ and (ii) ν_{it} is weakly stationary, then*

$$\begin{aligned}\operatorname*{plim}_{N_1\to\infty}\hat{\beta}_{AH}&=\beta+\tfrac{c_1}{c_2},\\ \operatorname*{plim}_{N_2\to\infty}\hat{\beta}_{AH}&=\beta+\tfrac{\eta_1}{\eta_2}\,,\end{aligned} \tag{4.10}$$

where c_1 and c_2 are constants and η_1 and η_2 are random variables specified in the proof.

Proof. See Appendix.

For general v_{it}, these results show that the simple IV estimator is inconsistent. The bias is a fixed value when $N_1 \to \infty$ but it is a random limit when $N_2 \to \infty$. In the special case of i.i.d. v_{it}, we furthermore have that $c_1 = 0$ and $\mathrm{E}(\eta_1) = 0$. These results show that in the case of neglected time-varying unobserved heterogeneity, the Arellano and Bond (1991) moment conditions for two-dimensional panel data become invalid and the resulting GMM estimators are inconsistent.

The main message of Proposition (1) is that the unobserved heterogeneities must be accommodated properly for the consistency of the GMM estimator. It is clear that one can obtain similar results for the other cases by following the same arguments as in the proof of Proposition (1); these derivations are therefore omitted for brevity.

For GMM estimation in three-dimensional panel data models, the Arellano and Bond (1991) moment conditions must adapt. The time varying effect v_{it} can be eliminated by the following Within transformation on the model in first differences

$$\Delta y_{ijt} - \Delta \bar{y}_{i.t} = \beta \left(\Delta y_{ijt-1} - \Delta \bar{y}_{i.t-1} \right) + \Delta \varepsilon_{ijt} - \Delta \bar{\varepsilon}_{i.t} \,, \tag{4.11}$$

which can be written more compactly as

$$\widetilde{\Delta y_{ijt}} = \beta \widetilde{\Delta y_{ijt-1}} + \widetilde{\Delta \varepsilon_{ijt}}, \quad t = 2, \ldots, T \,. \tag{4.12}$$

Under assumption (4.3), the following unconditional moment conditions are available:

$$\mathrm{E}\left(\mathbf{y}_{ij}^{t-2} \widetilde{\Delta \varepsilon_{ijt}} \right) = 0, \quad t = 2, \ldots, T \,. \tag{4.13}$$

Again, lagged levels of the endogenous variable can be used as instruments for current changes. As long as $N_1 \to \infty$ and/or $N_2 \to \infty$, the resulting GMM estimator is consistent.

One relevant aspect of the equation (4.12) is that the Within transformed error term $\widetilde{\Delta \varepsilon_{ijt}}$ exhibits cross-sectional dependence. For example, when $\varepsilon_{ijt} \sim i.i.d.(0, \sigma_\varepsilon^2)$ it is easy to see that

$$\mathrm{E}\left(\widetilde{\Delta \varepsilon_{ijt}} \widetilde{\Delta \varepsilon_{ikt}} \right) = -\frac{2\sigma^2}{N_2} \,. \tag{4.14}$$

Therefore, as long as N_2 is finite, the standard weight matrix based on cross-sectional independence will not lead to the optimal GMM estimator. This will have implications for calculating subsequently diagnostic test statistics, such as the overidentifying restrictions statistic. It is clear from equation (4.14) that when $N_2 \to \infty$, cross-sectional dependence vanishes and the aforementioned problems do not occur.

The discussion generalizes in a straightforward manner to models with both *it* and *jt* effects. Fixed effects GMM results when a set of $N_1(T-1) + N_2(T-1)$ (*it* and *jt*) time effects are substituted. Balazsi et al. (2015) show that the appropriate Within transformation to eliminate the *it* and *jt* effects is

$$\widetilde{\Delta y_{ijt}} = \Delta y_{ijt} - \Delta \bar{y}_{.jt} - \Delta \bar{y}_{i.t} + \Delta \bar{y}_{..t}. \tag{4.15}$$

Under assumption (4.3), the moment condition as defined in equation (4.13) is valid, where $\widetilde{\Delta\varepsilon_{ijt}}$ follows the same transformation as defined in equation (4.15). As long as $N_1 \to \infty$ and/or $N_2 \to \infty$, the resulting GMM estimator is consistent. The optimal weight matrix stays unchanged when both N_1 and N_2 are large. When N_1 or N_2 is small, however, cross-sectional error dependence occurs again.

From the discussion so far, provided there is an appropriate transformation, it is clear that equation (4.13) represents a set of valid moment conditions for purposes of estimating the autoregressive parameter, β. In certain cases, such as equation (2.14) when $\pi_{ijt} = \mu_{ij}$, additional moment conditions may also be available. Table (4.1) contains the appropriate transforms for each π_{ijt} specification listed in the Introduction, as well as their corresponding moment conditions.

Table 4.1 List of transforms for equation (4.13)

π_{ijt}	Transforms	Moment condition, $t = 2, \ldots, T$
2.11	$\Delta y_{ijt} - \Delta\bar{y}_{.jt}$ or $\Delta y_{ijt} - \Delta\bar{y}_{i.t}$	$\mathrm{E}\left(\mathbf{y}_{ij}^{t-2}\widetilde{\Delta\varepsilon_{ijt}}\right) = 0$
2.14	Δy_{ijt}	$\mathrm{E}\left(\mathbf{y}_{ij}^{t-2}\widetilde{\Delta\varepsilon_{ijt}}\right) = 0$
2.9	$\Delta y_{ijt} - \Delta\bar{y}_{.jt}$ or $\Delta y_{ijt} - \Delta\bar{y}_{i.t}$	$\mathrm{E}\left(\mathbf{y}_{ij}^{t-2}\widetilde{\Delta\varepsilon_{ijt}}\right) = 0$
2.5	$\Delta y_{ijt} - \Delta\bar{y}_{.jt} - \Delta\bar{y}_{i.t} + \Delta\bar{y}_{..t}$	$\mathrm{E}\left(\mathbf{y}_{ij}^{t-2}\widetilde{\Delta\varepsilon_{ijt}}\right) = 0$
2.2	$\Delta y_{ijt} - \Delta\bar{y}_{.jt} - \Delta\bar{y}_{i.t} + \Delta\bar{y}_{..t}$	$\mathrm{E}\left(\mathbf{y}_{ij}^{t-2}\widetilde{\Delta\varepsilon_{ijt}}\right) = 0$

4.2.2 Monte Carlo Experiments

To show the finite sample properties of estimators and corresponding test statistics, we conduct a small scale Monte Carlo experiment. We generate data according to (4.2) with the unobserved heterogeneity following case 2 (equation (2.14)) and case 5 (equation (2.2)). In the latter case we limit ourselves to ν_{it} only. We estimate (4.2) by the Within estimator and also exploit GMM using the Arellano and Bond (1991) moments in equation (4.5) and the adapted moment conditions in (4.13).

We choose $\beta = 0.5$, hence we do not consider the weak instruments case. We furthermore simulate under covariance stationarity, i.e.,

$$y_{i0} = \frac{\mu_{ij}}{1-\beta} + \frac{\nu_{it} + \varepsilon_{ijt}}{\sqrt{1-\beta^2}}\,, \tag{4.16}$$

where $\mu_{ij} \sim i.i.n.(0, \sigma_\mu^2)$ and $\varepsilon_{ijt} \sim i.i.n.(0, \sigma_\varepsilon^2)$. We also generate ν_{it} according to

$$\nu_{it} = \rho\,\nu_{it-1} + \xi_{it}\,, \tag{4.17}$$

where $\xi_{it} \sim i.i.n.(0, \sigma_\xi^2)$. We set $\sigma_\mu^2 = \sigma_\varepsilon^2 = 1$.

We choose two different data generating processes (labeled dgp) for the unobserved heterogeneity. We set $\sigma_\xi^2 = 0$ ($dgp = 2$) or $\sigma_\xi^2 = 1$ ($dgp = 5$). The former equals case 2, while the latter is case 5. We furthermore choose two values for ρ. When $\rho = 0$ omitting v_{it} does not lead to violation of the moments, while bias will occur in the case of $\rho = 0.5$.

We consider consistent estimation and inference in a small T, large N framework. We therefore set $T = 5$ and $N = 200$, and hence we avoid the issue of many instruments. We vary N_1 and N_2 and choose either 4 or 50. We estimate the model without and with dummy variables for v_{it} (labelled $uh = 2$ and $uh = 5$ respectively) i.e., using the moments (4.5) and (4.13), respectively. All experiments have 1000 replications.

Table 4.2 summarizes the parameter settings of all experiments in this section and reports the simulation results. We report bias and standard deviation of the GMM coefficient estimator, as well as actual rejection percentages at nominal 5% Wald t-statistics (labelled rp). We report the rejection percentage of the Hansen J test (labelled $rp\,J$) checking the validity of the moments. Unreported results show that, as expected, the Within estimator is biased downward with a relatively small standard deviation and the corresponding t-statistic is heavily oversized (up to 100%).

Table 4.2 Parameter values and simulation results for GMM

design	T	N_1	N_2	uh	dgp	ρ	β	mean of $\hat{\beta}$	sd	rp	$rp\,J$
1	5	4	50	2	2	0	0.5	0.471	0.092	6.25	5.10
2	5	50	4	2	2	0	0.5	0.481	0.092	5.70	5.40
3	5	4	50	2	5	0	0.5	0.343	0.215	52.40	97.95
4	5	50	4	2	5	0	0.5	0.471	0.098	10.00	26.95
5	5	4	50	5	5	0	0.5	0.472	0.093	6.35	6.10
6	5	50	4	5	5	0	0.5	0.474	0.106	9.65	17.60
7	5	4	50	2	2	0.5	0.5	0.471	0.092	6.25	5.10
8	5	50	4	2	2	0.5	0.5	0.481	0.092	5.70	5.40
9	5	4	50	2	5	0.5	0.5	0.568	0.264	43.10	96.95
10	5	50	4	2	5	0.5	0.5	0.786	0.126	74.25	32.20
11	5	4	50	5	5	0.5	0.5	0.472	0.093	6.35	6.10
12	5	50	4	5	5	0.5	0.5	0.474	0.106	9.65	17.60

Experiments 1 and 2 show that when only μ_{ij} are included in both data generating process (DGP) and estimation, standard GMM results for two-dimensional panel data are visible. There is no issue with weak or many instruments and there is only a small bias in each of the coefficient estimators, t-statistic and J statistic. As expected, the relative magnitude of N_1 and N_2 is immaterial for the finite sample properties.

In experiments 3 and 4, we add i.i.d. v_{it} to the DGP, but ignore these in the estimation. This has detrimental effects on the estimation bias of the GMM estimator, especially for N_1 small. This can be explained by the theoretical results of the

previous section. Although the unconditional moments in (4.5) are still valid, the conditional moments (4.9) are not zero. Only in the special case of i.i.d. ν_{it} and N_1 large does this not lead to asymptotic bias for the GMM estimator based on (4.5). When N_1 is small as in experiment 3, the GMM estimator converges to a random limit, hence no systematic pattern on the bias can be detected. Indeed the Monte Carlo standard deviation is much larger compared with experiment 4, and large size distortions can be seen in the t-test. This demonstrates the adverse impacts of misspecifying the error components in estimating a dynamic panel data model in higher dimensions.

In experiments 5 and 6, we add a set of $N_1(T-1)$ dummy variables to explicitly take into account ν_{it}. Including these additional fixed effects in the estimation, the GMM estimator now exploits the moment conditions in (4.13). Standard GMM results are visible again irrespective of the relative magnitude of N_1 and N_2. Only when N_1 is large and N_2 is small, does the J test show severe over-rejection. This can be explained by the fact that a suboptimal weight matrix has been used in this case since the cross-sectional dependence due to small N_2 has been ignored (see equation (4.14)).

Experiments 7 to 12 show similar results. In these experiments, time varying unobserved heterogeneity is autocorrelated instead of iid. Experiments 7 and 8 show standard GMM results. However, since ν_{it} is no longer i.i.d. even the unconditional moments in (4.5) are not valid when omitting ν_{it} in estimation (experiments 9 and 10). Finally, experiments 11 and 12 give the same numerical results as experiments 5 and 6 due to the fact that ν_{it} is considered fixed in the GMM estimation.

4.3 Reciprocity

Three-dimensional panel data are used extensively in social network analysis, which investigates relations between social entities. These entities can be individuals, firms, groups or even countries. The smallest group of social entities is a dyad. Consider the case of N_1 entities, then we have $N = \frac{1}{2}N_1(N_1-1)$ dyads.

In the statistical analysis of networks, the outcome variable of interest is typically the occurrence of a link which is measured with a binary indicator variable. The identification of true state dependence in the presence of unobserved heterogeneity can be achieved by considering random effects and fixed effects discrete choice methods from the two-dimensional dynamic panel data literature, see Graham (2015) for an overview. Here we will consider the simpler case of having a quantitative measure as the dependent variable, which allows the use of linear regression models.

In network analysis, reciprocity is defined as the double link (with opposite directions) between entities. Considering a relation y_{ij} between two actors i and j in the network, reciprocity implies the existence of the reverse relation y_{ji}. Lincoln (1984) considers the following cross-section model for reciprocity:

$$\begin{aligned} y_{ij} &= \alpha + \phi y_{ji} + \gamma' x_i + \delta' x_j + \varepsilon_{ij}, \\ y_{ji} &= \alpha + \phi y_{ij} + \delta' x_i + \gamma' x_j + \varepsilon_{ji}\,, \end{aligned} \tag{4.18}$$

where x_i and x_j are exogenous determinants. The assumed equality of coefficients across equations aids identification. Consistent estimation of the reciprocity effect ϕ, however, requires either a zero correlation between the error terms ε_{ij} and ε_{ji} or an exclusion restriction operating through γ and/or δ. Lincoln (1984) discusses in some detail the drawbacks of both approaches. A zero correlation between ε_{ij} and ε_{ji} implies that variables omitted from the model would affect the flow of ties in one direction only. An exclusion restriction would imply that some exogenous regressor affected the outcome in one direction only. Thus, the estimation problem associated with models similar to equation (4.18) requires a more thorough analysis, which is the focus of this subsection.

4.3.1 Within Estimator

Let us examine the estimation problem in a more general setting, specifically, consider,

$$y_{ijt} = \phi y_{jit} + \delta' w_{ijt} + u_{ijt}, \qquad i = 1,\dots N_*,\ t = 1,\dots,T, \tag{4.19}$$

where $N_1 = N_2 = N_*$ ($N = N_*^2$), $u_{ijt} = \pi_{ijt} + \varepsilon_{ijt}$ and w_{ijt} denotes a $(K \times 1)$ vector of covariates. Note that equation (4.19) implies that

$$\begin{pmatrix} y_{ijt} \\ y_{jit} \end{pmatrix} = \begin{pmatrix} y_{jit} \\ y_{ijt} \end{pmatrix} \phi + \begin{pmatrix} w'_{ijt} \\ w'_{jit} \end{pmatrix} \delta + \begin{pmatrix} u_{ijt} \\ u_{jit} \end{pmatrix}. \tag{4.20}$$

Given equation (4.20), it is obvious that y_{jit} is correlated with u_{ijt} since the y_{jit} equation contains y_{ijt} as a covariate. Thus, a reciprocity specification such as equation (4.19) induces another source of endogeneity similar to that caused by simultaneous equations. It should also be obvious that the endogeneity problem caused by reciprocity exists regardless of the specification of u_{ijt}. That is, regardless of the specification of π_{ijt}, pooled OLS and Within estimators as discussed in previous chapters will generally be inconsistent. This can be more formally expressed by the following proposition:

Proposition 2. *Consider equation (4.19), define the Within estimator of ϕ as*

$$\hat{\phi} = \left[y' \left(I_T \otimes K_{N_*} \right) M_D \left(I_T \otimes K_{N_*} \right) y \right]^{-1} y' \left(I_T \otimes K_{N_*} \right) M_D y \tag{4.21}$$

where y denotes the TN vector of y_{ijt} sorted first by i, then j, and then t. K_{N_*} denotes $(N \times N)$ commutation matrix such that $K_{N_*} vec\, A = vec\, A'$ for any $(N_* \times N_*)$ matrix A and M_D denotes the transformation so that $M_D D = 0$ with D being the partitioned matrix (W, Π). Under the assumption that $\phi \neq 1$, $E(y\varepsilon') < \infty$ and $E(yy') < \infty$ then

$$plim\,(\hat{\phi}-\phi)=\lim\frac{(1-\phi^2)\{tr[M_D(I_T\otimes K_{N_*})]+\phi tr(M_D)\}}{(1+\phi^2)tr(M_D)+2\phi tr[M_D(I_T\otimes K_{N_*})]}, \tag{4.22}$$

where the limit can be taken over N, T or (N,T).

Proof. See Appendix.

Note that the right hand side of equation (4.22) is generally not 0. This indicates that the Within estimator is biased even when all the unobserved heterogeneities have been appropriately accommodated. This is not surprising because the source of endogeneity did not come from the unobserved heterogeneities in the case of the reciprocity model.

The bias, however, does depend on the specification of the unobserved heterogeneities as reflected by the presence of M_D on the right hand side of equation (4.22). While it is difficult to generalise the magnitude of the bias, equation (4.22) suggests that the bias tends to be positive.

Table (4.3) provides the traces under the different specifications of π_{ijt}, whilst Table (4.4) provides the corresponding bias expressions asymptotically and semi-asymptotically. The case $\pi_{ijt}=0$ implies $M_D=I_{TN}$, which is equivalent to the OLS estimator.

Table 4.3 Values of $\mathrm{tr}(M_D)$ and $\mathrm{tr}\,(M_D\,(I_T\otimes K_{N_*}))$

π_{ijt}	$\mathrm{tr}(M_D)$	$\mathrm{tr}\,(M_D\,(I_T\otimes K_{N_*}))$
0	TN	TN_*
2.11	$TN_*^2-2N_*-T+2$	$T(N_*-1)$
2.14	$N_*^2(T-1)$	$N_*(T-1)$
2.9	$(N_*^2-1)(T-1)$	$(N_*-1)(T-1)$
2.5	$T(N_*-1)^2$	$T(N_*-1)$
2.2	$(N_*-1)^2(T-1)$	$(N_*-1)(T-1)$

Table 4.4 plim $\left(\hat{\phi}-\phi\right)$ under different specifications of π_{ijt}

π_{ijt}	$N_*\to\infty$	$T\to\infty$	$(T,N_*)\to\infty$
0	$\phi\,(1-\phi^2)\,(1+\phi^2)^{-1}$	$(1-\phi^2)\,(1+\phi N_*)\left[(1+\phi^2)N+2\phi\right]^{-1}$	$\phi\,(1-\phi^2)\,(1+\phi^2)^{-1}$
2.11	$\phi\,(1-\phi^2)\,(1+\phi^2)^{-1}$	$(1-\phi^2))\left[(1+\phi(N_*+1)\right]\left[(1+\phi^2)(N_*+1)+2\phi\right]^{-1}$	$\phi\,(1-\phi^2)\,(1+\phi^2)^{-1}$
2.14	$\phi\,(1-\phi^2)\,(1+\phi^2)^{-1}$	$(1-\phi^2)\,(1+\phi N_*)\left[(1+\phi^2)N_*+2\phi\right]^{-1}$	$\phi\,(1-\phi^2)\,(1+\phi^2)^{-1}$
2.9	$\phi\,(1-\phi^2)\,(1+\phi^2)^{-1}$	$(1-\phi^2)\left[1+\phi(N_*+1)\right]\left[(1+\phi^2)(N_*+1)+2\phi\right]^{-1}$	$\phi\,(1-\phi^2)\,(1+\phi^2)^{-1}$
2.5	$\phi\,(1-\phi^2)\,(1+\phi^2)^{-1}$	$(1-\phi^2)\left[1+\phi(N_*-1)\right]\left[(1+\phi^2)(N_*-1)+2\phi\right]^{-1}$	$\phi\,(1-\phi^2)\,(1+\phi^2)^{-1}$
2.2	$\phi\,(1-\phi^2)\,(1+\phi^2)^{-1}$	$(1-\phi^2)\left[1+\phi(N_*-1)\right]\left[(1+\phi^2)(N_*-1)+2\phi\right]^{-1}$	$\phi\,(1-\phi^2)\,(1+\phi^2)^{-1}$

4.3.2 GMM Estimation

Given the results above, we need to establish a consistent estimator for ϕ. Equation (4.20) provides some insights for obtaining such an estimator. It can be rewritten as a reduced-form:

$$\begin{pmatrix} y_{ijt} \\ y_{jit} \end{pmatrix} = \frac{1}{1-\phi^2} \begin{pmatrix} \delta' & \phi\delta' \\ \phi\delta' & \delta' \end{pmatrix} \begin{pmatrix} w_{ijt} \\ w_{jit} \end{pmatrix} + \frac{1}{1-\phi^2} \begin{pmatrix} 1 & \phi \\ \phi & 1 \end{pmatrix} \begin{pmatrix} u_{ijt} \\ u_{jit} \end{pmatrix} . \tag{4.23}$$

Equation (4.23) suggests that it is possible to consistently estimate ϕ and δ by regressing y_{ijt} on w_{ijt} and w_{jit}. Both ϕ and δ are identifiable under equation (4.23), but ϕ is not identifiable in the absence of w_{ijt} and w_{jit}. Therefore, it is not possible to estimate ϕ if the model contains only the reciprocal term without any additional explanatory variables.

An alternative is to consider IV estimation. This requires a variable z_{ijt} that is correlated with y_{jit} but not with u_{ijt}. Such variable may exist externally but its existence would depend on the context of the dependent variable. In terms of internal instruments, there are two potential candidates, specifically, y_{ijt} and w_{jit}. The first candidate is clearly invalid because y_{ijt}, while it is correlated with y_{jit}, is also correlated with u_{ijt}. Moreover, y_{kst} for $k, s \neq i, j$ cannot be valid instruments because they are not correlated with y_{ijt} by definition. In terms of using w_{jit} as an instrument, this is equivalent to the reduced form approach as discussed previously.

A potential drawback in using w_{jit} as an instrument is that π_{ijt} is unlikely to be zero in practice. It is also likely to be correlated with both w_{ijt} and w_{jit}, thus inducing another form of endogeneity. While it is possible to accommodate this by applying the appropriate transformations as discussed in Chap. 2, such transformations may remove w_{jit}. An alternative is to construct internal instruments following the approach as discussed in Chap. 3 but this is likely to require some strong assumptions about the correlations between the elements in w_{ijt}, w_{jit} and π_{ijt}.

Therefore, we will consider another route to estimate the parameters based, as before, on lagged internal instruments. To demonstrate the idea, consider a simple example:

$$\begin{aligned} y_{ijt} &= \phi y_{jit} + \delta' w_{ijt} + \mu_{ij} + \varepsilon_{ijt}, \\ y_{jit} &= \phi y_{ijt} + \delta w_{jit} + \mu_{ji} + \varepsilon_{jit} . \end{aligned}$$

Note that $\pi_{ijt} = \mu_{ij}$ in this case. We can estimate ϕ from the first equation by taking first differences to remove the fixed effects:

$$\Delta y_{ijt} = \phi \Delta y_{jit} + \delta_1' \Delta w_{it} + \delta_2' \Delta w_{jt} + \Delta \varepsilon_{ijt} . \tag{4.24}$$

The OLS estimator is inconsistent for the first-differenced equation because Δy_{jit} is an endogenous regressor. Assuming that the errors ε_{ijt} are not autocorrelated as in assumption (4.3), we can again use sequential moment conditions. For this model, they are defined as

$$\mathrm{E}(y_{jis}\Delta\varepsilon_{ijt}) = 0 \qquad s = 1,\ldots t-2; \quad t = 2,\ldots,T\,. \tag{4.25}$$

Additional moment conditions follow from the assumed exogeneity of Δw_{ijt} (the first difference of w_{ijt}). Specifically:

$$\mathrm{E}(\Delta w_{ijt}\Delta\varepsilon_{ijt}) = 0 \quad i = 1,\ldots,N_*.$$

These can be used in standard two-step GMM estimation to provide asymptotically valid inference for the reciprocity effect ϕ. The set of appropriate Within transformations for each specification of π_{ijt} is the same as in the dynamic case which can be found in Table 4.1. The set of moment conditions is also similar to those listed in Table 4.1 with $\mathbf{y}_{ij}^{t-2}$ being replaced by $\mathbf{y}_{ji}^{t-2}$.

4.3.3 No Self-flow

A special characteristic of a model with reciprocity is that y_{iit} does not generally exist and the transformations as listed in Table 4.1 are no longer applicable. In order to accommodate the situation of no self-flow, a different set of transformations is required. Similar to the approach above, the general idea is to first remove the static unobserved heterogeneities by taking the first difference, then derive a set of transformations to accommodate the remaining unobserved heterogeneities. The results on self-flow from Chap. 1, specifically equation (1.21), are particularly helpful in this case. Table 4.5 shows the appropriate transformations and the moment conditions for all five cases.

Table 4.5 Transformations and moment conditions for models without self-flow

π_{ijt}	Transformations	Moment Conditions, $t = 2,\ldots,T$
2.11	$\frac{\Delta y_{+jt}}{N-1}$ or $\frac{\Delta y_{i+t}}{N-1}$	$\mathrm{E}\left(\mathbf{y}_{ji}^{t-2}\widetilde{\Delta\varepsilon_{ijt}}\right) = 0$
2.14	Δy_{ijt}	$\mathrm{E}\left(\mathbf{y}_{ji}^{t-2}\widetilde{\Delta\varepsilon_{ijt}}\right) = 0$
2.9	$\frac{\Delta y_{+jt}}{N-1}$ or $\frac{\Delta y_{i+t}}{N-1}$	$\mathrm{E}\left(\mathbf{y}_{ji}^{t-2}\widetilde{\Delta\varepsilon_{ijt}}\right) = 0$
2.5	$\frac{(N-1)\left(\Delta y_{i+t}+\Delta y_{+jt}\right)}{N(N-2)} - \frac{\Delta y_{+it}+\Delta y_{j+t}}{N(N-2)} + \frac{\Delta y_{++t}}{(N-1)(N-2)}$	$\mathrm{E}\left(\mathbf{y}_{ji}^{t-2}\widetilde{\Delta\varepsilon_{ijt}}\right) = 0$
2.2	$\frac{(N-1)\left(\Delta y_{i+t}+\Delta y_{+jt}\right)}{N(N-2)} - \frac{\Delta y_{+it}+\Delta y_{j+t}}{N(N-2)} + \frac{\Delta y_{++t}}{(N-1)(N-2)}$	$\mathrm{E}\left(\mathbf{y}_{ji}^{t-2}\widetilde{\Delta\varepsilon_{ijt}}\right) = 0$

4.4 Combining Dynamics and Reciprocity

Short-run dynamics and reciprocity can be combined and analyzed jointly in one model. For example, in network analysis we are interested in both the persistence of dyads and reciprocity. First, short-run dynamics are present if links are persistent. If actors i and j are previously connected, they are more likely to be linked in subsequent periods. Second, reciprocity implies that a flow from actor i to j is matched with a flow in the opposite direction. A simple model incorporating both network features is

$$y_{ijt} = \phi y_{jit} + \beta y_{ijt-1} + \pi_{ijt} + \varepsilon_{ijt}, \quad i = 1,\ldots,N_1; \quad j > i; \quad t = 1,\ldots,T\ , \tag{4.26}$$

where the unobserved heterogeneity π_{ijt} can take any of the forms discussed previously. Model (4.26) assumes reciprocity only without any other spillover effects. Additional covariates do not materially change the analysis and are omitted for ease of exposition. Note that a relatively small number of entities N_1 can already result in a large number of cross-sectional observations N.

The reciprocity variable y_{jit} is an endogenous regressor and, hence, the standard fixed effects OLS estimator is inconsistent irrespective of the model for the unobserved heterogeneity. Even for large T, the inconsistency does not vanish because the origin of it is simultaneity bias rather than bias due to the predeterminedness of the regressors. As long as T is small and N is large, however, standard GMM methods for dynamic panel data models can be applied. Consider case 2, i.e., only time-invariant unobserved heterogeneity. Taking first differences, we have

$$\Delta y_{ijt} = \phi \Delta y_{jit} + \beta \Delta y_{ijt-1} + \Delta \varepsilon_{ijt}, \quad t = 2,\ldots,T\ . \tag{4.27}$$

Although we have removed the time-invariant incidental parameters μ_{ij}, we cannot apply OLS because the transformed regressors Δy_{jit} and Δy_{ijt-1} are both correlated with the transformed idiosyncratic error term $\Delta \varepsilon_{ijt}$. The conditional moment equation (4.3), however, implies the following $T(T-1)$ Arellano and Bond (1991) moment conditions for the equation in first differences:

$$\left.\begin{aligned} \mathrm{E}(y_{jis}\widetilde{\Delta \varepsilon_{ijt}}) &= 0 \\ \mathrm{E}(y_{ijs}\widetilde{\Delta \varepsilon_{ijt}}) &= 0 \end{aligned}\right\} \quad s = 1,\ldots t-2; \quad t = 2,\ldots,T\ . \tag{4.28}$$

Note that these moment conditions are the combination of those proposed separately for the dynamic and reciprocity models in earlier sections. These moment conditions can be applied in the standard two-step GMM estimation of the parameters of interest, ϕ and β. The assumption of uncorrelated errors ε_{ijt} is crucial for the validity of these moment conditions. Inclusion of additional short-run dynamics, i.e., further lagged values of the dependent variable, may therefore prove to be helpful.

Since the transformations required to eliminate the unobserved heterogeneities and the associated moment conditions are the same for both reciprocity and dynamic models for all error components specifications discussed in this chapter, it is not

surprising that the transformations and moment conditions required for the GMM estimator are also the same as those listed in Table (4.1) for the combined model.

When implementing the GMM estimator, one should take care of both many and weak instrument problems. The number of moment conditions in (4.28) is $O(T^2)$ for each endogenous regressor, hence the total number of orthogonality conditions increases rapidly with the number of time periods. Typically we economize on the number of moment conditions by taking nearest lags only, which leads to less finite sample bias (Bun and Kiviet, 2006). Alternatively, Roodman (2009) proposes collapsing the available moments.

When instruments are weak, that is, only lowly correlated with the endogenous variables, IV and GMM estimators may perform poorly in finite samples (see, e.g., Bound et al. (1995) and Staiger and Stock (1997)). With weak instruments, IV or GMM estimators for two-dimensional panel data models are biased in the direction of the least squares estimator, and their distributions are non-normal (Wansbeek and Knaap, 1999; Hahn et al., 2007; Kruiniger, 2009; Bun and Kleibergen, 2016), affecting inference for ϕ and β in (4.27) using standard $t-$ or Wald testing procedures.

The weak instrument problem in (4.27) occurs primarily through the value of β. From the literature on two-dimensional panel data, it is well known that $\beta \approx 1$ is a notoriously difficult case, as lagged levels are then weak predictors of future first differences. Additionally, $\beta \approx 0$ may also be problematic. To illustrate this, consider the first-differenced set of simultaneous equations:

$$\begin{aligned} \Delta y_{ijt} &= \phi \Delta y_{jit} + \beta \Delta y_{ijt-1} + \Delta \varepsilon_{ijt}, \\ \Delta y_{jit} &= \phi \Delta y_{ijt} + \beta \Delta y_{jit-1} + \Delta \varepsilon_{jit} \ . \end{aligned} \tag{4.29}$$

Suppose the true value of β is zero, then the reduced form becomes

$$\begin{aligned} \Delta y_{ijt} &= \tfrac{1}{1-\phi^2} \left(\phi \Delta \varepsilon_{jit} + \Delta \varepsilon_{ijt} \right), \\ \Delta y_{jit} &= \tfrac{1}{1-\phi^2} \left(\Delta \varepsilon_{jit} + \phi \Delta \varepsilon_{ijt} \right) \ . \end{aligned} \tag{4.30}$$

Now consider the moment conditions for the first equation

$$\begin{aligned} \mathrm{E}\left[y_{ijs} (\Delta y_{ijt} - \phi \Delta y_{jit-1} - \beta \Delta y_{ijt}) \right] &= 0, \\ \mathrm{E}\left[y_{jis} (\Delta y_{ijt} - \phi \Delta y_{jit-1} - \beta \Delta y_{ijt}) \right] &= 0 \, , \end{aligned} \tag{4.31}$$

for $s = t-2,\ t-3, \ldots$. Under $\varepsilon_{ijt} \sim iid(0, \sigma_\varepsilon^2)$, the strength of the instruments depends primarily on the following two covariances between regressors and instruments:

$$\mathrm{Cov}\left(y_{ijs}, \Delta y_{ijt-1} \right), \tag{4.32}$$

$$\mathrm{Cov}\left(y_{jis}, \Delta y_{jit} \right) \ . \tag{4.33}$$

The first moment condition is informative so long as the panel data are not persistent. A weak instrument problem will occur, however, when the true value of β is close to one. The same holds for the second moment condition, but there is an additional identification problem when β is close to zero. In that case, the covariance in (4.33)

is

$$\mathrm{Cov}\left(y_{jis}, \Delta y_{jit}\right) = \mathrm{Cov}\left(y_{jis}, \frac{1}{1-\phi^2}\left(\Delta\varepsilon_{jit} + \phi\Delta\varepsilon_{ijt}\right)\right) = 0$$

for all $s = t-2,\ t-3,\ldots$ In other words, instrumenting the endogenous reciprocity regressor with lagged values requires some autoregressive dynamics in network formation.

4.4.1 Monte Carlo Experiments

To show the finite sample properties of two-step GMM estimators and corresponding test statistics in the model with both dynamics and reciprocity, we conduct a small scale Monte Carlo experiment. We generate data according to (4.26) for the unobserved heterogeneity case 2. We estimate (4.26) using the Within estimator and exploit GMM estimator using the Arellano and Bond (1991) moments in (4.28).

We choose $\phi = \{0, 0.5\}$ and $\beta = \{0, 0.5\}$. The case $\beta = 0$ is potentially a weak instruments case. The model is simulated under weak stationarity with $\mu_{ij} \sim i.i.n.(0, \sigma_\mu^2)$ and $\varepsilon_{ijt} \sim i.i.n.(0, \sigma_\varepsilon^2)$; finally $\sigma_\mu^2 = \sigma_\varepsilon^2 = 1$.

We consider consistent estimation and inference in a small T, large N framework and therefore set $T = 5$, which already implies $T(T-1) = 20$ instruments. We vary N_1 and choose either 40 or 80 resulting in $N = \{780, 3160\}$. Compared with existing simulation results for two-dimensional panel data, these cross-section dimensions are relatively large. All experiments have 1000 replications.

Table 4.6 summarizes the parameter settings of all experiments in this section and reports the simulation results. We report bias and standard deviation of the GMM coefficient estimator, as well as actual rejection percentages at nominal 5% Wald t-statistics. We report the rejection percentage of the Hansen J test checking the validity of the moments. Unreported results show that, as expected, the Within estimator is biased downward for β with a relatively small standard deviation and the corresponding t-statistic is heavily oversized (up to 100%). Furthermore, bias in estimating ϕ is upward.

Experiment 1 shows that for $\phi = \beta = 0.5$ and $N_1 = 40$, standard GMM results are visible. There are no issues with weak or many instruments, hence, there is small bias in all of the coefficient estimator, t-statistic and J statistic. Experiment 2 shows that the value of ϕ seems largely immaterial for the finite sample properties. Experiment 3, however, confirms that weak instruments problems occur when $\beta = 0$. This has detrimental effects on the estimation bias of the GMM estimator of ϕ, which is biased towards the Within estimator. This bias will not disappear for larger N_1 as shown in experiment 4. Remarkably, there does not seem to be a weak instrument problem for β.

Table 4.6 Parameter values and simulation results for GMM

design	T	N_1	ϕ	mean of $\hat{\phi}$	sd	rp	β	mean of $\hat{\beta}$	sd	rp	$rp\,J$
1	5	40	0.5	0.541	0.055	13.60	0.5	0.459	0.054	14.30	5.80
2	5	40	0	-0.002	0.089	5.20	0.5	0.488	0.047	5.60	4.60
3	5	40	0.5	0.805	0.150	57.40	0	-0.001	0.017	5.30	2.70
4	5	80	0.5	0.811	0.145	61.60	0	-0.000	0.008	4.40	3.40

4.5 Extensions

This section discusses some possible extensions, including a generalisation of reciprocity as well as extensions to higher dimensions.

4.5.1 Generalized Reciprocity

The models so far assume reciprocity only without any other spillover effects. In fact, the first two index sets are restricted to be the same. That is, the index set for i is the same as the index set for j. It is possible to write this as a special case of

$$y_{ijt} = \beta y_{ijt-1} + \sum_{p \neq i} \sum_{q \neq j} \phi_{pq} y_{pqt} + \mu_{ij} + \varepsilon_{ijt} \,, \tag{4.34}$$

where the first index set does not have to be the same as the second index set. In fact, if we define the first (or second) index set as a set of geographical regions, then this specification can also be viewed as a generalisation of the Spatial Autoregressive (SAR) model.

Although the models so far do not include spatial spillovers at the ijt level, the possibility of controlling for cross-sectional dependencies increases in three dimensional panel data models. In the standard two-dimensional error components model, it is mainly the aggregated time effect λ_t which models the correlation between two cross-sectional units. Additionally, in fixed effects estimation we allow for arbitrary correlation between the time invariant μ_{ij} effects. In the three-dimensional model, however, we can also incorporate νit and ζ_{jt}. Therefore, in fixed effects GMM estimation, we can allow for arbitrary cross-sectional dependencies at the it and jt level.

If we want to allow for spillover effects at the ijt level, it is necessary to change the set of moment conditions. A relatively simple extension is the model of Baltagi et al. (2014). Abstracting from additional covariates, spatial autoregressive errors and adapting their set up to three dimensions, their model is equation (4.34) with $\phi_{pq} = \phi w_{pq}$ with w_{pq} being the known spatial weights. Note that model (4.26) results as a special case when

$$w_{pq} = \begin{cases} 1, & p = j, q = i, \\ 0, & \text{elsewhere.} \end{cases}$$

Taking first differences, we get

$$\Delta y_{ijt} = \beta \Delta y_{ijt-1} + \phi \sum_{p \neq i} \sum_{q \neq j} w_{pq} \Delta y_{pqt} + \Delta \varepsilon_{ijt} \ . \tag{4.35}$$

The Arellano and Bond (1991) moment conditions are then

$$\left.\begin{array}{r} \mathrm{E}(y_{jis} \Delta \varepsilon_{ijt}) = 0 \\ \mathrm{E}\left(\sum_{p \neq i} \sum_{p \neq j} w_{pq} y_{pqs} \Delta \varepsilon_{ijt}\right) = 0 \end{array}\right\} \quad s = 1, \ldots, t-2; \quad t = 2, \ldots, T \ . \tag{4.36}$$

Once we know the spatial weights, this is a straightforward implementation again of standard Arellano and Bond (1991) GMM.

It is also possible to allow for spatial dependence in the errors. Adapting the model of Baltagi et al. (2014) to the three-dimensional panel data model, spatial autoregressive errors can be specified as

$$\varepsilon_{ijt} = \rho \sum_{p \neq i} \sum_{q \neq j} m_{pq} \varepsilon_{pqt} + u_{ijt} \ ,$$

where m_{pq} is known. The set of moment conditions in (4.36) will stay valid, but efficiency gains can be achieved. Sarafidis (2016) shows that the spatial Arellano and Bond (1991) moment conditions are non-redundant and lead to an increase in the asymptotic efficiency of existing estimators if they are used in combination with the original moment conditions.

Further extensions to higher order (spatial) autoregressive models are provided in Lee and Yu (2014). An important limitation of the spatial approach is that, up to a scale parameter, the spatial interaction matrices are assumed to be known constants. Kuersteiner and Prucha (2015) relax this assumption and allow for data-dependent spatial weights. Furthermore, in their GMM framework, they allow for an error factor structure, which is an example of strong cross-sectional dependence.

4.5.2 Higher Dimensions

The methods proposed in this chapter extend naturally beyond three dimensions for the dynamic case. For example, consider the four dimensional dynamic panel data model:

$$y_{ijkt} = \beta y_{ijkt-1} + \delta' w_{ijkt} + \pi_{ijkt} + \varepsilon_{ijkt}.$$

As the number of dimensions increases, the theory on estimation with GMM and the proposed moment conditions remains applicable under a similar set of assumptions. For example,

$$\mathrm{E}\left(\mathbf{y}_{ijk}^{t-2} \widetilde{\Delta \varepsilon_{ijkt}}\right) = 0, \quad t = 2, ..., T \tag{4.37}$$

is a set of valid moment conditions provided the appropriate transformation is available for four dimensional data. Therefore, the potential difficulty in higher dimensions appears to be the derivation of the appropriate transformations to eliminate the unobserved heterogeneities. This has been discussed in some detail already in Chap. 1.

The definition of reciprocity is somewhat unclear in higher dimensions, as there are $m!$ possible ways to arrange the indices of the m individual units for $m+1$-dimensional data. Therefore the validity of the proposed approach would depend on the definition of reciprocity and the model specification in higher dimensions.

References

Alvarez, J. and Arellano, M. (2003). The time series and cross-sectional asymptotics of dynamic panel data estimators. *Econometrica*, 71:1121–1159.

Anderson, T. and Hsiao, C. (1981). Estimation of dynamic models with error components. *Journal of the American Statistical Association*, 76:598–606.

Anderson, T. and Hsiao, C. (1982). Formulation and estimation of dynamic models using panel data. *Journal of Econometrics*, 18:47–82.

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 Bover, O. (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics*, 68:29–51.

Balazsi, L., Matyas, L., and Wansbeek, T. (2015). The estimation of multidimensional fixed effects panel data models. *Econometric Reviews*, forthcoming, published online: April 2015.

Baltagi, B. H., Fingleton, B., and Pirotte, A. (2014). Estimating and forecasting with a dynamic spatial panel data model. *Oxford Bulletin of Economics and Statistics*, 76:112–138.

Blundell, R. and Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, 87:115–143.

Bound, J., Jaeger, D., and Baker, R. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association*, 90:443–450.

Bun, M. and Kiviet, J. (2006). The effects of dynamic feedbacks on LS and MM estimator accuracy in panel data models. *Journal of Econometrics*, 132:409–444.

Bun, M. and Klaassen, F. (2007). The euro effect on trade is not as large as commonly thought. *Oxford Bulletin of Economics and Statistics*, 69:473–496.

Bun, M., Klaassen, F., and Tan, R. (2009). Free trade areas and intra-regional trade: The case of ASEAN. *Singapore Economic Review*, 54:319–334.

Bun, M. and Kleibergen, F. (2016). Identification and inference in moments based analysis of linear dynamic panel data models. *UvA-Econometrics discussion paper, University of Amsterdam*, 2013/07.

Egger, P. (2001). European exports and outward foreign direct investment: A dynamic panel data approach. *Weltwirtschaftliches Archiv*, 137:427–449.

Graham, B. (2015). Methods of identification in social networks. *Annual Review of Economics*, 7:465–485.

Hahn, J., Hausman, J., and Kuersteiner, G. (2007). Long difference instrumental variables estimation for dynamic panel models with fixed effects. *Journal of Econometrics*, 140:574–617.

Hayakawa, K. (2015). The asymptotic properties of the system GMM estimator in dynamic panel data models when both N and T are large. *Econometric Theory*, 31:647–667.

Holtz-Eakin, D., Newey, W., and Rosen, H. (1988). Estimating vector autoregressions with panel data. *Econometrica*, 56:1371–1395.

Keister, L. (2001). Exchange structures in transition: Lending and trade relations in Chinese business groups. *American Sociological Review*, 66:336–360.

Kruiniger, H. (2009). GMM estimation and inference in dynamic panel data models with persistent data. *Econometric Theory*, 25:1348–1391.

Kuersteiner, G. and Prucha, I. (2015). Dynamic spatial panel models: Networks, common shocks, and sequential exogeneity. *Working paper, University of Maryland*.

Lee, L. and Yu, J. (2014). Efficient GMM estimation of spatial dynamic panel data models with fixed effects. *Journal of Econometrics*, 180:174–197.

Lee, M.-J. (2016). *Micro-Econometrics: Method of moments and limited dependent variables*. Springer.

Lincoln, J. (1984). Analyzing relations in dyads. *Sociological Methods & Research*, 13:45–76.

Lincoln, J., Gerlach, M., and Takahashi, P. (1992). Keiretsu networks in the Japanese economy: A dyad analysis of intercorporate ties. *American Sociology Review*, 57:561–585.

Maloney, W. and Montes Rojas, G. (2005). How elastic are sea, sand and sun? Dynamic panel estimates of the demand for tourism. *Applied Economics Letters*, 12:277–280.

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica*, 49:1417–1426.

Roodman, D. (2009). A note on the theme of too many instruments. *Oxford Bulletin of Economics and Statistics*, 71:135–158.

Ruyssen, I., Everaert, G., and Rayp, G. (2014). Determinants and dynamics of migration to OECD countries in a three-dimensional panel framework. *Empirical Economics*, 46:175–197.

Sarafidis, V. (2016). Neighbourhood GMM estimation of dynamic panel data models. *Computational Statistics & Data Analysis*, 100:526–544.

Sarafidis, V. and Robertson, D. (2009). On the impact of error cross-sectional dependence in short dynamic panel estimation. *Econometrics Journal*, 12:62–81.

Staiger, D. and Stock, J. (1997). Instrumental variables regression with weak instruments. *Econometrica*, 65:557–586.

Wansbeek, T. J. and Knaap, T. (1999). Estimating a dynamic panel data model with heterogenous trends. *Annales d'Economie et de Statistique*, 55–56:331–350.

Appendix

Proof of Proposition 1

The Anderson and Hsiao (1981, 1982) IV estimator is equal to

$$\hat{\beta}_{AH} = \beta + \sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\sum_{t=2}^{T} \frac{y_{ijt-2}\left(\Delta\nu_{it} + \Delta\varepsilon_{ijt}\right)}{y_{ijt-2}\Delta y_{ijt-1}} \,.$$

Under the assumptions of the proposition, model (4.2) can be rewritten as:

$$\begin{aligned} y_{ijt} &= \beta y_{ijt-1} + \nu_{it} + \mu_{ij} + \varepsilon_{ijt} = \tfrac{\mu_{ij}}{1-\beta} + \textstyle\sum_{s=0}^{\infty}\beta^j\nu_{it-s} + \sum_{s=0}^{\infty}\beta^j\varepsilon_{ijt-s} \\ &= \tfrac{\mu_{ij}}{1-\beta} + w_{it} + \textstyle\sum_{s=0}^{\infty}\beta^j\varepsilon_{ijt-s} \,, \end{aligned}$$

$$\begin{aligned} \Delta y_{ijt} &= \beta\Delta y_{ijt-1} + \Delta\nu_{it} + \Delta\varepsilon_{ijt} = \textstyle\sum_{s=0}^{\infty}\beta^j\Delta\nu_{it-s} + \sum_{s=0}^{\infty}\beta^j\Delta\varepsilon_{ijt-s} \\ &= \Delta w_{it} + \textstyle\sum_{s=0}^{\infty}\beta^j\Delta\varepsilon_{ijt-s} \,. \end{aligned}$$

Therefore, we have

$$\mathrm{E}\left(y_{ijt-2}\left(\Delta\nu_{it} + \Delta\varepsilon_{ijt}\right)|\nu_{it}, \nu_{it-1}, \ldots\right) = \Delta\nu_{it}\sum_{s=0}^{\infty}\beta^j\nu_{it-2-s} \neq 0 \,,$$

$$\begin{aligned} &= w_{it-2}\Delta w_{it-1} + \mathrm{E}\left(\textstyle\sum_{s=0}^{\infty}\beta^j\varepsilon_{ijt-2-s}\sum_{s=0}^{\infty}\beta^j\Delta\varepsilon_{ijt-1-s}\right) \\ &= w_{it-2}\Delta w_{it-1} - \tfrac{\sigma_\varepsilon^2}{1+\beta} \,. \end{aligned}$$

When $N_1 \to \infty$ we have that

$$\begin{aligned} \operatorname*{plim}_{N_1\to\infty} \tfrac{1}{N_1}\textstyle\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\sum_{t=2}^{T} y_{ijt-2}\left(\Delta\nu_{it} + \Delta\varepsilon_{ijt}\right) &= \operatorname*{plim}_{N_1\to\infty} \tfrac{1}{N_1}\textstyle\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\sum_{t=2}^{T} w_{it-2}\Delta\nu_{it} \\ &= N_2\mathrm{E}\left(\textstyle\sum_{t=2}^{T} w_{it-2}\Delta\nu_{it}\right) \\ &= c_1 \,, \end{aligned}$$

$$\begin{aligned} \operatorname*{plim}_{N_1\to\infty} \tfrac{1}{N_1}\textstyle\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\sum_{t=2}^{T} y_{ijt-2}\Delta y_{ijt-1} &= \operatorname*{plim}_{N_1\to\infty} \tfrac{1}{N_1}\textstyle\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\sum_{t=2}^{T} w_{it-2}\Delta w_{it-1} \\ &\quad + \operatorname*{plim}_{N_1\to\infty} \tfrac{1}{N_1}\textstyle\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\sum_{t=2}^{T}\left(\sum_{s=0}^{\infty}\beta^j\varepsilon_{ijt-2-s}\sum_{s=0}^{\infty}\beta^j\Delta\varepsilon_{ijt-1-s}\right) \\ &= N_2\mathrm{E}\left(\textstyle\sum_{t=2}^{T} w_{it-2}\Delta w_{it-1}\right) - \tfrac{N_2(T-1)\sigma_\varepsilon^2}{1+\beta} \\ &= c_2 \,, \end{aligned}$$

hence the asymptotic bias can be expressed as

$$\operatorname*{plim}_{N_1\to\infty}\left(\hat{\beta}_{AH}-\beta\right)=\frac{c_1}{c_2}\,.$$

Both c_1 and c_2 are nonrandom limits because we took large N_1 cross-sectional averages. In the special case that v_{it} is i.i.d. we furthermore have that $c_1=0$, hence $\hat{\beta}_{AH}$ is consistent for β.

When $N_2\to\infty$ we have that

$$\begin{aligned}\operatorname*{plim}_{N_2\to\infty}\tfrac{1}{N_2}\textstyle\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\sum_{t=2}^{T}y_{ijt-2}\left(\Delta v_{it}+\Delta\varepsilon_{ijt}\right)&=\operatorname*{plim}_{N_2\to\infty}\tfrac{1}{N_2}\textstyle\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\sum_{t=2}^{T}w_{it-2}\Delta v_{it}\\&=\textstyle\sum_{i=1}^{N_1}\sum_{t=2}^{T}w_{it-2}\Delta v_{it}\\&=\eta_1\,,\end{aligned}$$

$$\begin{aligned}\operatorname*{plim}_{N_2\to\infty}\tfrac{1}{N_2}\textstyle\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\sum_{t=2}^{T}y_{ijt-2}\Delta y_{ijt-1}&=\operatorname*{plim}_{N_2\to\infty}\tfrac{1}{N_2}\textstyle\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\sum_{t=2}^{T}w_{it-2}\Delta w_{it-1}\\&\quad+\operatorname*{plim}_{N_2\to\infty}\tfrac{1}{N_2}\textstyle\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\sum_{t=2}^{T}\left(\sum_{s=0}^{\infty}\beta^j\varepsilon_{ijt-2-s}\sum_{s=0}^{\infty}\beta^j\Delta\varepsilon_{ijt-1-s}\right)\\&=\textstyle\sum_{i=1}^{N_1}\sum_{t=2}^{T}w_{it-2}\Delta w_{it-1}-\frac{N_1(T-1)\sigma_\varepsilon^2}{1+\beta}\\&=\eta_2\,,\end{aligned}$$

hence the asymptotic bias can be expressed as

$$\operatorname*{plim}_{N_2\to\infty}\left(\hat{\beta}_{AH}-\beta\right)=\frac{\eta_1}{\eta_2}\,.$$

Both η_1 and η_2 are random limits now, however, because we took large N_2 cross-sectional averages. In the special case that v_{it} is i.i.d., we have that $\mathrm{E}(\eta_1)=0$, which explains that in the Monte Carlo simulations $\hat{\beta}_{AH}$ is still centered around β.

Proof of Proposition 2

Under the assumptions on the existence of moments, straightforward manipulation on equation (4.21) gives

$$\begin{aligned}\hat{\phi}=&\phi+\left[y'\left(I_T\otimes K_{N_*}\right)M_D\left(I_T\otimes K_{N_*}\right)y\right]^{-1}y'\left(I_T\otimes K_{N_*}\right)M_D\varepsilon\\\operatorname{plim}\hat{\phi}-\phi=&\operatorname{plim}\operatorname{tr}\left[y'\left(I_T\otimes K_{N_*}\right)M_D\left(I_T\otimes K_{N_*}\right)y\right]^{-1}y'\left(I_T\otimes K_{N_*}\right)M_D\varepsilon\\=&\operatorname{plim}\frac{\operatorname{tr}\left[y'\left(I_T\otimes K_{N_*}\right)M_D\varepsilon\right]}{\operatorname{tr}\left[y'\left(I_T\otimes K_{N_*}\right)M_D\left(I_T\otimes K_{N_*}\right)y\right]}\\=&\operatorname{plim}\frac{\operatorname{tr}\left[M_D\left(I_T\otimes K_{N_*}\right)y\varepsilon'\right]}{\operatorname{tr}\left[\left(I_T\otimes K_{N_*}\right)M_D\left(I_T\otimes K_{N_*}\right)yy'\right]}\\=&\lim\frac{\operatorname{tr}\left[M_D\left(I_T\otimes K_{N_*}\right)\mathrm{E}\left(y\varepsilon'\right)\right]}{\operatorname{tr}\left[M_D\left(I_T\otimes K_{N_*}\right)\mathrm{E}(yy')\left(I_T\otimes K_{N_*}\right)\right]}\,.\end{aligned}\tag{A2}$$

Under the assumption that $\varepsilon \sim iid(0, \sigma_\varepsilon^2 I)$, it is straightforward to show that

$$\mathrm{E}\left(y_{ijt}\varepsilon_{ijt}\right) = \begin{cases} \frac{\sigma^2}{1-\phi^2} & i \neq j \\ \frac{\sigma^2}{1-\phi} & i = j \end{cases} \qquad \mathrm{E}\left(y_{jit}\varepsilon ijt\right) = \frac{\phi\sigma^2}{1-\phi^2}$$

$$\mathrm{E}\left(y_{ijt}^2\right) = \begin{cases} \frac{\left(1+\phi^2\right)\sigma^2}{\left(1-\phi^2\right)^2} & i \neq j \\ \frac{(1+\phi)\sigma^2}{\left(1-\phi^2\right)^2} & i = j \end{cases} \qquad \mathrm{E}\left(y_{ijt}y_{jit}\right) = 2\frac{\phi\sigma^2}{\left(1-\phi^2\right)^2}.$$

Using these expressions, it is possible to show that

$$\mathrm{E}\left(y\varepsilon'\right) = \sigma^2\left[\frac{1}{1-\phi^2}I_{TN} + \frac{\phi}{1-\phi^2}\left(I_T \otimes K_{N_*}\right)\right]$$

$$\mathrm{E}\left(yy'\right) = \sigma^2\left[\frac{1+\phi^2}{\left(1-\phi^2\right)^2}I_{TN} + \frac{2\phi}{\left(1-\phi^2\right)^2}\left(I_T \otimes K_{N_*}\right)\right].$$

Substituting these results in the last line of (A2) gives the results. This completes the proof.

Chapter 5
Random Coefficients Models

Jaya Krishnakumar, Monika Avila Márquez, and Laszlo Balazsi

Abstract This chapter deals with specification, estimation and inference within the framework of a random coefficient model in the presence of higher dimensional panel data. Most of the chapter is concerned with a three-dimensional setting with an extension to higher dimensions at the end. We discuss several estimation methods, starting with the GLS made feasible by a new procedure for the estimation of the variance-covariance components, as well as an extension of the MINQUE approach for this setting. We also derive the full Maximum Likelihood and a Restricted Maximum Likelihood estimators involving the maximization of the log-likelihood in a subset of the parameter space for an independent estimation of the variance-covariance elements. Furthermore, we design specification tests that allow to determine if the response coefficients are constant or varying. Additionally, we present different extensions of the linear model including unbalanced panels, correlated random components and correlation of the stochastic elements with the regressors. Finally, the chapter ends with brief discussions of non-linear and higher dimensional extensions as well as a simulation experiment comparing the performance of the above methods in a finite sample setting.

5.1 Introduction

The main focus of this chapter is a Random Coefficient Linear Model (RCM) for three-dimensional panel data, within which we discuss the estimation of the unknown parameters, hypothesis testing and some special issues, such as correlated

Jaya Krishnakumar
University of Geneva, Geneva, Switzerland

Monika Avila Márquez
University of Geneva, Geneva, Switzerland

Laszlo Balazsi
Central European University, Budapest, Hungary

L. Matyas (ed.), *The Econometrics of Multi-dimensional Panels*, Advanced Studies in Theoretical and Applied Econometrics 50, DOI 10.1007/978-3-319-60783-2_5

effects and unbalanced data. After a detailed examination of this model, we present some extensions to higher-dimensional settings as well as nonlinear specifications. Our model is a generalization of the random coefficient linear model for the analysis of multilevel panel data allowing the intercept and slope coefficients to vary along the three dimensions. A typical three-dimensional panel data set consists of time series observations for different individuals within groups. For instance, one can think of time series data on GDP growth for different regions or states within various countries, or time series on investments of firms belonging to different industries, and so on. In this setting, the heterogeneity of responses at the three levels of the data can be captured by assuming that the coefficients vary over all three dimensions. Due to the potentially large number of unknown parameters that this entails, the coefficients are usually assumed to be random and made up of different components corresponding to the different dimensions.

The motivation to propose this study is twofold. First, there is an increasing availability of three-level panel data as well as an increasing use of such data in economic models. Secondly, varying coefficients improve the fitting and predictive power of the model when it is not appropriate to assume that the response coefficients are fixed along cluster, individual and time dimensions. Indeed, ignoring these sources of heterogeneity can lead to biased estimators (see, e.g., Skrondal and Rabe-Hesketh, 2010). Additionally, De Leeuw et al. (2008) explain that individuals who belong to a group can be correlated, while interclass correlation is less common. Modelling three-level heterogeneity can capture these intra and inter class dependence structures.

Three-level panel data have been widely used in several fields such as international trade, biostatistics and health economics. Egger and Pfaffermayr (2003) adopted nested fixed effects for three-level panel data to take into account bilateral interaction effects. Gibbons and Hedeker (1997) studied the consequences of smoking cessation interventions by fitting a random effects probit model to this type of panel data. One can also refer to Balazsi et al. (2015), who present a three-dimensional fixed effects model and a Within procedure to estimate it. The most common model includes fixed effects for each of the three dimensions and the most efficient Within estimator is the one that wipes out the three effects simultaneously. They also present another model that takes into account bilateral interactions by including nested fixed effects. Finally, they analyze the behaviour of the estimators under unbalanced panel data and dynamic autoregressive models. The models employed for the analysis of three-level panel data are principally linear with either fixed or random effects.

Random coefficient models go further than the linear introduction of fixed and random effects because they represent unobserved individual heterogeneity in responses to explanatory variables. Several empirical studies have shown that varying response coefficients are more appropriate in various practical situations. To illustrate, we can mention the study developed by Heckman and Vytlacil (1998) to estimate rates of return to schooling with a random coefficient model. Another example is the application done by Kwan (1991) for the estimation of interest rate sensitivity of commercial bank stock returns.

Although heterogeneous coefficient models exist for cross-sectional (Hildreth and Houck, 1968) and panel (Swamy, 1970; Hsiao, 1975) settings, such a specification has not yet been proposed for higher-dimensional data. Swamy (1970) specified an efficient estimation methodology for a model with random coefficients that change in the individual dimension, applying Aitken's Generalized Least Squares for the estimation of the coefficients. In order to make the estimator feasible, he proposed running linear regressions separately for each individual for estimating the variance components. He also developed a test to determine whether the coefficients are random or not. All in all, Swamy (1970) introduced the estimation strategy along with a hypothesis testing procedure.

Building on Hsiao (1975), Hsiao and Pesaran (2008) present two approaches for the estimation of coefficients in a Random Coefficient Model in the presence of two-dimensional panel data. The first is feasible generalized least squares (FGLS) using a minimum quadratic unbiased estimation (MINQUE) approach for estimating the variance components and the second is the maximum likelihood estimation (MLE). Interestingly, they show that this model has a Bayesian justification. The key advantage of this model lies in the fact that even allowing the coefficients to differ from unit to unit and over time, the number of parameters to be estimated is still small.

We generalise the above models in a three-dimensional setting by proposing random coefficient components in all three dimensions with *non-diagonal* variance-covariance matrices. We first discuss the application of the GLS method, paying attention to the computational aspect, as well as the estimation of the variance-covariance components. We propose two consistent methods for the latter. The first is a method that exploits either the sub-level Within variations or the overall variation of coefficients. The second is an extension of MINQUE methodology. Then we go on to examine MLE and its implementation using Anderson's (1971) algorithm. Using the asymptotic properties of the estimators, we develop tests for constant versus varying coefficients. A simulation experiment shows that FGLS performs better than OLS, but MLE beats FGLS. Furthermore, we analyse some important special issues, such as correlated components and unbalanced panels. Finally, we discuss a non-linear extension and propose a Monte-Carlo EM algorithm for its estimation.

The chapter is organized as follows: Sect. 5.2 sets out the model and presents FGLS and variance-covariance components estimation methods, Sects. 5.3 discusses MLE and Restricted MLE, Sect. 5.4 addresses inference questions, Sect. 5.5 derives predictors of the heterogenous coefficients, Sect. 5.6 presents a Bayesian approach, Sect. 5.7 discusses extensions within the linear model, Sect. 5.8 introduces non-linearity, Sect. 5.9 reports simulation results, and finally Sect. 5.10 draws some conclusions.

5.2 The Linear Model for Three Dimensions

5.2.1 The Model

We consider the following model for $t = 1,\ldots,T$ observations of $j = 1,\ldots,N_2$ individuals in $i = 1,\ldots,N_1$ groups:

$$y_{ijt} = x'_{ijt}\beta_{ijt} + \varepsilon_{ijt}\,, \tag{5.1}$$

where

$$\beta_{ijt} = (\bar{\beta} + \alpha_i + \gamma_j + \lambda_t)\,, \tag{5.2}$$

with x'_{ijt} being the $(1 \times K)$ vector of explanatory variables, $\bar{\beta}$ the $(K \times 1)$ vector of unobserved parameters, α_i, γ_j, λ_t are the $(K \times 1)$ random vectors, and finally, ε_{ijt} is an additive random disturbance. The following is assumed to hold.

Assumption 1

$$E(\varepsilon_{ijt}) = 0\,, \quad E(\varepsilon^2_{ijt}) = \sigma^2_\varepsilon\,.$$

Assumption 2

$$E(\alpha_i) = 0\,, \quad E(\gamma_j) = 0\,, \quad \textit{and} \quad E(\lambda_t) = 0$$

$$E(\alpha_i\alpha'_h) = \begin{cases} \Delta_\alpha & \textit{if} \quad i = h \\ 0 & \textit{otherwise} \end{cases}$$
$$E(\gamma_j\gamma'_l) = \begin{cases} \Delta_\gamma & \textit{if} \quad j = l \\ 0 & \textit{otherwise} \end{cases}$$
$$E(\lambda_t\lambda'_s) = \begin{cases} \Delta_\lambda & \textit{if} \quad t = s \\ 0 & \textit{otherwise} \end{cases}$$

$$E(\alpha_i\gamma'_j) = E(\alpha_i\lambda'_t) = E(\gamma_j\lambda'_t) = 0\,.$$

Note that our variance-covariance matrix is different from the ones found in the literature on RCM. We assume a *full* variance-covariance matrix for all the random components, while they are usually restricted to be diagonal (see e.g., Hsiao, 1974). A typical element of the k-th row and k'-th column of Δ_α is called $\sigma_{\alpha,kk'}$, while it is $\sigma^2_{\alpha,kk}$ when $k = k'$.

Assumption 3 *X is non-stochastic and $rank(X) = K$.*

Stacking the observations, we get:

$$y = X\bar{\beta} + X_1\alpha + X_2\gamma + X_3\lambda + \varepsilon\,, \tag{5.3}$$

with, using $\tilde{X} = \text{diag}(x'_{111},\ldots,x'_{N_1N_2T})$ of size $(N_1N_2T \times N_1N_2TK)$,

$$\begin{array}{lll} X_1 = \tilde{X}\left((I_{N_1} \otimes \iota_{N_2} \otimes \iota_T) \otimes I_K\right) & \text{of size} & (N_1 N_2 T \times N_1 K) \\ X_2 = \tilde{X}\left((\iota_{N_1} \otimes I_{N_2} \otimes \iota_T) \otimes I_K\right) & \text{of size} & (N_1 N_2 T \times N_2 K) \\ X_3 = \tilde{X}\left((\iota_{N_1} \otimes \iota_{N_2} \otimes I_T) \otimes I_K\right) & \text{of size} & (N_1 N_2 T \times TK)\,, \end{array}$$

and stacked parameters α, γ, λ of sizes $(N_1 K \times 1)$, $(N_2 K \times 1)$ and $(TK \times 1)$, respectively. The covariance matrix is then given by

$$\Omega = X_1(I_{N_1} \otimes \Delta_\alpha)X_1' + X_2(I_{N_2} \otimes \Delta_\gamma)X_2' + X_3(I_T \otimes \Delta_\lambda)X_3' + \sigma_\varepsilon^2 I_{N_1 N_2 T}\,. \tag{5.4}$$

The observations relate to different groups within which we observe cross-sectional units and for each unit we have time series information. However, it is not mandatory that the indices refer to time, individual and group. For example, in another setting, one could have workers in companies of different countries. Thus our general structure is suitable for multi-level data (perhaps with nested indices).

Assumption 1 implies that the disturbance term has constant variance along time, individuals and groups. Thus, there is no serial correlation or correlation among groups or individuals. Assumption 2 entails that the random elements added to the mean vector of coefficients are not correlated between each other but each vector has a full (non-diagonal) variance-covariance matrix as pointed out earlier. This leads to coefficients that are correlated among groups, individuals and time periods even though the random elements are not.

5.2.2 Feasible Generalized Least Squares (FGLS)

The GLS estimator of the model is given by

$$\hat{\tilde{\beta}}_{GLS} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y\,.$$

Since Ω is unknown, one needs a prior estimation of variance components for implementing GLS. But before addressing this question, we first examine a computational issue which is particularly relevant in our context due to the potentially huge size of Ω.

The dimension of Ω, whose inverse has to be calculated for obtaining the GLS estimator, is $(N_1 N_2 T \times N_1 N_2 T)$, consequently a direct computation is not generally recommended. We propose the following step-wise solution for computing its inverse.

First, let us rewrite

$$\Omega = I\sigma_\varepsilon^2 + X_1(I_{N_1} \otimes \Delta_\alpha)X_1' + X_2(I_{N_2} \otimes \Delta_\gamma)X_2' + X_3(I_T \otimes \Delta_\lambda)X_3'\,,$$

and call

$$P_1 = I\sigma_\varepsilon^2 + X_1(I_{N_1} \otimes \Delta_\alpha)X_1'\,.$$

It can be easily shown, that

$$P_1^{-1} = \text{diag}\{I_{N_2T}\sigma_\varepsilon^2 - X_i(X_i'X_i + \Delta_\alpha^{-1}\sigma_\varepsilon^2)^{-1}X_i'/\sigma_\varepsilon^2,\ i = 1\ldots N_1\},$$

with $X_i = (x_{i11}, \ldots x_{iN_2T})'$, a $(N_2T \times K)$ matrix. Now if we continue and define

$$P_2 = I\sigma_\varepsilon^2 + X_1(I_{N_1} \otimes \Delta_\alpha)X_1' + X_2(I_{N_2} \otimes \Delta_\gamma)X_2' = P_1 + X_2(I_{N_2} \otimes \Delta_\gamma)X_2',$$

its inverse can be obtained by applying the Woodbury matrix identity:

$$P_2^{-1} = P_1^{-1} - P_1^{-1}X_2R_1^{-1}X_2'P_1^{-1}\ ,$$

with

$$R_1^{-1} = \left(I_{N_2} \otimes \Delta_\gamma^{-1} + X_2'P_1^{-1}X_2\right)^{-1}$$

of size $(N_2K \times N_2K)$. Finally, using the Woodbury matrix identity again, to invert

$$\Omega = P_2 + X_3(I_T \otimes \Delta_\lambda)X_3'$$

gives

$$\Omega^{-1} = P_2^{-1} - P_2^{-1}X_3R_2^{-1}X_3'P_2^{-1}\ ,$$

with

$$R_2^{-1} = \left(I_T \otimes \Delta_\lambda^{-1} + X_3'P_2^{-1}X_3\right)^{-1}\ ,$$

a matrix of size $(TK \times TK)$.

As seen from the derivation, two inverses have to be computed directly: R_1^{-1} and R_2^{-1}, with orders N_2K and TK, respectively. Note that the inverse of P_1 can be performed analytically due to its block-diagonal nature. This observation offers a natural opportunity to reduce calculations by taking the largest random coefficient as P_1. Thus, for example, if a linked employer-employee data is such that it comprises $N_1 = 100,000$ employees, $N_2 = 1,000$ firms and $T = 10$ years with $K = 5$ regressors, the direct inverses to be taken are $(5,000 \times 5,000)$ and (50×50) matrices (doable by most computers), as opposed to the brute-force calculations on Ω, which in this case is a matrix of the order of $5,000,000,000$.

Now let us turn to to the estimation of variance components.

5.2.3 Method 1: Using Within Dimensions Variation

This method uses the variation within each dimension along the lines of Hsiao (1974). We fix one of the dimensions, say the group dimension i such that we have the following model:

$$y_{ijt} = x_{ijt}'\left(\bar{\beta} + \alpha_i\right) + x_{ijt}'\left(\gamma_j + \lambda_t\right) + \varepsilon_{ijt} \equiv x_{ijt}'\beta_i + u_{ijt}\ , \tag{5.5}$$

with the composite disturbance term

$$u_{ijt} = x_{ijt}'\left(\gamma_j + \lambda_t\right) + \varepsilon_{ijt}\ .$$

Then, the variance-covariance matrix of u_{ijt} is given by

$$\mathrm{E}\left(u_{ijt}u'_{ils}\right) = \begin{cases} 0 & \text{if} \quad j \neq l \text{ and } t \neq s \\ \sum_{k=1}^{K}\sum_{k'=1}^{K} x_{ijt,k}x_{ils,k'}\sigma_{\lambda,kk'} & \text{if} \quad j \neq l \text{ and } t = s \\ \sum_{k=1}^{K}\sum_{k'=1}^{K} x_{ijt,k}x_{ils,k'}\sigma_{\gamma,kk'} & \text{if} \quad j = l \text{ and } t \neq s \\ \sum_{k=1}^{K}\sum_{k'=1}^{K} x_{ijt,k}x_{ils,k'}\left(\sigma_{\lambda,kk'} + \sigma_{\gamma,kk'}\right) + \sigma_{\varepsilon}^2 & \text{if} \quad j = l \text{ and } t = s\,. \end{cases}$$

As the variance-covariance matrix is non-diagonal, we cannot follow Hsiao's (1974) approach from here on. Instead, we propose the following new procedure.

First, we can estimate the u_{ijt} error term of model (5.5) by running an OLS separately for each individual i

$$r_{ijt} = y_{ijt} - x'_{ijt}\hat{\beta}_i\,,$$

with

$$\hat{\beta}_i = \left(X'_iX_i\right)X'_iy_i\,, \quad y_i = \left(y_{i11}, y_{i12}, \ldots, y_{iN_2T}\right)'\,, \quad \text{and} \quad X_i = \left(x_{i11}, x_{i12}, \ldots, x_{iN_2T}\right)'\,.$$

Stacking over j and t, the residual can be expressed as

$$r_i = M_{X_i}u_i\,,$$

where the idempotent matrix M_{X_i} is

$$M_{X_i} = I_{N_2T} - X_i\left(X'_iX_i\right)^{-1}X'_i\,.$$

Taking the expectation of $r_ir'_i$ we have

$$\mathrm{E}\left(r_ir'_i\right) = M_{X_i}\mathrm{E}\left(u_iu'_i\right)M_{X_i} = M_{X_i}\Omega_iM_{X_i}\,, \tag{5.6}$$

where we know that the variance-covariance matrix can be written as

$$\Omega_i = \sum_{k=1}^{K}\sum_{k'=1}^{K}\sigma_{\gamma,kk'}H_{i_{\gamma,kk'}} + \sum_{k=1}^{K}\sum_{k'=1}^{K}\sigma_{\lambda,kk'}H_{i_{\lambda,kk'}} + \sigma_{\varepsilon}^2I_{N_2T} \tag{5.7}$$

and

$$H_{i_{\gamma,kk'}} = \tilde{X}_{i,k}\left(I_{N_2}\otimes\iota_T\right)\left(I_{N_2}\otimes\iota'_T\right)\tilde{X}'_{i,k'}\,; \qquad H_{i_{\lambda,kk'}} = \tilde{X}_{i,k}\left(I_T\otimes\iota_{N_2}\right)\left(I_T\otimes\iota'_{N_2}\right)\tilde{X}'_{i,k'}\,.$$

Thus, the expectation of the outer product of the vector of residuals can be expressed as:

$$\mathrm{E}\left(r_ir'_i\right) = \sum_{k=1}^{K}\sum_{k'=1}^{K}\sigma_{\gamma,kk'}M_{X_i}H_{i_{\gamma,kk'}}M_{X_i} + \sum_{k=1}^{K}\sum_{k'=1}^{K}\sigma_{\lambda,kk'}M_{X_i}H_{i_{\lambda,kk'}}M_{X_i} + \sigma_{\varepsilon}^2M_{X_i}\,. \tag{5.8}$$

Using the *vec* operator (where $\mathrm{vec}(A)$ gives the $(NM\times 1)$ column vector built up from arranging each column of the $(N\times M)$ matrix A below each other) and the *vech*

operator (where vech(G) gives the $(N(N+1)/2\times 1)$ column vector built up from stacking up the different elements of each column of the $(N\times N)$ symmetric matrix G)

$$\begin{aligned}\operatorname{vec}(\mathrm{E}(r_i r_i')) &= \textstyle\sum_{k=1}^{K}\sum_{k'=1}^{K}\sigma_{\gamma,kk'}\operatorname{vec}(M_{X_i}H_{i_{\gamma,kk'}}M_{X_i}) \\ &\quad + \textstyle\sum_{k=1}^{K}\sum_{k'=1}^{K}\sigma_{\lambda,kk'}\operatorname{vec}(M_{X_i}H_{i_{\lambda,kk'}}M_{X_i}) + \sigma_\varepsilon^2\operatorname{vec}(M_{X_i}) \\ &= A_{\gamma,i}\cdot\operatorname{vec}(\Delta_\gamma) + A_{\lambda,i}\cdot\operatorname{vec}(\Delta_\lambda) + \sigma_\varepsilon^2\operatorname{vec}(M_{X_i}) \\ &= A_{\gamma,i}D\cdot\operatorname{vech}(\Delta_\gamma) + A_{\lambda,i}D\cdot\operatorname{vech}(\Delta_\lambda) + \sigma_\varepsilon^2\operatorname{vec}(M_{X_i})\end{aligned} \tag{5.9}$$

for $i=1,\ldots,N_1$, with

$$\begin{aligned}A_{\gamma,i} &= \Big(\operatorname{vec}(M_{X_i}H_{i_{\gamma,11}}M_{X_i}),\ldots,\operatorname{vec}(M_{X_i}H_{i_{\gamma,kk}}M_{X_i})\Big) \\ A_{\lambda,i} &= \Big(\operatorname{vec}(M_{X_i}H_{i_{\lambda,11}}M_{X_i}),\ldots,\operatorname{vec}(M_{X_i}H_{i_{\lambda,kk}}M_{X_i})\Big) \\ D &= \textstyle\sum_{i\geqq j}\operatorname{vec}(L_{ij})u_{ij}' \\ L_{ij} &= E_{ji}+E_{ij} \\ E_{ij} &= e_i e_j' ,\end{aligned}$$

where e_i is a unit vector with one in the ith position and zeros otherwise, u_{ij} is a vector of order $\frac{1}{2}n(n+1)$ with unity in its $[(j-1)n+i-\frac{1}{2}j(j-1)]$-th position and zeros otherwise and D is the duplication matrix of size $(K^2\times K(K+1)/2)$ with full column-rank.

Writing (5.9) as

$$R_i = Z_i\eta \qquad \text{for} \quad i=1,\ldots,N_1\,, \tag{5.10}$$

with

$$R_i = \operatorname{vec}\big(\mathrm{E}(r_i r_i')\big)\,; \quad Z_i = \big(A_{\gamma,i}D, A_{\lambda,i}D, \operatorname{vec}(M_i)\big) \quad \text{with rank } 2K(K+1)/2+1$$

and

$$\eta' = (\operatorname{vech}(\Delta_\gamma)'\operatorname{vech}(\Delta_\lambda)'\sigma_\varepsilon^2)\,,$$

we can estimate the variance-covariance elements as

$$\hat{\eta}^{(i)} = \big(Z_i'Z_i\big)^{-1}Z_i'\hat{R}_i \qquad \text{for} \quad i=1,\ldots,N_1\,. \tag{5.11}$$

This is equivalent having the following setting:

$$\hat{R}_i = Z_i\eta + \mu_i\,, \tag{5.12}$$

where μ_i is an error term that represents the deviation of $\hat{R}_i$ from $\mathrm{E}(r_i r_i')$. For simplicity reasons, we use $r_i r_i'$ as an estimator of $\mathrm{E}(r_i r_i')$. This is unbiased, although not efficient, and additionally, it retrieves good final estimators of the coefficients. We could have also used the sample estimator

$$\widehat{\mathrm{E}(r_i r_i')} = \frac{1}{N_1-1}\sum_{i=1}^{N_1}(r_i-\bar{r}_i)(r_i-\bar{r}_i)'\,, \tag{5.13}$$

but this estimated matrix would be invariant over i, and so may not be appropriate for the regression (5.12).

Formulas (5.11)-(5.13) lead to N_1 estimators for Δ_γ, Δ_λ and σ_ε^2, so in order to use information more effectively, one can go ahead and take their means over i to get

$$\begin{aligned}\hat{\Delta}_\gamma &= \tfrac{1}{N_1}\textstyle\sum_{i=1}^{N_1}\hat{\Delta}_\gamma^{(i)}\\ \hat{\Delta}_\lambda &= \tfrac{1}{N_1}\textstyle\sum_{i=1}^{N_1}\hat{\Delta}_\gamma^{(i)}\\ \hat{\sigma}_\varepsilon^2 &= \tfrac{1}{N_1}\textstyle\sum_{i=1}^{N_1}\hat{\sigma}_\varepsilon^{2(i)}\,.\end{aligned} \tag{5.14}$$

We can repeat the same procedure for the other two dimensions, fixing them one by one. This process gives two estimators for each of the variance components Δ_γ, Δ_λ and Δ_α. We can once again take their means in order to have a single estimator at the end. Similarly, σ_ε^2 has three estimations (one from each estimation), so again we can do the trick and take their overall mean to get a single estimator. It can easily be shown that all the listed estimators are consistent.

5.2.4 Method 2: Using the Overall Variation

In this section, we propose another method for the estimation of the variance-covariance matrix which simultaneously estimates all three components using the residuals from the pooled OLS written as

$$r_{ijt} = y_{ijt} - x_{ijt}'\hat{\beta}\,, \tag{5.15}$$

where

$$\hat{\beta} = \left(X'X\right)^{-1}X'y \tag{5.16}$$

Stacking observations over i, j and t, the residual can be expressed as

$$r = M_X u \quad \text{with} \quad M_X = I_{N_1N_2T} - X\left(X'X\right)^{-1}X'\,. \tag{5.17}$$

Taking the expectation of rr', we have

$$\mathrm{E}\left(rr'\right) = M_X\mathrm{E}\left(uu'\right)M_X = M_X\Omega M_X\,. \tag{5.18}$$

Just as in Sect. 5.2.3, the variance-covariance matrix Ω can be written as

$$\Omega = \sum_{k=1}^{K}\sum_{k'=1}^{K}\sigma_{\alpha,kk'}H_{\alpha,kk'} + \sum_{k=1}^{K}\sum_{k'=1}^{K}\sigma_{\gamma,kk'}H_{\gamma,kk'} + \sum_{k=1}^{K}\sum_{k'=1}^{K}\sigma_{\lambda,kk'}H_{\lambda,kk'} + \sigma_\varepsilon^2 I_{N_1N_2T}\,, \tag{5.19}$$

where

$$\begin{aligned}H_{\alpha,kk'} &= \tilde{X}_k(I_{N_1}\otimes\iota_{N_2}\otimes\iota_T)(I_{N_1}\otimes\iota_{N_2}'\otimes\iota_T')\tilde{X}_{k'}'\\ H_{\gamma,kk'} &= \tilde{X}_k(\iota_{N_1}\otimes I_{N_2}\otimes\iota_T)(\iota_{N_1}'\otimes I_{N_2}\otimes\iota_T')\tilde{X}_{k'}'\\ H_{\lambda,kk'} &= \tilde{X}_k(\iota_{N_1}\otimes\iota_{N_2}\otimes I_T)(\iota_{N_1}'\otimes\iota_{N_2}'\otimes I_T)\tilde{X}_{k'}'\end{aligned}$$

and $\tilde{X_k}$ is the $(N_1N_2T \times N_1N_2T)$ diagonal matrix constructed from the k-th column of the data matrix X. As before, the expectation of the outer product of the vector of residuals can be expressed as

$$\begin{aligned}\mathrm{E}(rr') = &\textstyle\sum_{k=1}^{K}\sum_{k'=1}^{K}\sigma_{\alpha,kk'}M_X H_{\alpha,kk'}M_X + \sum_{k=1}^{K}\sum_{k'=1}^{K}\sigma_{\gamma,kk'}M_X H_{\gamma,kk'}M_X \\ &+\textstyle\sum_{k=1}^{K}\sum_{k'=1}^{K}\sigma_{\lambda,kk'}M_X H_{\lambda,kk'}M_X + \sigma_\varepsilon^2 M_X\ .\end{aligned} \tag{5.20}$$

Applying the *vec* operator and *vech* operator

$$\begin{aligned}\mathrm{vec}(\mathrm{E}(rr')) &= \textstyle\sum_{k=1}^{K}\sum_{k'=1}^{K}\sigma_{\alpha,kk'}\mathrm{vec}(M_X H_{\alpha,kk'}M_X) \\ &\quad+\textstyle\sum_{k=1}^{K}\sum_{k'=1}^{K}\sigma_{\gamma,kk'}\mathrm{vec}(M_X H_{\gamma,kk'}M_X) \\ &\quad+\textstyle\sum_{k=1}^{K}\sum_{k'=1}^{K}\sigma_{\lambda,kk'}\mathrm{vec}(M_X H_{\lambda,kk'}M_X) + \sigma_\varepsilon^2\mathrm{vec}(M_X) \\ &= B_\alpha\mathrm{vec}(\Delta_\alpha) + B_\gamma\mathrm{vec}(\Delta_\gamma) + B_\lambda\mathrm{vec}(\Delta_\lambda) + \sigma_\varepsilon^2\mathrm{vec}(M_X) \\ &= B_\alpha D\mathrm{vech}(\Delta_\alpha) + B_\gamma D\mathrm{vech}(\Delta_\gamma) + B_\lambda D\mathrm{vech}(\Delta_\lambda) + \sigma_\varepsilon^2\mathrm{vec}(M_X)\ ,\end{aligned}$$

with the appropriate definitions

$$\begin{aligned}B_\alpha &= (\mathrm{vec}(M_X H_{\alpha,11}M_X), \ldots, \mathrm{vec}(M_X H_{\alpha,KK}M_X)) \\ B_\gamma &= \big(\mathrm{vec}(M_X H_{\gamma,11}M_X), \ldots, \mathrm{vec}(M_X H_{\gamma,KK}M_X)\big) \\ B_\lambda &= \big(\mathrm{vec}(M_X H_{\lambda,11}M_X), \ldots, \mathrm{vec}(M_X H_{\lambda,KK}M_X)\big)\ .\end{aligned}$$

From this, the variance-covariance components can be estimated as

$$\hat{\eta} = \left(Z'Z\right)^{-1}Z'\hat{R} \tag{5.21}$$

with

$$\begin{aligned}\hat{R} &= \mathrm{vec}\,(rr') \\ Z &= \big(B_\alpha D,\ B_\gamma D,\ B_\lambda D,\ \mathrm{vec}(M_X)\big) \quad \text{of rank} \quad 3K(K+1)/2+1 \\ D &\text{ is the duplication matrix defined in Sect. 5.2.3} \\ \hat{\eta}' &= \big(\mathrm{vech}(\hat{\Delta}_\alpha)'\ \mathrm{vech}(\hat{\Delta}_\lambda)'\ \mathrm{vech}(\hat{\Delta}_\gamma)'\ \hat{\sigma}_\varepsilon^2\big)\end{aligned}$$

This is equivalent to having the following setting

$$\hat{R} = Z\eta + \mu\ , \tag{5.22}$$

where μ is the error term representing the deviation of rr' from $\mathrm{E}(rr')$. It can be shown that the above procedure leads to consistent estimators of Δ_α, Δ_γ, Δ_λ and σ_ε^2. The caveat of the method above is that it does not necessarily lead to a positive definite estimated variance-covariance matrix. In order to overcome this problem, we propose to do a re-parametrization. We set the diagonal elements as

$$\sigma_{\alpha,kk}^2 = \exp\{\tau_{\alpha,kk}\} \quad k \in 1,2,\ldots,K\ ,$$

similarly for γ, λ and σ_ε^2. Let us denote by ϑ the new vector of parameters with all the elements of the variance-covariance matrix as before, except for the diagonal components that are replaced by the new parametrization. We propose a constrained

non-linear least squares estimation by minimizing the sum of squared errors under positivity restrictions, as positive diagonal elements are not sufficient to guarantee positive definiteness. Thus, the problem to be solved is

$$\begin{array}{l} \min\limits_{\tau_l,\sigma_{l,kk}} (R-Z\vartheta)'(R-Z\vartheta)\,, \quad \text{such that} \\ \quad \Delta_\alpha,\, \Delta_\gamma,\, \Delta_\lambda \quad \text{are positive definite}\,. \end{array} \tag{5.23}$$

Following Benson and Vanderbei (2003), we can use an interior-point method for non-linear programming to solve our problem. In this case, we apply the spectral decomposition to each matrix, and our restrictions are equivalent to imposing that all the eigenvalues of each matrix are positive. Although implementing the programming exercise (5.23) would take us beyond the scope of the chapter, it should be noted that our simulation experiment in Sect. 5.9 produces a relatively high proportion of non-positive estimated variance-covariance matrices.

5.2.5 *Minimum Norm Quadratic Unbiased Estimation (MINQUE)*

In this section, we apply the Minimum Norm Quadratic Unbiased Estimation (hereafter, MINQUE) method (see Rao, 1970) for estimating the variance components of our model. Fixing the group and individual dimensions, and letting the coefficients vary randomly over time, we can write

$$y_{ij} = X_{ij}(\beta + \alpha_i + \gamma_j) + X_{3,ij}\lambda + \varepsilon_{ij} \equiv X_{ij}\beta_{ij} + u_{ij}\,, \tag{5.24}$$

with y_{ij}, X_{ij} and $X_{3,ij}$ being the $(T\times 1)$, $(T\times K)$, and $(T\times TK)$ matrices of y, X and X_3 for individual pair ij. Now, the variance-covariance matrix is

$$\Omega_{ij} = X_{3,ij}(I_T \otimes \Delta_\lambda)X'_{3,ij} + \sigma_\varepsilon^2 I_T\,. \tag{5.25}$$

We define the problem as that of finding the matrix A that minimizes $||U'AU||$ with $U = (X_{3,ij}, I_T)$, subject to

1. Unbiasedness condition: $\mathrm{E}(y'_{ij}Ay_{ij}) = \mathrm{tr}(I_T \otimes \Delta_\lambda)(X_{3,ij}AX_{3,ij}) + \sigma_\varepsilon^2\mathrm{tr}(A)$
2. Invariability condition: $X'_{3,ij}AX_{3,ij} = 0$.

Using the results of Rao (1970), the matrix A is given by:

$$A = R_{ij}\tilde{X}_{ij}\Lambda\tilde{X}'_{ij}R_{ij}\,, \tag{5.26}$$

with Λ as a solution of $\tilde{X'}_{ij}R_{ij}I_T\Lambda I_T R_{ij}\tilde{X}_{ij} + R_{ij}\tilde{X'}_{ij}\Lambda\tilde{X}_{ij}R_{ij} = M$.

Using this, the MINQUE estimators of Δ_λ and σ_ε^2 are given by the solution to the following system of equations:

$$\begin{array}{l} v_{ij}v'_{ij} = \tilde{X}'_{ij}R_{ij}\tilde{X}_{ij}(I_T \otimes \Delta_\lambda)\tilde{X}'_{ij}R_{ij}\tilde{X}_{ij} + \sigma_\varepsilon^2\tilde{X}'_{ij}R_{ij}R_{ij}\tilde{X}_{ij} \\ u'_{ij}u_{ij} = \mathrm{tr}R_{ij}\tilde{X}_{ij}(I_T \otimes \Delta_\lambda)\tilde{X}'_{ij}R_{ij} + \sigma_\varepsilon^2\mathrm{tr}R_{ij}R_{ij}\,, \end{array} \tag{5.27}$$

where

$$R_{ij} = F_{ij}^{-1}(I - P_{ij}) \,, \quad P_{ij} = X_{ij}(X_{ij}' F_{ij}^{-1} X_{ij})^{-} X_{ij}' F_{ij}^{-1} \,,$$
$$v_{ij} = \tilde{X}_{ij}' R_{ij} Y_{ij} \,, \quad \text{and} \quad u_{ij} = R_{ij} Y_{ij} \,.$$

And we can use two matrices for F_{ij}, the first approximates Ω_{ij} by taking $\Delta_\lambda = I_K$ and $\sigma_\varepsilon^2 = 1$ and the second uses an initial estimate of it:

$$F_{ij} = \sum_k \sum_{k'} \tilde{X}_{ij,k} \tilde{X}_{ij,k'}' + I_{ij}$$
$$F_{ij} = \Omega_{o,ij} \,.$$

We thus obtain $N_1 N_2$ estimators of Δ_λ and σ_ε^2, and take their average as our final estimator. Finally, the procedure is repeated for each dimension.

According to Callanan (1985), one problem of MINQUE estimation is that in spite of being unbiased, it can retrieve estimators that are not within the parameter space. This means that we can end up with a non-positive definite matrix, facing the same problem as we did in the previous section. A possible solution to this is obtaining the estimators under a constraint for keeping the estimations in the parameter space, which ensures that the variance-covariance matrix is positive definite. Nevertheless, the drawback of this approach is that the estimators may be biased.

5.2.6 *Properties of the Estimators*

Given that the elements of the variance-covariance matrix can be estimated in a consistent way, we can easily show that the FGLS estimator of $\bar{\beta}$ is consistent with the asymptotic variance given by $\text{Var}(\hat{\bar{\beta}}) = \left(X' \Omega^{-1} X\right)^{-1}$. Recall that for Hsiao's (1974) 2D model both N and T asymptotics are required, whereas in our case the growth of *any two* indices is sufficient.[1] This is a huge improvement over earlier results, as consistent estimation is now feasible for short panels as well, with T fixed and small. Further, if $N_1, N_2 \to \infty$ holds, $\hat{\bar{\beta}}$ is asymptotically normally distributed with

$$\sqrt{N_1 N_2} \left(\hat{\bar{\beta}} - \bar{\beta} \right) \sim N \left(0, \left(X' \Omega^{-1} X \right)^{-1} \right) \,. \tag{5.28}$$

[1] Strictly speaking, $N_2 \to \infty$ *or* $T \to \infty$ is sufficient for the consistent estimation of $\hat{\sigma}_{\alpha k}^2$, $N_1 \to \infty$ *or* $T \to \infty$ is sufficient for the consistent estimation of $\hat{\sigma}_{\gamma k}^2$, and $N_1 \to \infty$ *or* $N_2 \to \infty$ for the consistent estimation of $\hat{\sigma}_{\lambda k}^2$. Putting these three conditions together gives the overall condition.

5.3 Maximum Likelihood Estimation

5.3.1 The Unrestricted Maximum Likelihood

Under the assumption that the error term and the random components follow a normal distribution with zero mean and σ_ε^2, Δ_α, Δ_γ and Δ_λ respective variance and variance-covariance matrices, the log-likelihood is given by

$$\ln L(\beta, \sigma_\alpha, \sigma_\gamma, \sigma_\lambda, \sigma_\varepsilon | y) = -\frac{N_1 N_2 T}{2} \ln(2\pi) - \frac{1}{2}|\Omega| - \frac{1}{2}(y - X\beta)' \Omega^{-1} (y - X\beta) \,. \tag{5.29}$$

Following Hsiao (1974), we can express the variance-covariance matrix as a linear combination of the variance components as

$$\begin{aligned} \Omega \; = & \textstyle\sum_{k=1}^{K} \sigma_{\alpha,kk}^2 H_{\alpha,kk} + \sum_{k,k';k\neq k'} \sigma_{\alpha,k'k} (H_{\alpha,k'k} + H_{\alpha,kk'}) \\ & + \textstyle\sum_{k=1}^{K} \sigma_{\gamma,kk}^2 H_{\gamma,kk} + \sum_{k,k';k\neq k'} \sigma_{\gamma,k'k} (H_{\gamma,k'k} + H_{\gamma,kk'}) \\ & + \textstyle\sum_{k=1}^{K} \sigma_{\lambda,kk}^2 H_{\lambda,kk} + \sum_{k,k';k\neq k'} \sigma_{\lambda,k'k} (H_{\lambda,k'k} + H_{\lambda,kk'}) \\ & + \sigma_\varepsilon^2 I_{N_1 N_2 T} \,, \end{aligned} \tag{5.30}$$

which is naturally equivalent to (5.19) derived in Sect. 5.2.4, only now are diagonal elements in Δ_α etc. are taken out from the sums. Now, following Anderson (1971), the term $(y - X\beta)' \Omega^{-1} (y - X\beta)$ can be re-expressed as $\mathrm{tr}(\Omega^{-1})(y - X\beta)'(y - X\beta)$. The first order conditions of the likelihood function, using expression (5.30) are given by

$$\begin{aligned} X' \Omega^{-1} X\beta &= X' \Omega^{-1} y \\ \mathrm{tr}(\Omega^{-1}) H_{m,kk} &= \mathrm{tr}(\Omega^{-1}) H_{m,kk} \Omega^{-1} (y - X\beta)'(y - X\beta) \\ \mathrm{tr}(\Omega^{-1})(H_{m,k'k} + H_{m,kk'}) &= \mathrm{tr}(\Omega^{-1})(H_{m,k'k} + H_{m,kk'}) \Omega^{-1} (y - X\beta)'(y - X\beta) \\ \mathrm{tr}(\Omega^{-1}) &= \mathrm{tr}(\Omega^{-1}) \Omega^{-1} (y - X\beta)'(y - X\beta) \end{aligned} \tag{5.31}$$

with the index m referring to α, γ or λ, giving a system of total 8 equations. As the system in (5.31) is non-linear, we have to linearize it using, for example, Anderson's (1971) algorithm. Solutions are then collected from solving the linear system of equations for β and the variance-covariance components.

As previously mentioned, a major computational problem is the inversion of the variance-covariance matrix. Luckily, the step-wise procedure to construct the inverse, presented in Sect. 5.2.2, can also be applied for this MLE case.

Another MLE can be designed by fixing two of the dimensions for the random components and estimate the variance-covariance elements for the remaining one. For instance, if we fix time and individual j, that is components λ_t and γ_j, the variance covariance matrix is given by

$$X_1 \left(I_{N_1} \otimes \Delta_\alpha \right) X_1' + \sigma_\varepsilon^2 I_{N_2 T} \,. \tag{5.32}$$

The likelihood function is written as before, and the FOC equations are solved to obtain Δ_α and σ_ε^2. This process should be repeated for each dimension. As a result, we have three estimates for σ_ε^2 and once again we can take their average.

The MLE estimator has the usual drawbacks that it relies on a parametric specification of the data generating process and the variance covariance estimators are not corrected for the loss of degrees of freedom. Notwithsanding, Harville (1977) argues that even when the true distribution is not normal, the estimations obtained under this assumption may still be suitable.

5.3.2 Restricted Maximum Likelihood Estimation

Patterson and Thompson (1971) developed an alternative method called the Restricted Maximum Likelihood Estimation (RMLE) to obtain unbiased estimators. This method maximizes the likelihood over a restricted parameter space by splitting the log-likelihood into two orthogonal components such that we can estimate the elements of the variance-covariance matrix independently of those of $\bar{\beta}$. We apply this to our model.

First, we rewrite the probability density function of y as:

$$f(y|\bar{\beta},\Omega) = f(U|\Omega) f(\hat{\beta}_G|\bar{\beta},\Omega)|J_{h^{-1}(y)}|\,, \tag{5.33}$$

where U is defined as

$$U = L'y \quad \text{with} \quad LL' = I - X(X'X)^{-1}X' = M_X \quad \text{and} \quad L'L = I\,, \tag{5.34}$$

and

$$J_{h^{-1}(y)} = \frac{\partial f(U,\hat{\beta})}{\partial y}\,. \tag{5.35}$$

From the properties of L ,it is implied that U is independent of $\hat{\beta}_G$, as U lies in the orthogonal column space of X.

The estimation of the variance-covariance components is carried out using the conditional likelihood $L(\Omega|U)$, first by observing

$$f(U|\Omega) = \frac{f(y|\bar{\beta},\Omega)}{f(\hat{\beta}_G|\bar{\beta},\Omega)}|J_{h^{-1}(y)}|^{-1}\,, \tag{5.36}$$

and that

$$\begin{aligned} f(U|\Omega) = {} & (2\pi)^{(-N_1N_2T+K)}|\Omega|^{-1/2}|(X'(\Omega)^{-1}X)^{-1}|^{-1/2}\times \\ & \times \exp\left\{-\tfrac{1}{2}(Y - X\hat{\beta}_G)'\Omega^{-1}(Y - X\hat{\beta}_G)\right\}\,, \end{aligned} \tag{5.37}$$

which in turn gives

$$\begin{aligned}\ln L(\Omega|U) = &(-N_1N_2T+K)\ln(2\pi)-(1/2)\ln|(\Omega)|-(1/2)\times \\ &\times\ln|(X'(\Omega)^{-1}X)|-1/2(y-X\hat{\beta}_G)'\Omega^{-1}(y-X\hat{\beta}_G)\,.\end{aligned} \tag{5.38}$$

In line with Callanan (1985), the solution to the FOC can be obtained using various algorithms, such as the *EM* algorithm, the *Newthon–Raphson* method or the method of *Scoring*. According to Jiang (1996), the estimators are asymptotically normal and have similar properties to the ones obtained by MLE under some regularity conditions; however, the rates of convergence are already corrected for the loss of degrees of freedom for this restricted case.

5.4 Inference: Varying (Random) or Constant Coefficients?

In order to use the random coefficient specification, it is necessary to test whether the variance-covariance matrices for the random components are significantly different from zero, that is, (some or all of them) should be treated as random variables, or simply as constants. In this section, we propose several specification tests to investigate this. Different testing procedures are derived for the different underlying estimation methods used for the variance-covariance components. Such tests employ nulls, like $\Delta_\alpha = 0$ and/or $\Delta_\lambda = 0$ and/or $\Delta_\gamma = 0$.

5.4.1 Testing for Methods 1 and 2

Suppose we want to test the following null:

$$H_0: \quad \Delta_\lambda = 0\,,$$

assuming that Δ_γ, Δ_α and σ_ε^2 are known. In order to test this hypothesis, we need an estimator for Δ_λ, so let's write model (5.1) as

$$y_{ijt} = x'_{ijt}\left(\bar{\beta}+\lambda_t\right)+x'_{ijt}\alpha_i+x'_{ijt}\gamma_j+\varepsilon_{ijt} \equiv x'_{ijt}\beta_t+u_{ijt}\,, \tag{5.39}$$

with an obvious change of notations in the second part of the equation. Fixing t, we can estimate β_t with the number of N_1N_2 observations for each period

$$\hat{\beta}_t = \left(X'_tX_t\right)^{-1}X'_ty_t\,,$$

where

$$X_t = (x_{11t},\,\ldots,\,x_{N_1N_2t})'\,,$$

similarly for y. It can be shown that the above estimator is consistent for N_1, N_2 or both, and asymptotically normally distributed

$$\sqrt{N_1N_2}(\hat{\beta}_t-\beta_t) \sim N(0,\Sigma_t)\,,$$

with

$$\Sigma_t = \left(X_t'X_t\right)^{-1} X_t' \left(X_{1,t}(I_{N_1} \otimes \Delta_\alpha)X_{1,t}' + X_{2,t}(I_{N_2} \otimes \Delta_\gamma)X_{2,t}' + \sigma_\varepsilon^2 I_{N_1N_2}\right) X_t \left(X_t'X_t\right)^{-1}$$

with $X_{1,t}$ and $X_{2,t}$ being the $(N_1N_2 \times N_1K)$ and $(N_1N_2 \times N_2K)$ sub-matrices of X_1 and X_2 where only the N_1N_2 rows corresponding to time t are preserved.

Given that $\text{Var}(\beta_t) = \Delta_\lambda$, a feasible candidate for estimating Δ_λ can be constructed as follows

$$\begin{aligned}\hat{\Delta}_\lambda = {} & \tfrac{1}{T-1}\left(\textstyle\sum_{t=1}^T \hat{\beta}_t\hat{\beta}_t' - \tfrac{1}{T}\sum_{t=1}^T \hat{\beta}_t \sum_{t=1}^T \hat{\beta}_t'\right) - \tfrac{1}{T}\sigma_\varepsilon^2 \left(X_t'X_t\right)^{-1} \\ & - \tfrac{1}{T-1}\textstyle\sum_{t=1}^T \left(X_t'X_t\right)^{-1} X_t' \left(X_{1,t}(I_{N_1} \otimes \Delta_\alpha)X_{1,t}' + X_{2,t}(I_{N_2} \otimes \Delta_\gamma)X_{2,t}'\right) X_t \left(X_t'X_t\right)^{-1} \\ & + \tfrac{1}{T(T-1)}\textstyle\sum_{t=1}^T \left(X_t'X_t\right)^{-1} X_t'X_{1,t}(I_{N_1} \otimes \Delta_\alpha) \sum_{t=1}^T X_{1,t}'X_t \left(X_t'X_t\right)^{-1} \\ & + \tfrac{1}{T(T-1)}\textstyle\sum_{t=1}^T \left(X_t'X_t\right)^{-1} X_t'X_{2,t}(I_{N_2} \otimes \Delta_\gamma) \sum_{t=1}^T X_{2,t}'X_t \left(X_t'X_t\right)^{-1} .\end{aligned} \tag{5.40}$$

Note that all the terms after the first one are correcting for the bias. It can be shown that the above estimator follows a *Wishart* distribution provided Δ_γ, Δ_α and σ_ε^2 are all known. However, in practice we rarely know their true values, hence we can hardly implement any test based on this *Wishart* distribution. If on the other hand, we replace the unknown variance-covariance elements by their consistently estimated counterparts, we end up with a sum of correlated random variables whose distribution is unknown. Consequently, it is impossible to directly test for $H_0 : \Delta_\lambda = 0$. Similar reasoning holds for Δ_α and Δ_γ.

5.4.1.1 Test in Steps

The failure of the direct test leads us to propose an alternative testing procedure which assesses the equality of the coefficients (against variability) taking a "fixed-effects" perspective. Following Hsiao (1974), we propose a test in steps: first, test variability in one of the dimensions, specifying the heterogeneity in this particular dimension as varying coefficients while removing the other coefficient components to the composite error term, as done in (5.39). Specifically, our first null is given by

$$H_0: \quad \beta_1 = \beta_2 = \cdots = \beta_T \tag{5.41}$$

for model (5.39).

Note that H_0 can be written as $R_1\beta_{(t)} = 0$, where

$$\beta_{(t)} = (\beta_1, \beta_2, \ldots, \beta_T)'$$

and

$$R_1 = \begin{pmatrix} I_K & -I_K & 0 & 0 & 0 & 0 \\ 0 & I_K & -I_K & 0 & 0 & 0 \\ & & & \ddots & & \\ 0 & 0 & 0 & 0 & I_K & -I_K \end{pmatrix} \quad \text{of size} \quad (K(T-1) \times KT)$$

is a matrix of ones and zeros aiding to formulate the $K(T-1)$ linear constraints. With this clever notation we have the following statistic

$$\hat{\beta}'_{(t)} R'_1 \left(R_1 \Sigma_1 R'_1\right)^{-1} R_1 \hat{\beta}_{(t)} \sim \chi^2_{K(T-1)} \,,$$

where Σ_1 is the asymptotic variance of $\hat{\beta}_{(t)}$.

If we do not reject the null, we can go on to test the variability of the effects over the second dimension, say j. Once again we test this through assessing the equality of the coefficients across this dimension on model

$$y_{ijt} = x'_{ijt}(\bar{\beta} + \gamma_j) + x'_{ijt}(\alpha_i) + \varepsilon_{ijt} = x'_{ijt}\beta_j + u_{ijt} \,, \tag{5.42}$$

with the following null hypothesis:

$$H_0 : \quad \beta_1 = \beta_2 = \cdots = \beta_{N_2} \,.$$

Note that now λ_t is eliminated from the model as the result of the first test.

The OLS estimator for each individual j is given by

$$\hat{\beta}_j = \left(X'_j X_j\right)^{-1} X'_j y_j \,, \tag{5.43}$$

where

$$X_j = \left(x_{1j1}, \ldots, x_{N_1 jT}\right)' \,,$$

similarly for y_j. The variance-covariance matrix of estimator (5.43), under the null hypothesis, and given the non-rejection of the first test, is

$$\Sigma_j = \left(X'_j X_j\right)^{-1} X'_j \left(X_{1,j}(I_{N_1} \otimes \Delta_\alpha) X'_{1,j} + \sigma^2_\varepsilon I_{N_1 T}\right) X_j \left(X'_j X_j\right)^{-1} \,, \tag{5.44}$$

where $X_{1,j}$ is the $(N_1 T \times N_1 K)$ sub-matrix of X_1 with all columns and rows only corresponding to individual j. Stacking all N_2 individual estimators, the vector

$$\hat{\beta}_{(j)} = \left(\hat{\beta}_{.1.}, \hat{\beta}_{.2.}, \ldots, \hat{\beta}_{.N_2.}\right)'$$

is asymptotically distributed as

$$\sqrt{N_1 T}(\hat{\beta}_{(j)} - \beta_{(j)}) \sim N(0, \Sigma_2) \,,$$

with which the test statistic is constructed as,

$$\hat{\beta}'_{(j)} R'_2 \left(R_2 \Sigma_2 R'_2\right)^{-1} R_2 \hat{\beta}_{(j)} \sim \chi^2_{(N_2-1)K} \,,$$

where

$$R_2 = \begin{pmatrix} I_K & -I_K & 0 & 0 & 0 & 0 \\ 0 & I_K & -I_K & 0 & 0 & 0 \\ & & & \ddots & & \\ 0 & 0 & 0 & 0 & I_K & -I_K \end{pmatrix} \quad \text{of size} \quad (K(N_2-1) \times KN_2) \, .$$

Finally, if we fail to reject the above null, we are left with testing the variability over the last remaining dimension i.e., at the group level i in an exactly similar manner to the other two. The null hypothesis on model

$$y_{ijt} = x'_{ijt}(\bar{\beta} + \alpha_i) + \varepsilon_{ijt} = x'_{ijt}\beta_i + \varepsilon_{ijt} \tag{5.45}$$

is

$$H_0: \quad \beta_1 = \beta_2 = \cdots = \beta_{N_1} \, .$$

Under the null hypothesis, pooling provides a consistent way of estimating the vector of parameters. Thus, the following statistic can be constructed:

$$\sum_{i}^{N_1} \frac{\left(\hat{\beta}_i - \hat{\beta}\right)' (X_i'X_i)^{-1} \left(\hat{\beta}_i - \hat{\beta}\right)}{\hat{\sigma}_{(i)}} \, , \tag{5.46}$$

where

$$X_i = (x_{i11}, \ldots, x_{iN_2T})' \qquad \hat{\beta} = (X'X)^{-1}X'y \quad \text{and} \quad \hat{\sigma}_{(i)} = \frac{(y_i - X_i\beta_i)'(y_i - X_i\beta_i)}{N_2T - K} \, .$$

The asymptotic distribution is χ^2 with degrees of freedom equal to $(N_1 - 1)K$.

Naturally, the order of the testing is arbitrary.

5.4.1.2 Joint Test for the Presence of Random Elements

We can also go ahead and jointly test for the presence of two random elements. This means that we test, let's say, whether $\Delta_\alpha = \Delta_\gamma = 0$, but clearly the nullity of any combination of the random elements can be tested. Once more, we set it as an indirect test of equality of coefficients in the corresponding dimensions, say

$$H_0: \quad \beta_{11} = \beta_{12} = \cdots = \beta_{1N_2} = \cdots = \beta_{21} = \cdots = \beta_{2N_2} = \cdots = \beta_{N_1N_2}$$

on model

$$y_{ijt} = x'_{ijt}(\bar{\beta} + \alpha_i + \gamma_j) + x'_{ijt}\lambda_t + \varepsilon_{ijt} = x'_{ijt}\beta_{ij} + u_{ijt} \, . \tag{5.47}$$

Following the logic of the previous tests, we end up with the statistic

$$\hat{\beta}'_{(ij)} R_3' \left(R_3 \Sigma_3 R_3'\right)^{-1} R_3 \hat{\beta}_{(ij)} \sim \chi^2_{(N_1N_2-1)K} \, ,$$

where $\hat{\beta}_{(ij)}$ is the column of the N_1N_2 individual estimators, estimated by OLS using T observations for each ij pair, Σ_3 its variance, and

$$R_3 = \begin{pmatrix} I_K & -I_K & 0 & 0 & 0 & 0 \\ 0 & I_K & -I_K & 0 & 0 & 0 \\ & & & \ddots & & \\ 0 & 0 & 0 & 0 & I_K & -I_K \end{pmatrix} \quad \text{of size} \quad (K(N_1N_2-1)\times KN_1N_2)\,.$$

If the null hypothesis is not rejected, we can test the variability of the effects over the remaining dimension.

5.4.2 Testing in the Case of MLE

In order to test the hypothesis of the presence of random components in the coefficients for the maximum likelihood estimators we use the asymptotic distribution of the estimators. Indeed, if we stack all the estimators in a column of size $3K(K+1)/2+1$,

$$\hat{\theta} = \left(\text{vech}\,\hat{\Delta}_\alpha,\ \text{vech}\,\hat{\Delta}_\gamma,\ \text{vech}\,\hat{\Delta}_\lambda\,,\ \hat{\sigma}^2_\varepsilon\right)\,.$$

Denoting the Fisher Information matrix as

$$\mathbb{I}_\theta = \underset{N_1N_2T\to\infty}{\text{plim}} \frac{1}{N_1N_2T}\kappa_\theta$$

and typical elements of κ_θ as

$$\kappa_{kk'gg'} = \begin{cases} \frac{1}{2}\text{tr}\,\Sigma^{-1}H_{kk}\Sigma^{-1}H_{gg} & \text{if} \quad k=k';g=g' \\ \frac{1}{2}\text{tr}\,\Sigma^{-1}(H_{k'k}+H_{kk'})\Sigma^{-1}(H_{g'g}+H_{gg'}) & \text{otherwise,} \end{cases}$$

with $k,g \in 1,2,\ldots,3K(K+1)/2+1$. The first case is for the diagonal elements of the variance-covariance matrix, while the second is for the off-diagonal elements. The limiting distributions of the elements of the vector θ are given by:

$$\sqrt{N_1}\left(\hat{\sigma}^2_{\alpha,kk}-\sigma^2_{\alpha,kk}\right) \sim N\left(0,\mathbb{I}^{-1}_{\theta,kkgg}\right) \tag{5.48}$$

$$\sqrt{N_2}\left(\hat{\sigma}^2_{\gamma,kk}-\sigma^2_{\gamma,kk}\right) \sim N\left(0,\mathbb{I}^{-1}_{\theta,kkgg}\right) \tag{5.49}$$

$$\sqrt{T}\left(\hat{\sigma}^2_{\lambda,kk}-\sigma^2_{\lambda,kk}\right) \sim N\left(0,\mathbb{I}^{-1}_{\theta,kkgg}\right) \tag{5.50}$$

$$\sqrt{N_1N_2T}\left(\hat{\sigma}^2_{\varepsilon}-\sigma^2_{\varepsilon}\right) \sim N\left(0,\mathbb{I}^{-1}_{\theta,kkgg}\right)\,. \tag{5.51}$$

Knowing these asymptotic distributions, we can test the variability of coefficients along one or more of the three dimensions. A generic null would be

$$H_0: \quad G_l\theta = 0\,,$$

where l can be 1, 2 or 3 defining the three cases below. The corresponding Wald statistic is given by

$$\theta' G_l' \left(G_l \mathbb{I}_\theta^{-1} G_l'\right)^{-1} G_l\theta \sim \chi^2_{df_l}\,,$$

where df_l is the number of degrees of freedom corresponding to the number of restrictions for the cases $l = 1,2,3$. The three possible nulls are as follows.

1. **Simultaneous test for the equality of coefficients across all dimensions**
   In this case the null is:

$$\begin{aligned} H_0: \quad \sigma^2_{\alpha,11} = \sigma_{\alpha,12} = \cdots = \sigma^2_{\alpha,KK} = \sigma^2_{\gamma,11} = \sigma_{\gamma,12} = \cdots = \sigma^2_{\gamma,KK} \\ = \sigma^2_{\lambda,11} = \sigma_{\lambda,12} = \cdots = \sigma^2_{\lambda,KK} = 0 \end{aligned}$$

   with $G_1 = \left(I_{3K(K+1)/2},\, 0_{(3K(K+1)/2\times 1)}\right)$ and $df_1 = 3K(K+1)/2$.

2. **Joint test for equality over two dimensions**
   Let's say we want to test if the coefficients are constant at the group and individual levels, at the same time. This is equivalent to the following null:

$$H_0: \quad \sigma^2_{\alpha,11} = \sigma_{\alpha,12} = \cdots = \sigma^2_{\alpha,KK} = \sigma^2_{\gamma,11} = \sigma_{\gamma,12} = \cdots = \sigma^2_{\gamma,KK} = 0$$

   with $G_2 = \left(I_{2K(K+1)/2},\, 0_{(2K(K+1)/2\times K(K+1)/2+1)}\right)$ and $df_2 = 2K(K+1)/2$.

3. **Test for constancy along one dimension**
   In the case we want to test the equality of the coefficients along a single dimension, like i, a null is formulated as

$$H_0: \quad \sigma^2_{\alpha,11} = \sigma_{\alpha,12} = \cdots = \sigma^2_{\alpha,KK} = 0$$

   with $G_3 = \left(I_{K(K+1)/2},\, 0_{(K(K+1)/2\times 2K(K+1)/2+1)}\right)$ and $df_3 = K(K+1)/2$.

If the above tests present problems due to the null being at the boundary of the parameter space, we could go for a likelihood ratio-based approach

$$l = 2(\ln L_u - \ln L_r) \sim \chi^2_p\,,$$

where p is the number of restrictions and L_u and L_r are the unrestricted and restricted MLE, respectively. For example, to test the first hypothesis that there is no variability in the coefficients, $p = 3K^2$.

5.5 Prediction of the Coefficients

In many cases, we may also be interested in predicting the random coefficients themselves. This may be the case when a particular individual carries some importance or we wish to predict future values of the dependent variable for a given individual. Here we extend an approach developed by Lee and Griffiths (1979) to our 3D setting. We write the model as

$$y = \tilde{X}\tilde{\beta} + \varepsilon$$

$$\tilde{\beta} = \iota_{N_1N_2T} \otimes \bar{\beta} + (I_{N_1} \otimes \iota_{N_2T} \otimes I_K)\alpha + (\iota_{N_1} \otimes I_{N_2} \otimes \iota_T \otimes I_K)\gamma + (\iota_{N_1N_2} \otimes I_T \otimes I_K)\lambda$$

with the variance-covariance matrices of $\tilde{\beta}$ and ε given by

$$\Omega = \sigma_\varepsilon^2 I_{N_1N_2T}$$

and

$$\begin{aligned}\Sigma = {} & (I_{N_1} \otimes \iota_{N_2T} \otimes I_K)(I_{N_1} \otimes \Delta_\alpha)(I_{N_1} \otimes \iota'_{N_2T} \otimes I_K) \\ & + (\iota_{N_1} \otimes I_{N_2} \otimes \iota_T \otimes I_K)(I_{N_2} \otimes \Delta_\gamma)(\iota'_{N_1} \otimes I_{N_2} \otimes \iota'_T \otimes I_K) \\ & + (\iota_{N_1N_2} \otimes I_T \otimes I_K) \otimes (I_T \otimes \Delta_\lambda) \otimes (\iota'_{N_1N_2} \otimes I_T \otimes I_K)\,,\end{aligned}$$

respectively.

The coefficient vector is estimated by minimizing

$$Q(\tilde{\beta}, \bar{\beta}) = \frac{1}{\sigma_\varepsilon^2}(y - \tilde{X}\tilde{\beta})'(y - \tilde{X}\tilde{\beta}) + (\tilde{\beta} - \iota_{N_1N_2T} \otimes \bar{\beta})'\Sigma^{-1}(\tilde{\beta} - \iota_{N_1N_2T} \otimes \bar{\beta})$$

with respect to $\tilde{\beta}$. The solution is simply

$$\tilde{\beta} = (\frac{1}{\sigma_\varepsilon^2}\tilde{X}'\tilde{X} + \Sigma^{-1})^{-1}(\frac{1}{\sigma_\varepsilon^2}\tilde{X}'y + \Sigma^{-1}\iota_{N_1N_2T} \otimes \hat{\bar{\beta}}_{GLS})\,. \tag{5.52}$$

Although we have a solution which varies in all dimensions, the problem with this method is that we cannot identify the three components inside the coefficient vector, but only the sum total of all the components. Additionally, we need to replace the variance-covariance components by the corresponding estimates. The following procedure gives a way out for predicting individual components. Let us rewrite model (5.1) by merging α and $\bar{\beta}$ to some δ,

$$y = X_1\delta + X_2\gamma + X_3\lambda + \varepsilon = X_1\delta + u\,, \qquad u \sim (0, C_1)$$

where δ is a stochastic vector of length N_1K with

$$\delta = (\iota_{N_1} \otimes I_K)\bar{\beta} + \alpha \qquad \alpha \sim (0, C_2)$$

with

$$C_1 = X_2(I_{N_2} \otimes \Delta_\gamma)X_2' + X_3(I_T \otimes \Delta_\lambda)X_3' + I_{N_1N_2T}\sigma_\varepsilon^2 \quad \text{and} \quad C_2 = (I_{N_1} \otimes \Delta_\alpha)$$

being the covariance matrices at the two levels. Lee and Griffiths (1979) prove that the optimization problem

$$\min_{\delta,\bar{\beta}} (y - X\delta)' C_1^{-1} (y - X\delta) + (\delta - (\iota_{N_1} \otimes I_K)\bar{\beta})' C_2^{-1} \left(\delta - (\iota_{N_1} \otimes I_K)\bar{\beta}\right) , \quad (5.53)$$

which combines the sample and the prior information on the coefficient vector δ gives BLUE estimators for $\bar{\beta}$ (which is actually numerically the same as the $\hat{\bar{\beta}}_{GLS}$) and δ, given C_1 and C_2. This yields

$$\hat{\delta} = \left(X_1' C_1^{-1} X_1 + C_2^{-1}\right)^{-1} \left(X_1' C_1^{-1} y + C_2^{-1} (\iota_{N_1} \otimes I_K)\hat{\bar{\beta}}_{GLS}\right) .$$

This can be proven using the Woodbury matrix identity that $\left(X_1' C_1^{-1} X_1 + C_2^{-1}\right) = \Omega$, corresponding to (5.4), whose inverse has already been obtained whilst computing the GLS estimator. Regarding the two remaining inverses, one is trivial ($C_2^{-1} = (I_{N_1} \otimes \Delta_\alpha^{-1})$), while C_1^{-1} can be reached in steps similarly to Ω^{-1} as outlined in Sect. 5.2.2. The last step to get the individual predictor $\hat{\alpha}_i$ for a given individual i is the removal of the mean from $\hat{\delta}_i$, that is

$$\hat{\alpha}_i = \hat{\delta}_i - \hat{\bar{\beta}}_{GLS} ,$$

where $\hat{\delta}_i$ denotes the $(K \times 1)$ long i-th subvector of $\hat{\delta}$. By repeating the calculations with $\delta = (\iota_{N_2} \otimes I_K)\bar{\beta} + \gamma$ and $\delta = (\iota_T \otimes I_K)\bar{\beta} + \lambda$, we obtain the predictors for γ_j and for λ_t.

5.6 Bayesian Approach

Hsiao and Pesaran (2008) show that the RCM can also be imagined from a Bayesian angle. This is due to the fact that in RCM the coefficients are considered as random variables. Indeed, we can view the parameter β_{ijt} as stochastic with mean $\bar{\beta}$. Under the assumption that

$$\beta_{ijt} = \bar{\beta} + \gamma_j + \alpha_i + \lambda_t ,$$

and knowing that the random coefficients satisfy the distributional assumptions 1 and 2, we can settle with the following.

1. The prior distribution of $\bar{\beta}$ is diffuse.
2. The prior distributions of the error components are

$$\begin{aligned} \alpha &\sim N(0, (I_{N_1} \otimes \Delta_\alpha)) \\ \gamma &\sim N\left(0, (I_{N_2} \otimes \Delta_\gamma)\right) \\ \lambda &\sim N(0, (I_T \otimes \Delta_\lambda)) . \end{aligned} \quad (5.54)$$

3. The joint distribution of the observations, given $\bar{\beta}$, α, γ and λ, is

$$f_{y|\bar{\beta},\alpha,\gamma,\lambda}\left(y|\bar{\beta},\alpha,\gamma,\lambda\right) \sim N\left(X\bar{\beta}+\tilde{X}\left(I_{N_1}\otimes\iota_{N_2T}\otimes I_K\right)\alpha \right. \\ \left. +\tilde{X}\left(I_{N_2}\otimes\iota_{N_2T}\otimes I_K\right)\gamma+\tilde{X}\left(\iota_{N_1N_2}\otimes I_T\otimes I_K\right)\lambda,\sigma_\varepsilon^2 I_{N_1N_2T}\right) . \tag{5.55}$$

Now,

a. The posterior distribution of $\bar{\beta}$ given y is

$$f_{\bar{\beta}|y}\left(\bar{\beta}|y\right) \sim N\left(\left(X'\Omega^{-1}X\right)^{-1}\left(X'\Omega^{-1}y\right),\left(X'\Omega^{-1}X\right)^{-1}\right) . \tag{5.56}$$

b. The posterior distribution of α given y is

$$f_{\alpha|y}\left(\alpha|y\right)=\int\int\int f_{\alpha,\gamma,\bar{\beta},\lambda|y}\left(\alpha,\gamma,\bar{\beta},\lambda|y\right)d\gamma d\bar{\beta}d\lambda . \tag{5.57}$$

This is proportional to

$$f_{\alpha|y}\left(\alpha|y\right)\propto f_{y|\bar{\beta},\gamma,\alpha,\lambda}\left(y|\bar{\beta},\gamma,\alpha,\lambda\right)f_{\bar{\beta},\gamma,\alpha,\lambda}\left(\bar{\beta},\gamma,\alpha,\lambda\right) , \tag{5.58}$$

finally,

$$f_{\alpha|y}\left(\alpha|y\right)\propto f_{y|\bar{\beta},\gamma,\alpha,\lambda}\left(y|\bar{\beta},\gamma,\alpha,\lambda\right)f_{\alpha}\left(\alpha\right) . \tag{5.59}$$

Then, the posterior is given by

$$f_{\alpha|y}\left(\alpha|y\right)\sim N\left(\bar{\alpha},\Sigma_4\right) , \tag{5.60}$$

with an appropriate $\bar{\alpha}$ mean and Σ_4 variance.

The expressions for λ and γ are obtained following the same process that we used for α. Then, adding them to $\hat{\bar{\beta}}_{GLS}$ yields the same estimator as the one obtained using the extension of Lee and Griffiths (1979) proposed before. When the variance-covariance matrices are unknown, we have to add the assumptions on their prior distributions. In this case, we assume that Δ_α,Δ_γ and Δ_λ follow a Wishart distribution and σ_ε^2 a χ^2 distribution.

5.7 Extensions within the Linear Model

5.7.1 Alternative Model Formulations

Model (5.1) with coefficients as in (5.2) is, of course, only one model formulation among many possibilities for 3D data. One of the key characteristics of multidimensional data, however, is that heterogeneity may take on more complicated forms, and so can be incorporated into a random coefficient model in several alternative ways. Here, with no claim to being exhaustive, we collect alternative model specifications (which we represent with their coefficient structures) along with the

matrices necessary for their GLS estimation, estimation of their covariance matrices, and for the prediction of their coefficients.

The model (5.1) could have bilateral coefficients (i.e., varying simultaneously in the first two dimensions i and j) as follows:

$$\beta_{ijt} = \bar{\beta} + \gamma_{ij} \tag{5.61}$$

or bilateral and time coefficients as:

$$\beta_{ijt} = \bar{\beta} + \gamma_{ij} + \lambda_t \,. \tag{5.62}$$

Similar assumptions to 1-3 also hold for these models. Notice that both models can be thought of as straight 2D models if we let the (ij) pairs represent the individuals.[2] As all properties of these models are known from the existing literature, they enjoy limited attention here.

Instead, let us turn to models of interaction coefficients, like

$$\beta_{ijt} = \bar{\beta} + \alpha_{it} + \alpha^*_{jt} \,, \tag{5.63}$$

or to the all-encompassing parameter structure

$$\beta_{ijt} = \bar{\beta} + \gamma_{ij} + \alpha_{it} + \alpha^*_{jt} \,, \tag{5.64}$$

or

$$\beta_{ijt} = \bar{\beta} + \alpha_i + \gamma_{ij} + \lambda_t \,. \tag{5.65}$$

Table 5.1 collects the variable matrices and the covariance matrices to represent each model, while Table 5.2 collects matrices needed for the prediction of the random coefficients.

Table 5.1 Matrices to support the GLS estimation of models (5.61)–(5.64)

Model	β_{ijt}	Z
(5.61)	$\bar{\beta} + \gamma_{ij}$	$(X, X_4) = \tilde{X}\,[\iota_{N_1N_2T}, (I_{N_1N_2} \otimes \iota_T)] \otimes I_K$
(5.62)	$\bar{\beta} + \gamma_{ij} + \lambda_t$	$(X, X_3, X_4) = \tilde{X}\,[\iota_{N_1N_2T}, (I_{N_1N_2} \otimes \iota_T), (\iota_{N_1N_2} \otimes I_T)] \otimes I_K$
(5.63)	$\bar{\beta} + \alpha_{it} + \alpha^*_{jt}$	$(X, X_5, X_6) = \tilde{X}\,[\iota_{N_1N_2T}, (I_{N_1} \otimes \iota_{N_2} \otimes I_T), (\iota_{N_1} \otimes I_{N_2T})] \otimes I_K$
(5.64)	$\bar{\beta} + \gamma_{ij} + \alpha_{it} + \alpha^*_{jt}$	$(X, X_4, X_5, X_6) = \tilde{X}\,[\iota_{N_1N_2T}, (I_{N_1N_2} \otimes \iota_T), (I_{N_1} \otimes \iota_{N_2} \otimes I_T), (\iota_{N_1} \otimes I_{N_2T})] \otimes I_K$
(5.65)	$\bar{\beta} + \alpha_i + \gamma_{ij} + \lambda_t$	$(X, X_1, X_4, X_3) = \tilde{X}\,[\iota_{N_1N_2T}, (I_{N_1} \otimes \iota_{N_2T}), (I_{N_1N_2} \otimes \iota_T), (\iota_{N_1N_2} \otimes I_T)] \otimes I_K$

As seen in Table 5.1, we have all the ingredients necessary for the GLS estimation of models (5.61)–(5.64). Specifically, for model (5.64) for example, the GLS estimator is done with

$$\Omega = X_4 \mathrm{E}(\gamma\gamma')X_4' + X_5 \mathrm{E}(\alpha\alpha')X_5' + X_6 \mathrm{E}(\alpha^*\alpha^{*'})X_6' + \sigma_\varepsilon^2 I \,,$$

[2] Model (5.61) corresponds to Swamy's (1970) model, while (5.62) to Hsiao (1974).

where X_4, X_5 and X_6 are defined in Table 5.1, and we keep assuming the pairwise uncorrelatedness of the random coefficients. As computational difficulties are likely to step in again, the step-wise technique to invert Ω outlined in Sect. 5.2.2 comes handy.

To make the GLS feasible, the estimation of the variance components is to be done, which turns out to be no more difficult than it was for model (5.1). For each model, we just need to set the appropriate design matrices and all the procedures for the estimation of the variance-components presented before can be easily adapted to the specific characteristics of each model.

Predicting the random coefficients, along with the estimation of $\bar{\beta}$ may also be equally important in models (5.61)–(5.64). Luckily, the prediction technique outlined in Sect. 5.5 can be directly applied to these models with δ, C_1 and C_2 collected in Table 5.2.

Table 5.2 Matrices to support the prediction of the random coefficients for models (5.61)–(5.64)

Model	Predict	δ	C_1	C_2
(5.61)	γ_{ij}	$(\iota_{N_1N_2} \otimes I_K)\bar{\beta} + \gamma$	$I\sigma_\varepsilon^2$	$(I_{N_1N_2} \otimes \Delta_\gamma)$
(5.62)	γ_{ij}	$(\iota_{N_1N_2} \otimes I_K)\bar{\beta} + \gamma$	$X_3(I_T \otimes \Delta_\lambda)X_3' + I\sigma_\varepsilon^2$	$(I_{N_1N_2} \otimes \Delta_\gamma)$
	λ_t	$(\iota_T \otimes I_K)\bar{\beta} + \lambda$	$X_4(I_{N_1N_2} \otimes \Delta_\gamma)X_4' + I\sigma_\varepsilon^2$	$(I_T \otimes \Delta_\lambda)$
(5.63)	α_{it}	$(\iota_{N_1T} \otimes I_K)\bar{\beta} + \alpha$	$X_6(I_{N_2T} \otimes \Delta_{\alpha^*})X_6' + I\sigma_\varepsilon^2$	$(I_{N_1T} \otimes \Delta_\alpha)$
	α_{jt}^*	$(\iota_{N_2T} \otimes I_K)\bar{\beta} + \alpha^*$	$X_5(I_{N_1T} \otimes \Delta_\alpha)X_5' + I\sigma_\varepsilon^2$	$(I_{N_2T} \otimes \Delta_{\alpha^*})$
(5.64)	γ_{ij}	$(\iota_{N_1N_2} \otimes I_K)\bar{\beta} + \gamma$	$X_5(I_{N_1T} \otimes \Delta_\alpha)X_5' + X_6(I_{N_2T} \otimes \Delta_{\alpha^*})X_6' + I\sigma_\varepsilon^2$	$(I_{N_1N_2} \otimes \Delta_\gamma)$
	α_{it}	$(\iota_{N_1T} \otimes I_K)\bar{\beta} + \alpha$	$X_4(I_{N_1N_2} \otimes \Delta_\gamma)X_4' + X_6(I_{N_2T} \otimes \Delta_{\alpha^*})X_6' + I\sigma_\varepsilon^2$	$(I_{N_1T} \otimes \Delta_\alpha)$
	α_{jt}^*	$(\iota_{N_2T} \otimes I_K)\bar{\beta} + \alpha^*$	$X_4(I_{N_1N_2} \otimes \Delta_\gamma)X_4' + X_5(I_{N_1T} \otimes \Delta_\alpha)X_5' + I\sigma_\varepsilon^2$	$(I_{N_2T} \otimes \Delta_{\alpha^*})$
(5.65)	α_i	$(\iota_{N_1} \otimes I_K)\bar{\beta} + \alpha$	$X_3(I_T \otimes \Delta_\lambda)X_3' + X_4(I_{N_1N_2} \otimes \Delta_\gamma)X_4' + I\sigma_\varepsilon^2$	$(I_{N_1} \otimes \Delta_\alpha)$
	γ_{ij}	$(\iota_{N_1N_2} \otimes I_K)\bar{\beta} + \gamma$	$X_1(I_{N_1} \otimes \Delta_\alpha)X_1' + X_3(I_T \otimes \Delta_\lambda)X_3' + I\sigma_\varepsilon^2$	$(I_{N_1N_2} \otimes \Delta_\gamma)$
	λ_t	$(\iota_T \otimes I_K)\bar{\beta} + \lambda$	$X_1(I_{N_1} \otimes \Delta_\alpha)X_1' + X_4(I_{N_1N_2} \otimes \Delta_\gamma)X_4' + I\sigma_\varepsilon^2$	$(I_T \otimes \Delta_\lambda)$

5.7.2 *Incomplete Panels*

So far we have assumed that the data is complete and all variables span the three-dimensional space. We know, however, that real life data, especially multi-dimensional ones, are almost always of an incomplete nature. In the presence of such incompleteness, usual estimators might fail, and the identification of some parameters requires stronger assumptions. To formulate unbalanced observations, consider the following. For each ij-pair of individuals, instead of having T data points, observations are made on a $T_{ij} \in \{1 \dots T\}$ set, with cardinality (i.e., number of elements) $|T_{ij}|$.

Incomplete data during the GLS estimation means that we can no longer represent the covariance matrix Ω with kronecker products. While this is certainly inconvenient from mathematical and modelling points of view, it takes nothing away from the practical feasibility and the simplicity of the estimator as long as there is

no selection process causing the missing observations. Regardless of the incomplete nature of the data, we can assume without loss of generality that the number of total individuals and total time periods is still N_1, N_2 and T. With this observation, the covariance formulation (5.4) is simply given by

$$\Omega = I\sigma_\varepsilon^2 + \bar{X}_1(I_{N_1} \otimes \Delta_\alpha)\bar{X}_1' + \bar{X}_2(I_{N_2} \otimes \Delta_\gamma)\bar{X}_2' + \bar{X}_3(I_T \otimes \Delta_\lambda)\bar{X}_3',$$

where the "$-$" matrices are obtained from their complete data counterparts X_1, X_2 and X_3 by leaving out the rows corresponding to missing observations. In this way, these data matrices will uniformly shrink vertically until they have precisely $\sum_{ij}|T_{ij}|$ rows. This vertical, but not horizontal shrinkage holds as long as no "full" individual or time period is removed from the data, in which case we simply redefine N_1, N_2 or T. Luckily, the inverse of the covariance matrix is obtained in the same way as Sect. 5.2.2 suggested, with the same computational costs.

The same also holds for the estimation of the covariance matrix and the prediction of the random coefficients, where, if we adjust the matrices to incomplete data, that is, we leave out the rows corresponding to missing observations, the estimations and the predictions are done analogously to the complete data case.

5.7.3 Cross-Sectional Dependence

As with two-dimensional panels, sometimes we cannot rule out the possible cross-sectional dependence between the error terms. In the case of three-dimensional panels this could mean $O(N_1^2)$ ($O(N_2^2)$) extra parameters if dependence is solely assumed between individuals i (j), but can even mean controlling for $O(N_1^2N_2^2)$ extra unknown parameters if we feel that dependence between cross-sectional pairs should be addressed. To reduce the number of parameters to be incorporated, a convenient parametrization of the dependence structure can be proposed taking the form

$$\varepsilon_{ijt} = \frac{\sigma_{ij}}{\sqrt{1+\delta_{ij}'\delta_{ij}}}(\delta_{ij}' f_t + \vartheta_{ijt}), \tag{5.66}$$

where f_t and δ_{ij} are some $(s \times 1)$ latent factors and individual-pair-specific factor loadings, and ϑ_{ijt} is the idiosyncratic error term. It is easy to show that the cross-correlation between the ij and kl pairs of individuals is given by

$$\rho_{ijkl} = \rho_{klij} = \frac{\delta_{ij}'\delta_{kl}}{\sqrt{1+\delta_{ij}'\delta_{ij}} \cdot \sqrt{1+\delta_{kl}'\delta_{kl}}}.$$

Naturally, we do not always need this fully specified correlation structure for *pairs of individuals*. In many examples it is enough to address correlation between individual i-s or j-s. Using the example of linked employer-employee data, unobserved characteristics of firms might as well be cross-sectionally correlated, at least such

an idea is more reasonable than assuming the same for individuals. A simplified structure of (5.66) reflecting this is

$$\varepsilon_{ijt} = \frac{\sigma_j}{\sqrt{1+\delta_j'\delta_j}}(\delta_j' f_t + \vartheta_{ijt}),$$

where now only a number of N_2 unknown parameters are incorporated into the model. Either way, Pesaran (2006) shows that the focus parameters can be consistently estimated in both scenarios ($N_1 \to \infty$ or $N_2 \to \infty$ is necessary in the case of δ_{ij}-type factor loadings, while $N_1 \to \infty$ is necessary for δ_i-type factor loadings) when the observed regressors are augmented with the cross-sectional averages of the dependent variables and the regressors

$$\bar{y}_t = \sum_{ij} w_{ij} y_{ijt}; \qquad \bar{x}_t = \sum_{ij} w_{ij} x_{ijt},$$

with any weights satisfying

$$w_{ij} = O(1/N_1N_2), \quad \sum_{ij} w_{ij} = 1, \quad \sum_{ij} |w_{ij}| < a \text{ for some finite } a.$$

The non-weighted average $w_{ij} = 1/N_1N_2$ is of course a natural candidate.

We can also estimate the unobserved factors (and the factor loadings in turn) by factor analysis, as shown in Bai and Ng (2002).[3] The estimation takes the objective

$$V = \min_{\delta, f}\{\frac{1}{N_1N_2T}\sum_{ijt}(\varepsilon_{ijt} - \delta_{ij}' f_t)^2\} \quad \text{s.t.} \quad \frac{F'F}{T} = I_s,$$

with being the $(s \times T)$ matrix stacked version of f_t. This minimization problem is in fact identical to

$$\max_F\{\text{tr}(F'(\varepsilon^*\varepsilon^{*'})F)\}, \quad \text{with} \quad \varepsilon^* = \begin{pmatrix} \varepsilon_{111} & \dots & \varepsilon_{N_1N_21} \\ \vdots & \ddots & \vdots \\ \varepsilon_{11T} & \dots & \varepsilon_{N_1N_2T} \end{pmatrix}$$

from which the estimator for F is actually $\sqrt{T}$ times the eigenvectors corresponding to the s highest eigenvalues of the $(T \times T)$ matrix $\varepsilon^*\varepsilon^{*'}$. Once $\hat{F}$ is computed, the estimator for the factor loadings is obtained as in

$$\hat{\delta}' = (\hat{F}'\hat{F})^{-1}\hat{F}'\varepsilon^* = \frac{\hat{F}'\varepsilon^*}{T}.$$

[3] The method had originally been proposed in Connor and Korajzcyk (1986, 1988) for short panels (small T), and was later extended to large panels by Forni et al. (2000) and Stock and Watson (1998).

The factors are estimated consistently if $T \to \infty$, and if either $N_1 \to \infty$ or $N_2 \to \infty$ holds.

The above minimization problem is suited to cases when T is much smaller than $N_1(N_2)$, while its alternative, when instead of F, δ is constrained, is more efficient for panels where T is larger than the individual dimensions. Here we prefer the first option, as multi-dimensional panels tend to comprise many more individual units, than time periods.

5.7.4 Random Coefficients Correlated with the Explanatory Variables

If we cannot maintain the assumption regarding the mean independence of the random coefficients of the explanatory variables, the problem of the correlated effects emerges. Indeed, one could postulate

$$\begin{aligned} \mathrm{E}(\alpha_i|X_{ijt}) &\neq 0 \\ \mathrm{E}(\gamma_j|X_{ijt}) &\neq 0 \\ \mathrm{E}(\lambda_t|X_{ijt}) &\neq 0 \\ \mathrm{E}(\varepsilon_{ijs}|X_{ijt}) &= 0\,. \end{aligned}$$

In this case, one can use Mundlak's approach and model the dependency as follows:

$$\begin{aligned} \alpha_i &= (\iota_K \otimes \bar{x}'_{i..})\zeta_1 + \upsilon_i \\ \gamma_j &= (\iota_K \otimes \bar{x}'_{.j.})\zeta_2 + \tau_j \\ \lambda_t &= (\iota_K \otimes \bar{x}'_{..t})\zeta_3 + \pi_t\,, \end{aligned}$$

where the new stochastic components are independent and identically distributed as well as orthogonal to each other. This allows us to decompose the original stochastic components into two parts: one that is dependent on the explanatory variables, and another that is orthogonal to them.

Now, plugging these three equations into the original model and regrouping all the coefficients in one vector called Φ we get an augmented model of the form

$$y = S\Phi + \tilde{\varepsilon}\,, \tag{5.67}$$

which can be estimated with (F)GLS with the techniques outlined in Sect. 5.2.2. Additionally, we can test for the absence of correlation by testing for the significance of the corresponding coefficient, e.g., $\zeta_1 = 0$ for the absence of correlation between x_{ijt} and α_i.

5.7.5 Some Random and Some "Fixed" Coefficients?

In this chapter, we have covered the cases when the varying coefficients are exclusively thought of as random. The case of fixed coefficient components is rather straightforward as far as the estimation procedure is concerned if we make the *i.i.d.* assumption for the idiosyncratic error term. Hence, we do not devote much space to it in our chapter. However, the major problem in this case is the number of parameters to be estimated, which is potentially large taking away plenty of degrees of freedom. In fact, we did assume non-randomness of *some* components in many first stage estimations proposed for FGLS estimation and testing.

Now, it is not at all necessary to assume that either all are random or all are fixed. As it may very well happen in practice, it makes sense to incorporate different types of effects at different levels. Such "mixed" models can be specified in an analogous manner to model (5.1)-(5.2), where, for example α_i and γ_j are considered as random, but λ_t's are "fixed" coefficients.

This interesting model formulation is only briefly mentioned in theoretical works, see Hsiao (2003), but is fully absent from empirical works. This is hardly surprising for two main reasons. First, there is an enormous number of possible model specifications. We have already shown five economically meaningful random coefficient models – now imagine how this number grows when each coefficient can be either fixed or random. There is no testing tool constructed at the moment which can help decide between the numerous possible specifications. Second, computational difficulties are already present for many pure random and fixed coefficient models, and also arise (as we will see shortly) for mixed models.

Due to the aforementioned number of specifications and the scope of this chapter, we only briefly discuss the essentials of the estimation of "mixed" models, rather than excessively (and possibly dauntingly) carrying out a lengthy analysis.

Let us rewrite model (5.1)-(5.2) as

$$y_{ijt} = x'_{ijt}(\bar{\beta} + \lambda_t) + u_{ijt}\,, \quad \text{with} \quad u_{ijt} = x'_{ijt}(\alpha_i + \gamma_j) + \varepsilon_{ijt}\,. \tag{5.68}$$

Note that now $\bar{\beta}$ and λ_t are estimable, fixed parameters, and α_i and γ_j are pairwise uncorrelated random coefficients satisfying Assumption 2 of Sect. 5.2. The first observation to be made is that $\bar{\beta}$ and λ cannot be separated, and so are not identified.[4] To be able to estimate out $\bar{\beta}$ and λ_t, K parameter restrictions have to be imposed. An "even" restriction is to normalize the mean of λ_t, as

$$\sum_{t=1}^{T} \lambda_t = 0 \quad (K \times 1)\,.$$

The restricted model from this point behaves exactly as a pure random coefficients model, where now both $\bar{\beta}$ and λ_t are estimated from an FGLS, performed taking the covariance structure $\Omega = \mathrm{E}(u_{ijt}u'_{ijt})$ into account. Once the variance-covariance

[4] This is so, as the data matrix matching with $(\bar{\beta}, \lambda)$ has a rank deficiency of K.

matrix is inverted (using the step-wise inverse outlined in Sect. 5.2.2), the FGLS should have no further computational burden, as long as T is moderately small, which is typically the case.

In theory, any "mixed" model can be estimated using the following recipe: (i) identify the number and the form of parameter restrictions necessary to identify the model, and incorporate them (ii) derive the variance-covariance matrix Ω (more precisely its inverse), (iii) perform the GLS, and (iv) estimate the covariance matrix to make the GLS feasible. Although this is easily described in theory, the size and the number of fixed–random coefficients can strongly discourage its practical application. For example, when both α_i and λ_t are considered fixed along with $\bar{\beta}$, the number of parameters to be estimated directly (in the restricted model) is $K+(N_1-1)K+(T-1)K=(N_1+T-1)K$, which is infeasible with any statistical package even for moderately large N_i.

A possible solution to this dimensionality issue is to estimate the incomplete rank model with FGLS, then line up the parameter restriction. This might be more convenient, as long as $\Omega^{-1/2}$ can be attained at reasonable costs, as then we first transform model (5.68) by pre-multiplying with $\Omega^{-1/2}$, then estimate the transformed model with least squares of incomplete rank. Intuitively, picking up the notation of Section 5.2,

$$\Omega^{-1/2}y=\Omega^{-1/2}X\bar{\beta}+\Omega^{-1/2}X_3\lambda+\Omega^{-1/2}u$$

is simply some "pure" fixed coefficient model

$$\tilde{y}=\tilde{X}\bar{\beta}+\tilde{X}_3\lambda+\tilde{u}\,.$$

5.7.6 Higher Dimensions

We have seen how to formulate and estimate random coefficient models on three-dimensional panels. It is more and more typical, however, that the data at hand is four or even higher-dimensional, like industry, firm, or product level bilateral trade panels. In principle, higher-dimensional random coefficient models are just as easily analysed as their 3D counterparts, however, their investigation is subject to two key characteristics. One, due to several possible (semi-)asymptotic cases, the conditions needed for consistent estimation, prediction, or testing are highly non-trivial and need constant attention. Conveniently, for some models, especially for those with random coefficients only depending on a few indices, consistency is actually guaranteed with only one or two indices tending towards infinity. Second, the proposed estimation techniques become computationally forbidding. As four or even more indices are present, the few million observations usually contained in three-dimensional panels can now reach tens or even hundreds of millions of data points.

As an example, a typical four-dimensional model can be specified as

$$y_{ijst}=x'_{ijst}(\bar{\beta}+\alpha_i+\gamma_j+\zeta_s+\lambda_t)+\varepsilon_{ijst}\,. \tag{5.69}$$

with the additional index going from 1 to N_3, and conditions analogous to Assumptions 1 to 3 of Sect. 5.2. In order to obtain the GLS estimator of $\bar{\beta}$ given in Sect. 5.2.2, a matrix as large as the second highest dimension needs to be directly inverted. While this is most probably doable if the dataset is not large in all directions, the tables are turned once interaction effects are controlled for in the models. The same reasoning also holds for prediction and testing for the existence of the random components.

5.8 Non-linear Extension: RC Probit Model

This section explores the extension of the RCM framework to dichotomous outcomes. Let us consider a probit specification. We set the latent variable model as follows:

$$y^*_{ijt} = x'_{ijt}\left(\bar{\beta} + \alpha_i + \gamma_j + \lambda_t\right) + \varepsilon_{ijt}\ .$$

The observed variable y_{ijt} takes the value 1, if $y^*_{ijt} > 0$ and 0, if $y^*_{ijt} \leq 0$. This is equivalent to saying:

$$y_{ijt} = \begin{cases} 1, & \text{if} \quad \varepsilon_{ijt} > -x'_{ijt}\left(\bar{\beta} + \alpha_i + \gamma_j + \lambda_t\right) \\ 0, & \text{otherwise.} \end{cases}$$

If we assume that ε_{ijt} follows a standard normal distribution we are in the Probit setting. Fixing the random coefficient components, the likelihood is the product of the probability function distribution of each observation:

$$P(y_{ijt} = 1|\alpha_i, \gamma_j, \lambda_t) = P\left(\varepsilon_{ijt} < x'_{ijt}(\bar{\beta} + \alpha_i + \gamma_j + \lambda_t)|\alpha_i, \gamma_j, \lambda_t\right)\ .$$

However, the distribution of y_{ijt} without fixing the random coefficient components is not identical nor independent across observations. Thus, the likelihood can no longer be the product of individual likelihoods.

As a result, the estimation of this model is computationally demanding because it requires the evaluation of N_1N_2T integrals. Following Chib and Greenberg (1998), we can use a Markov Chain Monte Carlo EM algorithm in order to estimate the parameter vector β and the variance-covariance matrix Ω. Denoting the conditional distribution of y^* given y as $g(y^*|y,\theta)$ where θ is the vector containing β and 3K(K+1)/2 +1 parameters in the covariance matrix Ω, the algorithm runs as follows.

1. *E-step*: Estimation of the conditional expectation in iteration r.

$$Q(\theta, \hat{\theta}^r) = \int \ln \mathrm{f}(y, y^*|\hat{\theta}^r) g(y^*|y, \hat{\theta}^r) dy^* = \int \ln \mathrm{f}(y^*|\hat{\theta}^r) g(y^*|y, \hat{\theta}^r) dy^*\,.$$

The calculation of $Q(\theta, \hat{\theta}^r)$ is not possible, hence we estimate it with

$$Q(\theta, \hat{\theta}^r) = \frac{1}{M} \sum_{h=1}^{M} \ln \mathrm{f}(Y_h^* | \theta) \ ,$$

where the M values of y_h^* are drawn from the distribution using a Gibbs sampling technique which has the truncated multivariate normal distribution as the equilibrium distribution

$$\phi_{N_1 N_2 T}(y^* | X\hat{\bar{\beta}}^r, \hat{\Sigma}^r) I(y^* \in B) \tag{5.70}$$

$I(y^* \in B)$ indicates the vector of events that $y_{ijt}^* > 0$ if $y_{ijt} = 1$ and $y_{ijt}^* \leq 0$ if $y_{ijt} = 0$. To sample from this distribution, we follow Chib and Greenberg (1998) which suggests to sample each observation from the untruncated normal distribution and then use inverse sampling to get data within the event $y^* \in B$. Finally, $f(y^* | \theta)$ is the $N_1 N_2 T$ multivariate normal with mean $X\bar{\beta}$ and covariance matrix Ω.

Furthermore, in order to reduce the computational cost of doing a Gibbs Sampling in each *E-step*, Levine and Casella (2001) suggest replacing it by importance sampling. Thus, in the first step $Q(\theta, \hat{\theta}^0)$ is obtained by sampling from the truncated multivariate normal distribution with a given starting value for the parameters. Then, at each iteration instead of sampling again, an update is done on the conditional mean in the *E-step* given by

$$Q(\theta, \hat{\theta}^r) = \frac{\sum_{h=1}^{M} w_h \ln \mathrm{f}(Y_h^* | \theta)}{\sum_{h=1}^{M} w_h} \ ,$$

where the weights in each iteration are:

$$w_h = \frac{g(Y^* | Y, \hat{\theta}^r)}{g(Y^* | Y, \hat{\theta}^0)} \ .$$

2. *M-Step*: Maximize the function $Q(\theta, \hat{\theta}^r)$ for θ.
   The maximization has to be done in two stages. In the first, we obtain the estimator of $\bar{\beta}$, then we plug this value back into $Q(\theta, \hat{\theta}^r)$. Then we maximize this function for the elements in the covariance matrix Ω using a Newthon-Raphson algorithm.

5.9 A Simulation Experiment

In order to assess the performance of the proposed estimation methods, we conducted a Monte Carlo experiment generating 100 samples in 10 different scenarios (See Table 5.3).

In 8 of the 10 cases, the data were generated from a normally distributed disturbance term, whereas in the last 2 cases the stochastic elements followed a t-location

Table 5.3 Various scenarios used for the simulation exercise

Scenario	Data Generating Process			Value		Size		
	RCC	Disturbance Term	Indep. Var.	Interc.	Slope	N_1	N_2	T
1	a	$N(0,0.25)$	$N(0,25)$	2	3	20	20	5
2	a	$N(0,0.25)$	$N(0,25)$	2	3	10	10	5
3	a	$N(0,0.25)$	$N(0,25)$	2	0.7	20	20	5
4	a	$N(0,0.25)$	$N(0,25)$	2	0.7	10	10	5
5	a	$N(0,0.25)$	$U(0,25)$	2	3	20	20	5
6	a	$N(0,0.25)$	$U(0,25)$	2	3	10	10	5
7	a	$N(0,0.25)$	$U(0,25)$	2	0.7	20	20	5
8	a	$N(0,0.25)$	$U(0,25)$	2	0.7	10	10	5
9	b	T-scale location: $\mu = 0$, $\sigma = 0.5$, $df = 3$	$U(0,25)$	2	3	20	20	5
10	b	T-scale location: $\mu = 0$, $\sigma = 0.5$, $df = 3$	$U(0,25)$	2	3	10	10	5

Notes: RCC=Random Coefficient Components. Interc.: Intercept. a: $N\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & 0.5\\0.5 & 1\end{pmatrix}\right)$. b: t-location scale $\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & 0.5\\0.5 & 1\end{pmatrix}\right)$, $df = 3$.

scale distribution. We implement the following three estimation methods using Matlab:[5]

- FGLS using an estimated variance-covariance with Method 1.
- FGLS using an estimated variance-covariance with Method 2.
- MLE: Anderson algorithm.

The performance of the different coefficient estimators (including OLS) is presented in Table 5.4 for three small sample scenarios.[6]

Table 5.4 Small sample results for the slope parameters

	Scenario	OLS	Method 1	Method 2	MLE
Relative Bias (%)	4	9.9143	11.8143	11.8000	9.4714
	8	-6.9857	-7.4429	-7.6571	-7.5000
	10	1.1033	-0.0400	0.1133	-0.0100
RMSE	4	0.5226	0.5128	0.5169	0.4995
	8	0.6244	0.6477	0.6593	0.6462
	10	0.1742	0.0353	0.0271	0.0052

Considering the Root Mean Square Error (RMSE), we conclude that the Feasible Generalised Least Squares (FGLS) and the Maximum Likelihood (ML) estimates

[5] Program codes are available with the authors on request.

[6] We do not present all the tables and graphs obtained for RMSE and Relative Bias for length reasons but all of them are available with the authors on request.

are better than Ordinary Least Squares (OLS). MLE performs better than FGLS in all scenarios. Additionally, using within dimensions variability (Method 1) is better than using overall variation (Method 2) for the estimation of the variance-covariance components. This is always the case, except when the regressors are random draws from a uniform distribution. In this case, it is better to estimate them using the overall variation. MLE performs better than FGLS, which is expected since the true data generating process is a Normal distribution and this is the likelihood that is maximized. However, MLE beats FGLS even when the data generating process is not normal.

Concerning the estimated parameters of the variance-covariance matrix, MLE outperforms Methods 1 and 2 in all scenarios, although Methods 1 and 2 retrieve good estimators of the variance-covariance elements in all scenarios. However, we can see that the RMSE increases considerably when the estimation is done with a small sample.

All in all, the results for the estimation methods proposed suggest that the best one is MLE, followed by Method 1 (using within variation) and finally Method 2 (with overall variability).

Looking at the computational time, it is interesting to note the low computing time of Methods 1 and 2 compared to MLE. Both these methods are easy to implement. Method 1 using within variability has the lowest computing time: 0.5 seconds for the small sample and 7.5 seconds for the larger one. Method 2 requires 0.7 seconds for a total sample size of 500 observations. The computing time rises exponentially with an increase of the sample size. In our case, an increase to 2,000 units leads to a time of 20 seconds, which is still fast. The implementation of the MLE is more cumbersome and requires more computing power. The time required for the small sample is approximately 0.8 seconds and for the bigger one 40. The convergence with Anderson algorithm is fast reaching a maximum with three or four iterations. However, the linearization used in the algorithm can retrieve less accurate estimators.

Methods 1 and 2 suffer from the usual problems:

1. The estimated variance-covariance matrix can be negative definite as anticipated in Sect. 5.2.4. A possible solution is the positivity restricted non-linear least squares outlined in Sect. 5.2.5.
2. There is a 5% probability of having a non-invertible estimated variance-covariance matrix. A possible solution is to use the Moore-Penrose Pseudo inverse. This non-invertibility issue is also present in the case of MLE.

5.10 Conclusions

Higher dimensional data are becoming more and more available and used in empirical studies. Such data allow us to explicitly take into account possible heterogeneity in response coefficients across different dimensions. This chapter proposes a random coefficient framework as a way of modelling response heterogeneity, and examines

the specification, estimation and inference in random coefficient models for three dimensional panel data, typically made up of individuals within groups followed over time. This model is particularly useful for policy evaluation when treated individuals have heterogeneous responses to the treatment. Adding another dimension can also allow us to have a randomized setting *within* groups. The chapter ends with extensions to non-linear and higher dimensional settings.

We derive feasible generalized least squares and maximum likelihood estimators of the coefficients involving a prior or simultaneous estimation of the variance components. For the first method, we not only generalise existing two-dimensional approaches but also develop two new estimation methods for the variance-covariance components, while for the latter we adapt an algorithm based on Anderson (1971). We also develop tests for the presence/absence of heterogeneity in response coefficients. We show the infeasibility of a direct test for the coefficient variance-covariance matrix equal to zero, and propose an alternative procedure testing equality of coefficients across various dimensions. We go on to examine the prediction of the random components, incomplete panels, cross-sectional dependence as well as possible correlation of the random coefficients with the explanatory variables.

We present two major extensions of the three-dimensional linear model: (i) a probit setting with random coefficients, for which we suggest a Monte Carlo Expectation Maximisation algorithm for solving the maximum likelihood problem to estimate the coefficients as well as the variance components; and (ii) possible specifications for higher dimensional data.

Finally, running a small simulation experiment for the three-dimensional linear model, we show that FGLS has a lower RMSE than OLS and MLE beats FGLS. We also observe that FGLS can often yield non-positive definite matrices, while MLE is computationally costlier.

References

Anderson, T. (1971). Estimation of covariance matrices with linear structure and moving average processes of finite order. Technical report, DTIC Document.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70:191–221.

Balazsi, L., Matyas, L., and Wansbeek, T. (2015). The estimation of multidimensional fixed effects panel data models. *Econometric Reviews*, forthcoming, published online: April 2015.

Benson, H. Y. and Vanderbei, R. J. (2003). Solving problems with semidefinite and related constraints using interior-point methods for nonlinear programming. *Mathematical Programming*, 95(2):279–302.

Callanan, T. (1985). Restricted maximum likelihood estimation of variance components: Computational aspects. *Retrospective Theses and Dissertations. Paper 8682*.

Chib, S. and Greenberg, E. (1998). Analysis of multivariate probit models. *Biometrika*, 85(2):347–361.

Connor, G. and Korajzcyk, R. (1986). Performance measurement with the arbitrage pricing theory: A new framework for analysis. *Journal of Financial Economics*, 15:373–394.

Connor, G. and Korajzcyk, R. (1988). Risk and return in an equilibrium APT: Application to a new test methodology. *Journal of Financial Economics*, 21:255–289.

De Leeuw, J., Meijer, E., and Goldstein, H. (2008). *Handbook of Multilevel Analysis*. New York, Springer.

Egger, P. and Pfaffermayr, M. (2003). The proper panel econometric specification of the gravity equation: A three-way model with bilateral interaction effects. *Empirical Economics*, 28(3):571–580.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2000). The generalized dynamic factor model: Identification and estimation. *Review of Economics and Statistics*, 82:540–554.

Gibbons, R. D. and Hedeker, D. (1997). Random effects probit and logistic regression models for three-level data. *Biometrics*, pages 1527–1537.

Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, 72(358):320–338.

Heckman, J. and Vytlacil, E. (1998). Instrumental variables methods for the correlated random coefficient model: Estimating the average rate of return to schooling when the return is correlated with schooling. *Journal of Human Resources*, pages 974–987.

Hildreth, C. and Houck, J. P. (1968). Some estimators for a linear model with random coefficients. *Journal of the American Statistical Association*, 63(322):584–595.

Hsiao, C. (1974). Statistical inference for a model with both random cross-sectional and time effects. *International Economic Review*, 15(1):12–30.

Hsiao, C. (1975). Some estimation methods for a random coefficient model. *Econometrica*, pages 305–325.

Hsiao, C. (2003). *Analysis of Panel Data*. Cambridge University Press, 2 edition.

Hsiao, C. and Pesaran, M. (2008). Random coefficient models. In Matyas, L. and Sevestre, P., editors, *The Econometrics of Panel Data*, pages 185–213. Springer.

Jiang, J. (1996). REML estimation: Asymptotic behavior and related topics. *The Annals of Statistics*, 24(1):255–286.

Kwan, S. H. (1991). Re-examination of interest rate sensitivity of commercial bank stock returns using a random coefficient model. *Journal of Financial Services Research*, 5(1):61–76.

Lee, L. F. and Griffiths, W. E. (1979). The prior likelihood and best linear unbiased prediction in stochastic coefficient linear models. *Working Papers in Econometrics and Applied Statistics, No. 1. University of New England*.

Levine, R. A. and Casella, G. (2001). Implementations of the Monte Carlo EM algorithm. *Journal of Computational and Graphical Statistics*, 10(3):422–439.

Patterson, H. D. and Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika*, 58(3).

Pesaran, M. H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica*, 74:967–1012.

Rao, C. R. (1970). Estimation of heteroscedastic variances in linear models. *Journal of the American Statistical Association*, 65(329):161–172.

Skrondal, A. and Rabe-Hesketh, S. (2010). *Multilevel Modelling*. London: Sage.

Stock, J. H. and Watson, M. (1998). Diffusion indexes. *NBER Working Paper 6702*.

Swamy, P. (1970). Efficient inference in a random coefficient regression model. *Econometrica*, 33:311–323.

Chapter 6
Discrete Response Models

Balazs Kertesz

Abstract In this chapter, we consider the estimation of binary and multiple choice fixed effects models in a multi-dimensional setting. Sufficient statistics are put into action to get rid of the fixed effects, considered as nuisance or incidental parameters. We show how multiple choice models can, in fact, be viewed and treated as binary choice ones with an added dimension. We also deal with the issue of selectivity in a "Heckit" approach.

6.1 Introduction

There are many applications where one would like to infer effects based on qualitative data. In such situations, the researcher is often faced with a discrete rather than a continuous outcome variable discussed in the previous chapters. In general, we focus on models where the outcome variable takes the values of $0, 1, 2, \ldots$. However, the order of values representing an economic outcome might not necessarily be meaningful, for example, in consumer choice settings. A special case of this type is when the outcomes are binary, and the dependent variable of the model is in fact an indicator variable describing an event happening or not, labelled by 1 and 0, respectively. As in all previous chapters, in most of our discussion we use the notation i representing an individual-i (for instance a consumer), j also denoting an individual-j (for example a brand) and t for time, representing a three-dimensional panel model. As seen in the previous chapters, heterogeneity can be viewed as random or fixed.

The standard approach for estimating models with a discrete dependent variable is the Maximum Likelihood Estimation (or some variant of it), although there is considerable development in the two-dimensional panel model literature regarding semi-parametric approaches. The theoretical foundation of discrete choice models

Balazs Kertesz
Central European University, Budapest, Hungary

L. Matyas (ed.), *The Econometrics of Multi-dimensional Panels*, Advanced Studies in Theoretical and Applied Econometrics 50, DOI 10.1007/978-3-319-60783-2_6

is connected to the axiomatic approach discussed by McFadden (1973). Chamberlain (1980) introduced the fixed effect approach in non-linear panel models and one of the possible estimation procedures, the conditional likelihood estimation. Manski (1975) considered a binary choice model with individual fixed effects estimated by a semi-parametric approach, the Maximum Score Estimator. Among other unfavorable properties, this procedure is of limited use in multi-dimensional panels, because it is applicable for only a narrow set of fixed effects, as discussed by Charbonneau (2012). Mundlak (1978) introduced the correlated random effects framework, further developed by Chamberlain (1984), which is a kind of competing approach to the fixed effect models. Correlated random effect models relax the independence assumption between the unobserved heterogeneity and the covariates, which is a key assumption in random effects models. Pakes et al. (2015) capture the behavioral choice model with moment inequalities without specifying the distribution of disturbance terms parametrically. Bonhomme (2012) extends the approach with moment conditions to continuous dependent variables. The other strand of the literature tries to reduce the bias caused by the omitted variables, initially discussed by Hahn and Newey (2004) and further developed by Arellano and Bonhomme (2009) and Fernandez-Val and Weidner (2016).

In this chapter, we discuss the fixed effects approach for static non-linear binary and multiple choice multi-dimensional panel models, using conditional Maximum Likelihood techniques. The general outline of the estimation procedure is as follows. As in Chap. 1, we would like to get rid of the fixed effects, as their number is large, and estimating them in most cases is neither necessary nor feasible. After specifying the model including the fixed effects, we derive a sufficient statistic for the fixed effect parameters using the Neyman-Fisher factorization theorem, which enables us to eliminate the heterogeneity factors (this in fact corresponds to the orthogonal projections used for linear models). By conditioning on the sufficient statistic, the likelihood can be maximized by standard procedures to get consistent estimates for the parameter of interest. In the case of the presence of more than one fixed effect, we rely on the solution that Charbonneau (2012, 2014) proposes, she repeatedly applies the sufficient statistic approach to eliminate the fixed effects one by one. In constructing the sufficient statistic, the free dimension through which the fixed effects are held constant plays a crucial role, as one has to consider the sum of the outcome variable in this dimension.

6.2 Fixed Effects Binary Choice Models

We begin our discussion with the binary choice models. Throughout this chapter, we still think about the data generating process as linear, though one is restricted to observing only binary outcomes, usually denoted by a sequence of zeroes and ones. We are looking for the estimators of the same models as covered in Chap. 1, and here we give the modeling equations in a compressed form.

In a three-dimensional setting, these models look like

$$y_{ijt} = \mathbf{1}\{x'_{ijt}\beta + \mathrm{FE}_{ijt} + \varepsilon_{ijt} \geq 0\},$$

where $\mathbf{1}\{Cond\}$ is the indicator function taking value 1 whenever the condition *Cond* is satisfied and 0 otherwise. x_{ijt} is a $(p \times 1)$ vector of explanatory variables, and β is its $(p \times 1)$ coefficient vector. The FE_{ijt} are the fixed effect parameters and the ε_{ijt} are the disturbance terms. Here we are covering the same fixed effects specifications as in Chap. 1

$$\mathrm{FE}_{ijt} = \gamma_{ij} \tag{6.1}$$
$$\mathrm{FE}_{ijt} = \alpha_{it} \tag{6.2}$$
$$\mathrm{FE}_{ijt} = \alpha_{it} + \alpha^*_{jt} \tag{6.3}$$
$$\mathrm{FE}_{ijt} = \gamma_{ij} + \lambda_t \tag{6.4}$$
$$\mathrm{FE}_{ijt} = \alpha_i + \gamma_j + \lambda_t \tag{6.5}$$
$$\mathrm{FE}_{ijt} = \alpha_{it} + \alpha^*_{jt} + \gamma_{ij}\,. \tag{6.6}$$

The interpretation of such various fixed effect parameters FE_{ijt}, like in the 2D case, is that we account for a deterministic shift (the constant term of the regression model) being constant along the dimensions not appearing in the index of the fixed effect parameters that otherwise vary through the dimensions as indicated. The role of including these fixed effects into a regression model is to account for the heterogeneity of the data, while omitting them would result in the well-known omitted variable bias.

6.2.1 Model Assumptions

The theoretical grounding of non-linear fixed effects models is discussed in a two-dimensional framework by Lechner et al. (2008). However, we view the fixed effect parameters, as discussed in Chap. 1, as parameters to be estimated. Following Chamberlain (1980), we assume in the case of the first model (6.1) that the binary outcomes of the variables y_{ijt} are independent conditional on the explanatory variables x, their coefficients β and the fixed effects γ both within the group formed by the pairs of individual-i and individual-j and across groups, which is the *conditional independence assumption*, with the specification

$$P(y_{ijt} = 1 \mid \{\gamma_{ij}\}_{ij}, \beta, \{x_{ijt}\}_{ijt}) = F(\gamma_{ij} + x'_{ijt}\beta).$$

The latter is referred to as the *strict exogeneity assumption*. We use the notation $\{a_{nm}\}_n$ to indicate that the object a_{nm}'s indices run through index n. In the case of fixed effects, for example, γ_{nm} represents a bilateral individual fixed effect (individual-i, individual-j fixed effect) for the nth and mth individuals, respectively.

It can be seen from this formulation that every model has its own set of assumptions, and when using them, one needs to make sure that they are met. The

assumptions of the second model (6.2) can be written symmetrically, but now, the binary outcome y_{ijt} needs to be independent within and between groups conditional on $(\{\alpha_{it}\}_{it}, \beta, \{x_{ijt}\}_{ijt})$, where groups are formed by the pairs of individual-i and time t, with the specification

$$P(y_{ijt} = 1 \mid \{\alpha_{it}\}_{it}, \beta, \{x_{ijt}\}_{ijt}) = F(\alpha_{it} + x'_{ijt}\beta).$$

To estimate model (6.3), we require the independence of the outcome variable y_{ijt} conditional on the fixed effects, slope parameters and the covariates $(\{\alpha_{it}\}_{it}, \{\alpha^*_{jt}\}_{jt}, \beta, \{x_{ijt}\}_{ijt})$ both between and within groups formed by individual-i and time t as well as individual-j and time t, with the specification

$$P(y_{ijt} = 1 \mid \{\alpha_{it}\}_{it}, \{\alpha^*_{jt}\}_{jt}, \beta, \{x_{ijt}\}_{ijt}) = F(\alpha_{it} + \alpha^*_{jt} + x'_{ijt}\beta).$$

To estimate model (6.4) we make very similar assumptions as in (6.1) but conditional on $(\{\gamma_{ij}\}_{ij}, \{\lambda_t\}_t, \beta, \{x_{ijt}\}_{itj})$, with the specification

$$P(y_{ijt} = 1 \mid \{\gamma_{ij}\}_{ij}, \{\lambda_t\}_t, \beta, \{x_{ijt}\}_{itj}) = F(\gamma_{ij} + \lambda_t + x'_{ijt}\beta).$$

Similarly, the last two models (6.5) and (6.6) require the same independence assumptions, namely that the binary outcome variable y_{ijt} is independent across all the dimensions, but with conditions $(\{\alpha_i\}_i, \{\gamma_j\}_j, \{\lambda_t\}_t, \beta, \{x_{ijt}\}_{ijt})$ in the case of the fifth model and $(\{\alpha_{it}\}_{it}, \{\alpha_{jt}\}^*_{jt}, \{\gamma\}_{ij}, \beta, \{x_{ijt}\}_{ijt})$ in the case of the last model with the specifications

$$P(y_{ijt} = 1 \mid \{\alpha_i\}_i, \{\gamma_j\}_j, \{\lambda_t\}_t, \beta, \{x_{ijt}\}_{ijt}) = F(\alpha_i + \gamma_j + \lambda_t + x'_{ijt}\beta)$$

and

$$P(y_{ijt} = 1 \mid \{\alpha_{it}\}_{it}, \{\alpha^*_{jt}\}_{jt}, \{\gamma\}_{ij}, \beta, \{x_{ijt}\}_{ijt}) = F(\alpha_{it} + \alpha^*_{jt} + \gamma_{ij} + x'_{ijt}\beta),$$

respectively.

In fact, we are going to treat the fixed effects as nuisance (or incidental) parameters. We start by assuming a balanced panel, that is we have a panel of individual-is running through $i = 1, \ldots, N_1$, within each individual-i, we observe individual-js running through $j = 1, \ldots, N_2$ tracked over time $t = 1, \ldots, T$. In practice, N_1 and N_2 are often large relative to the length of time T. At the end of this chapter, we relax this assumption and show that it has practically no effect on the estimation.

6.2.2 Problems with Non-linear Fixed Effects Models

A crucial difficulty arises with respect to the identification in these models. As usual in regression estimation, variability is required for parameter identification. Therefore, for example in model (6.1), it is impossible to identify the effect of time-invariant observables. In the case of model (6.2), explanatory variables being con-

stant over individual-j dimension invalidates the identifiability of their effects. In general, the variability of the explanatory variables along the free dimension of a fixed effect is a necessary condition for identifying the impact of a covariate (explanatory variable). Otherwise, we are unable to compare choice probabilities for various values of the covariates, which means that we cannot use these models to identify causal effects unless some distributional assumptions are made.

Therefore, as another issue caused by the discreteness of the dependent variable y, when individuals (individual-(ij)s in the case of model (6.1)) stay in the same state, they do not provide any information for the estimation of β. This is the so-called mover-stayer problem. Assuming that the fixed effects γ_{ij} in the complete period of time $t = 1,\dots,T$ were such that they produced "only" outcomes $y_{ijt} = 1$, they would also produce the same for any other fixed effect $\bar{\gamma}_{ij} > \gamma_{ij}$. Therefore, estimations must be carried out on individuals which change states over time (moving individuals).

Our main aim is to deliver an estimation procedure for the models discussed at the beginning of this chapter. However, we come across two difficulties when trying to do so. Since the fixed effects are parameters, we need to take care of them. Basically, there are two options at hand. One could either consistently estimate these parameters, or eliminate them assuming that they are nuisance. Consistently estimating these parameters is hardly achievable in general, because it requires asymptotics in the free dimension of the fixed effect, T along with both N_1 and N_2 for instance in model (6.1). Despite having $N_1 \to \infty$ and $N_2 \to \infty$, the finite number of observations in the time dimension invalidates the consistency of the nuisance parameter estimates and their inconsistency eventually also contaminates the estimates of the slope parameters β. Standard techniques of orthogonal projections covered in Chap. 1 cannot be applied directly to non-linear models, hence we cannot directly eliminate the fixed effect parameters. As a consequence, we turn to Maximum Likelihood techniques based on distributional assumptions on the disturbance terms ε_{ijt}.

For illustrative purposes, consider model (6.1), which can in fact be reduced to a two-dimensional model treating the individual index-pairs (ij) as one, k. The model then can be transformed into a familiar form of a two-dimensional panel model

$$\begin{aligned} y_{(ij)t} &= \mathbf{1}\{\gamma_{(ij)} + x'_{(ij)t}\beta + \varepsilon_{(ij)t} \geq 0\} \\ y_{kt} &= \mathbf{1}\{\gamma_k + x'_{kt}\beta + \varepsilon_{kt} \geq 0\}\,. \end{aligned}$$

Nonetheless, issues we come across in two-dimensional models are amplified in a general higher dimensional setting.

For the likelihood, first we derive the probabilities of the event $\{y_{kt} = 1\}$ occurring and its complement $\{y_{kt} = 0\}$ conditional on the data we observe and the model parameters

$$\begin{aligned} P(y_{kt}=1 \mid x_{kt},\gamma_k,\beta) &= P(\gamma_k + x'_{kt}\beta + \varepsilon_{kt} \geq 0 \mid x_{kt},\gamma_k,\beta) \\ &= P(-\varepsilon_{kt} \leq \gamma_k + x'_{kt}\beta \mid x_{kt},\gamma_k,\beta) \\ &= P(\varepsilon_{kt} \leq \gamma_k + x'_{kt}\beta \mid x_{kt},\gamma_k,\beta) \\ &= F(\gamma_k + x'_{kt}\beta), \end{aligned}$$

where we assumed the disturbance term to be symmetrically distributed with a cumulative distribution function F. Furthermore, throughout the discussion of the fixed effects estimation in this section, we assume that the cumulative distribution function of the disturbance terms can be given by the logistic distribution, so $F(u) = \Lambda(u) := \exp(u)/(1+\exp(u))$. We call the models logit when the disturbance term is distributed logistically $(F(u) = \Lambda(u))$, and probit when this distribution is the standard normal $(F(u) = \Phi(u))$. Therefore,

$$P(y_{kt}=1 \mid x_{kt},\gamma_k,\beta) = \frac{\exp(\gamma_k + x'_{kt}\beta)}{1+\exp(\gamma_k + x'_{kt}\beta)},$$

while

$$P(y_{kt}=0 \mid x_{kt},\gamma_k,\beta) = \frac{1}{1+\exp(\gamma_k + x'_{kt}\beta)}.$$

Since the binary choice model is in fact a Bernoulli scheme, by slightly abusing the notation, we can write the likelihood function as

$$\begin{aligned} L &= L(\gamma_k,\beta;y_{kt},x_{kt}) \\ &= \prod_k \prod_{t=1}^{T} P(y_{kt}=1 \mid x_{kt},\gamma_k,\beta)^{y_{kt}} P(y_{kt}=0 \mid x_{kt},\gamma_k,\beta)^{1-y_{kt}} \\ &= \prod_k \prod_{t=1}^{T} F(\gamma_k + x'_{kt}\beta)^{y_{kt}} \left[1 - F(\gamma_k + x'_{kt}\beta)\right]^{1-y_{kt}}. \end{aligned}$$

Taking logs, we get the log-likelihood function as

$$\ln L(\gamma_k,\beta;y_{kt},x_{kt}) = \sum_k \sum_{t=1}^{T} y_{kt} \ln\left[F(\gamma_k + x_{kt})'\beta\right] + (1-y_{kt}) \ln\left[1 - F(\gamma_k + x'_{kt}\beta)\right].$$

Essentially, as the time dimension is often limited in practice, we cannot consistently estimate the fixed effect parameters γ_k. More crucially, inconsistency in the fixed effect parameter estimation transmits to the slope parameter β and the bias might be as large as the parameter itself; in the case of $T = 2$ and $N_2 = 1$, $\hat{\beta} = 2\beta$, see also Andersen (1970) and Arellano and Honore (2001). This issue is known as the incidental parameter problem (Neyman and Scott, 1948).

6.2.3 Elimination of Fixed Effects

As seen in Chap. 1, in a linear model one can wipe out the fixed effects by appropriate orthogonal projections. However, although we do not have such a direct tool in non-linear models, we are able to eliminate these parameters considered as nuisance or incidental. This procedure was developed and first applied by Rasch (1960, 1961), Andersen (1970), Chamberlain (1980) and for a good summary, see also Arellano and Honore (2001). As soon as we eliminate the fixed effects, we are able to consistently estimate the parameters of interest β. In order to do so, we derive a sufficient statistic for the fixed effects. Then, instead of direct Maximum Likelihood Estimation (that should otherwise be avoided in the case of small T), conditional likelihood procedures are used. Conditional Maximum Likelihood is proved to be consistent for the slope parameters regardless of the number of time periods T. Moreover, under appropriate regularity conditions, the resulting estimator is asymptotically normal (see Andersen, 1970).

Let us first define what a sufficient statistic is.

Definition 1. Let Z be a random variable. $r = R(Z)$ is a sufficient statistic for the parameter θ that characterizes the random variable Z if the conditional probability distribution of the random variable Z does not depend on the parameter θ given the sufficient statistic $r = R(Z)$, that is

$$f(z \mid r, \theta) = f(z \mid r).$$

The definition simply suggests that one cannot give a better estimation of the distribution function of a random variable Z knowing the theoretical distribution parameters θ than just knowing the sufficient statistic $r = R(Z)$.

Furthermore, we have a constructive theorem that provides a simple device for obtaining a sufficient statistic in some cases (see Fisher, 1922).

Theorem 1. *Let Z be a random variable with probability density function $f(z \mid \theta)$, where θ is the parameter that characterizes the distribution of Z. Then $r = R(Z)$ is a sufficient statistic for the parameter θ if and only if there exist non-negative functions g and h such that*

$$f(z \mid \theta) = h(z)\, g(R(z), \theta).$$

In other words, the probability distribution function of Z can be factored into two multiplicatively separable non-negative parts, for which h does not depend on the parameter θ, while g depends on the parameter θ and depends on the realizations z only through the statistic $R(z)$. We will use this theorem to find a sufficient statistic for the incidental parameters.

Unfortunately, probit models cannot be estimated by a Conditional Maximum Likelihood procedure as opposed to the logit models derived below. If we started to derive the sufficient statistic, we would be faced with the problem of non-separability of the outcome variable from the fixed effects and, therefore, we would be unable to obtain the sufficient statistic:

$$
\begin{aligned}
&f(\{y_{ijt}\}_{ijt} \mid \{\gamma_{ij}\}_{ij}; \beta, \{x_{ijt}\}_{ijt}) \\
&\quad = \prod_{i=1}^{N_1}\prod_{j=1}^{N_2}\prod_{t=1}^{T} \Phi(\gamma_{ij} + x'_{ijt}\beta)^{y_{ijt}} \left[1 - \Phi(\gamma_{ij} + x'_{ijt}\beta)\right]^{1-y_{ijt}} \\
&\quad = \prod_{i=1}^{N_1}\prod_{j=1}^{N_2}\prod_{t=1}^{T} \left[\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(\gamma_{ij} + x'_{ijt}\beta)^2}{2}\right)\right]^{y_{ijt}} \\
&\qquad\qquad \times \left[1 - \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(\gamma_{ij} + x'_{ijt}\beta)^2}{2}\right)\right]^{1-y_{ijt}}
\end{aligned}
$$

from which we see that no part can be factored out containing only the data and a function of the data as desired. Hence our attention is limited to the logit framework here.[1]

Let us first consider the model described in equation (6.1). We would like to derive the sufficient statistic for the incidental parameter γ_{ij}, for which we apply the results of the Fisher-Neyman factorization theorem. For simplicity, let us have 2 periods of time, $T = 2$. Then

$$
\begin{aligned}
&f(\{y_{ijt}\}_{ijt} \mid \{\gamma_{ij}\}_{ij}; \beta, \{x_{ijt}\}_{ijt}) \\
&\quad = \prod_{i=1}^{N_1}\prod_{j=1}^{N_2}\prod_{t=1}^{2} \Lambda(\gamma_{ij} + x'_{ijt}\beta)^{y_{ijt}} \left[1 - \Lambda(\gamma_{ij} + x'_{ijt}\beta)\right] \\
&\quad = \prod_{i=1}^{N_1}\prod_{j=1}^{N_2}\prod_{t=1}^{2} \left[\frac{\exp(\gamma_{ij} + x'_{ijt}\beta)}{1 + \exp(\gamma_{ij} + x'_{ijt}\beta)}\right]^{y_{ijt}} \left[\frac{1}{1 + \exp(\gamma_{ij} + x'_{ijt}\beta)}\right]^{1-y_{ijt}} \\
&\quad = \prod_{i=1}^{N_1}\prod_{j=1}^{N_2} \left[\exp\left(\gamma_{ij}\sum_{t=1}^{2} y_{ijt} + (x_{ij1}y_{ij1} + x_{ij2}y_{ij2})'\beta\right)\right. \\
&\qquad\quad + \exp\left(\gamma_{ij}\left(\sum_{t=1}^{2} y_{ijt} + 1\right) + (x_{ij1}y_{ij1} + x_{ij2}y_{ij2} + x_{ij1})'\beta\right) \\
&\qquad\quad + \exp\left(\gamma_{ij}\left(\sum_{t=1}^{2} y_{ijt} + 1\right) + (x_{ij1}y_{ij1} + x_{ij2}y_{ij2} + x_{ij2})'\beta\right) \\
&\qquad\quad \left. + \exp\left(\gamma_{ij}\left(\sum_{t=1}^{2} y_{ijt} + 2\right) + (x_{ij1}y_{ij1} + x_{ij2}y_{ij2} + x_{ij1} + x_{ij2})'\beta\right)\right] \\
&\qquad\quad \times \left[1 + \exp(\gamma_{ij})(1 + \exp(\gamma_{ij}))\exp((x_{ij1} + x_{ij2})'\beta)\right]^{-1} \\
&\quad = \prod_{i=1}^{N_1}\prod_{j=1}^{N_2} h_{ij}(\{y_{ijt}\}_t; \beta, \{x_{ijt}\}_t)\, g_{ij}(R(\{y_{ijt}\}_t), \gamma_{ij}; \beta, \{x_{ijt}\}_t),
\end{aligned}
$$

[1] Let us note, however, that it is possible to derive the Conditional Maximum Likelihood estimation for Poisson models, for example.

where

$$h_{ij}(\{y_{ijt}\}_t;\beta,\{x_{ijt}\}_t) = \exp\left((x_{ij1}y_{ij1}+x_{ij2}y_{ij2})'\beta\right)$$
$$g_{ij}(R(\{y_{ijt}\}_t),\gamma_{ij};\beta,\{x_{ijt}\}_t) = \exp\left(\gamma_{ij}\sum_{t=1}^{2} y_{ijt}\right)$$
$$\times\left[1+\exp(\gamma_{ij}+x'_{ij1}\beta)+\exp(\gamma_{ij}+x'_{ij2}\beta)+\exp(2\gamma_{ij}+(x_{ij1}+x_{ij2})'\beta)\right]$$
$$\times\left[1+\exp(\gamma_{ij})(1+\exp(\gamma_{ij}))\exp((x_{ij1}+x_{ij2})'\beta)\right]^{-1}$$

and

$$h(\{y_{ijt}\}_{ijt};\beta,\{x_{ijt}\}_{ijt}) = \prod_{i=1}^{N_1}\prod_{j=1}^{N_2} h_{ij}(\{y_{ijt}\}_t;\beta,\{x_{ijt}\}_t)$$
$$g(R(\{y_{ijt}\}_{ijt}),\{\gamma_{ij}\}_{ij};\beta,\{x_{ijt}\}_{ijt}) = \prod_{i=1}^{N_1}\prod_{j=1}^{N_2} g_{ij}(R(\{y_{ijt}\}_t),\gamma_{ij};\beta,\{x_{ijt}\}_t).$$

Therefore, in this case, the sufficient statistic for γ_{ij} is $R(\{y_{ijt}\}_t) = \sum_{t=1}^{2} y_{ijt}$, i.e., it is enough to know what the sum of the outcome variable across time is for a specific individual unit ij. The previous calculation can be generalized to any period of time T.

As we derived the sufficient statistic for the fixed effect parameters, we can calculate the conditional probability

$$P(y_{ijt}=1 \mid \{\gamma_{ij}\},\beta,\{x_{ijt}\}_{ijt},y_{ijt}+y_{ijs}=1) = \Lambda\left[(x_{ijt}-x_{ijs})'\beta\right].$$

From this formula we could construct the conditional likelihood function to be maximized, but rather, we apply this formula to every cell of the data resulting in the following function to be maximized

$$\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\sum_{t=1}^{T}\sum_{s\in B_{ijt}} \ln\left(\frac{\exp\left[(x_{ijt}-x_{ijs})'\beta\right]}{1+\exp\left[(x_{ijt}-x_{ijs})'\beta\right]}\right),$$

where $B_{ijt} = \{s : y_{ijt}+y_{ijs}=1\}$ for each $t = 1,\ldots,T$.

It is not surprising that the estimator for this model is quite similar to Chamberlain (1980), because this model is purely two-dimensional as already discussed.

Let us now turn to the model (6.2). In this case, we would like to construct a sufficient statistic for α_{it}. We proceed as before and for simplicity let us have only 2 individual-js, $N_2 = 2$. Now we get

$$
\begin{aligned}
&f(\{y_{ijt}\}_{ijt} \mid \{\alpha_{it}\}_{it};\beta,\{x_{ijt}\}_{ijt}) \\
&\quad= \prod_{i=1}^{N_1}\prod_{j=1}^{2}\prod_{t=1}^{T}\Lambda(\alpha_{it}+x'_{ijt}\beta)^{y_{ijt}}\left[1-\Lambda(\alpha_{it}+x'_{ijt}\beta)\right] \\
&\quad= \prod_{i=1}^{N_1}\prod_{j=1}^{2}\prod_{t=1}^{T}\left[\frac{\exp(\alpha_{it}+x'_{ijt}\beta)}{1+\exp(\alpha_{it}+x'_{ijt}\beta)}\right]^{y_{ijt}}\left[\frac{1}{1+\exp(\alpha_{it}+x'_{ijt}\beta)}\right]^{1-y_{ijt}} \\
&\quad= \prod_{i=1}^{N_1}\prod_{t=1}^{T}\left[\exp\left(\alpha_{it}\sum_{j=1}^{2}y_{ijt}+(x_{i1t}y_{i1t}+x_{i2t}y_{i2t})'\beta\right)\right. \\
&\qquad\quad+\exp\left(\alpha_{it}\left(\sum_{j=1}^{2}y_{ijt}+1\right)+(x_{i1t}y_{i1t}+x_{i2t}y_{i2t}+x_{i1t})'\beta\right) \\
&\qquad\quad+\exp\left(\alpha_{it}\left(\sum_{j=1}^{2}y_{ijt}+1\right)+(x_{i1t}y_{i1t}+x_{i2t}y_{i2t}+x_{i2t})'\beta\right) \\
&\qquad\quad\left.+\exp\left(\alpha_{it}\left(\sum_{j=1}^{2}y_{ijt}+2\right)+(x_{i1t}y_{i1t}+x_{i2t}y_{i2t}+x_{i1t}+x_{i2t})'\beta\right)\right] \\
&\qquad\quad\times\left[1+\exp(\alpha_{it})(1+\exp(\alpha_{it}))\exp((x_{i1t}+x_{i2t})'\beta)\right]^{-1} \\
&\quad= \prod_{i=1}^{N_1}\prod_{t=1}^{T}h_{it}(\{y_{ijt}\}_j;\beta,\{x_{ijt}\}_j)\,g_{it}(R(\{y_{ijt}\}_j),\alpha_{it};\beta,\{x_{ijt}\}_j),
\end{aligned}
$$

where

$$
\begin{aligned}
h_{it}(\{y_{ijt}\}_j;\beta,\{x_{ijt}\}_j) &= \exp\left((x_{i1t}y_{i1t}+x_{i2t}y_{i2t})'\beta\right) \\
g_{it}(R(\{y_{ijt}\}_j),\alpha_{it};\beta,\{x_{ijt}\}_j) &= \exp\left(\alpha_{it}\sum_{j=1}^{2}y_{ijt}\right) \\
&\quad\times\left[1+\exp(\alpha_{it}+x'_{i1t}\beta)+\exp(\alpha_{it}+x'_{i2t}\beta)+\exp(2\alpha_{it}+(x_{i1t}+x_{i2t})'\beta)\right] \\
&\quad\times\left[1+\exp(\alpha_{it})(1+\exp(\alpha_{it}))\exp((x_{i1t}+x_{i2t})'\beta)\right]^{-1}
\end{aligned}
$$

and

$$
\begin{aligned}
h(\{y_{ijt}\}_{ijt};\beta,\{x_{ijt}\}_{ijt}) &= \prod_{i=1}^{N_1}\prod_{t=1}^{T}h_{it}(\{y_{ijt}\}_j;\beta,\{x_{ijt}\}_j) \\
g(R(\{y_{ijt}\}_{ijt}),\{\alpha_{it}\}_{it};\beta,\{x_{ijt}\}_{ijt}) &= \prod_{i=1}^{N_1}\prod_{t=1}^{T}g_{it}(R(\{y_{ijt}\}_j),\alpha_{it};\beta,\{x_{ijt}\}_j).
\end{aligned}
$$

We can see that in this model, the sufficient statistic for the fixed effect parameter α_{it} is unsurprisingly $R(\{y_{ijt}\}_j)=\sum_{j=1}^{2}y_{ijt}$, because of the similarity of the model to (6.1). As before, the conditional probability becomes

$$
P(y_{ijt}=1 \mid \{\alpha_{it}\},\beta,\{x_{ijt}\}_{ijt},y_{ijt}+y_{ilt}=1)=\Lambda\left[(x_{ijt}-x_{ilt})'\beta\right].
$$

Similarly to the previous model, we obtain the function to be maximized with respect to the slope parameters β by applying this formula to every cell of the data

$$\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\sum_{t=1}^{T}\sum_{l\in B_{ijt}} \ln\left(\frac{\exp\left[(x_{ijt}-x_{ilt})'\beta\right]}{1+\exp\left[(x_{ijt}-x_{ilt})'\beta\right]}\right),$$

where $B_{ijt}=\{l : y_{ijt}+y_{ilt}=1\}$ for each $j=1,\ldots,N_2$. Essentially this estimator is the same as before except that it is computationally more demanding since the individual-j dimension is usually greater than the number of time periods, $T \ll N_2$.

As seen, the fixed effects can be easily eliminated by deriving a simple sufficient statistic. In particular, the sufficient statistic we need to condition for becomes the sum of the dependent variable along the dimension in which the fixed effect parameter is invariant. Keeping this in mind, we skip the derivations on how to obtain sufficient statistics for models (6.3)–(6.6). We use the same approach to extend the method of conditional likelihood techniques, and get a function to be maximized over the parameter space of β.

Let us now continue with the conditional likelihood for model (6.3). As we allow for multiple fixed effects, we extend this logic following Charbonneau (2012), who introduced the approach in a two-dimensional setting with multiple fixed effects for different distributional assumptions on the disturbance terms (logit, poisson, negative binomial and gamma). First, let us note that model (6.3) can in fact be viewed as a sequence of two-dimensional models over time, for which in every period the modeling equation must be satisfied. To eliminate one of the fixed effects α_{kt} for all $k=1,\ldots,N_1$, we write up the conditional probability of the dependent variable $y_{kjt}=1$ conditional on the sufficient statistic derived above, that is we consider the summation over the individual-j dimension. Therefore,

$$\begin{aligned} P(y_{kjt}=1 \mid \{x_{kjt}\}_{kjt},\{\alpha_{kt}\}_{kt},\{\alpha^*_{jt}\}_{jt},y_{kjt}+y_{klt}=1) \\ =\Lambda\left[(\alpha^*_{jt}-\alpha^*_{lt})+(x_{kjt}-x_{klt})'\beta\right], \end{aligned}$$

which does not depend on α_{kt}. As this relationship holds for any $k=1,\ldots,N_1$, hence for i as well,

$$\begin{aligned} P(y_{ijt}=1 \mid \{x_{ijt}\}_{ijt},\{\alpha_{it}\}_{it},\{\alpha^*_{jt}\}_{jt},y_{ijt}+y_{ilt}=1) \\ =\Lambda\left[(\alpha^*_{jt}-\alpha^*_{lt})+(x_{ijt}-x_{ilt})'\beta\right]. \end{aligned}$$

Because these conditional probabilities still depend on the other fixed effect parameters α^*_{jt}, we need to make sure we eliminate them as well. Note that now we have a binary choice logit model on the explanatory variables $x_{ijt}-x_{ilt}$ and fixed effect parameters $\alpha^*_{jt}-\alpha^*_{lt}$. Thus, applying the usual trick will wipe out the fixed effects. Let

$$C:=\{y_{kjt}+y_{klt}=1, y_{ijt}+y_{ilt}=1\}.$$

Let us augment the set of criteria C we are conditioning upon with one more element. Then we can write up the conditional probability as

$$P(y_{kjt}=1 \mid \{x_{kjt}\}_{kjt},\{\alpha_{kt}\}_{kt},\{\alpha^*_{jt}\}_{jt},C,y_{ijt}+y_{kjt}=1)$$
$$=\Lambda\left(\left[(x_{kjt}-x_{klt})-(x_{ijt}-x_{ilt})\right]'\beta\right),$$

which does not depend on α^*_{jt}, hence we are able to estimate the slope parameters by maximizing the following function

$$\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\sum_{t=1}^{T}\sum_{(k,l)\in B_{ijt}}\ln\left(\frac{\exp\left[(x_{kjt}-x_{klt})-(x_{ijt}-x_{ilt})\right]'\beta}{1+\exp\left[(x_{kjt}-x_{klt})-(x_{ijt}-x_{ilt})\right]'\beta}\right),$$

where $B_{ijt}=\{(k,l):y_{kjt}+y_{klt}=1,y_{ijt}+y_{ilt}=1,y_{ijt}+y_{kjt}=1\}$ for each $i=1,\ldots,N_1$ and $j=1,\ldots,N_2$.

Let us now consider model (6.4), which is really two-dimensional and as such its estimation function is derived by Charbonneau (2012). When we eliminate the individual-ij fixed effects γ_{ij} by conditioning on the sum of the outcome variable across time, and time effects λ_t by conditioning on the sum of the outcome variable across the unified individual dimension (ij) just like we did before, we get

$$\sum_{(ij)}\sum_{t=1}^{T}\sum_{((ij)',s)\in B_{(ij)t}}\ln\left(\frac{\exp\left(\left[(x_{(ij)'t}-x_{(ij)'s})-(x_{(ij)t}-x_{(ij)s})\right]'\beta\right)}{1+\exp\left(\left[(x_{(ij)'t}-x_{(ij)'s})-(x_{(ij)t}-x_{(ij)s})\right]'\beta\right)}\right),$$

where $B_{(ij)t}=\{((ij)',s):y_{(ij)'t}+y_{(ij)'s}=1,y_{(ij)t}+y_{(ij)s}=1,y_{(ij)t}+y_{(ij)s}=1\}$, (ij) runs through the pairs and $t=1,\ldots,T$.

In the next model (6.5), we cannot exploit the additive nature of the individuals-i, individuals-j and the time effects, hence we end up getting exactly the same result as before in model (6.3). This is because once we condition on the sum of the dependent variable across one of the dimensions, it eliminates two kinds of fixed effects.

$$P(y_{kjt}=1 \mid \{x_{kjt}\}_{kjt},\{\alpha_k\}_k,\{\gamma_j\}_j,\{\lambda_t\}_t,y_{kjt}+y_{klt}=1)$$
$$=\Lambda\left[(\gamma_j-\gamma_l)+(x_{kjt}-x_{klt})'\beta\right]$$

for all $k=1,\ldots,N_1$ and similarly holds for i as well

$$P(y_{ijt}=1 \mid \{x_{ijt}\}_{ijt},\{\alpha_i\}_i,\{\gamma_j\}_j,\{\lambda_t\}_t,y_{ijt}+y_{ilt}=1)$$
$$=\Lambda\left[(\gamma_j-\gamma_l)+(x_{ijt}-x_{ilt})'\beta\right],$$

thus

$$P(y_{kjt}=1 \mid \{x_{kjt}\}_{kjt},\{\alpha_k\}_k,\{\gamma_j\}_j,\{\lambda_t\}_t,C,y_{ijt}+y_{kjt}=1)$$
$$=\Lambda\left(\left[(x_{kjt}-x_{klt})-(x_{ijt}-x_{ilt})\right]'\beta\right),$$

where $C=\{y_{kjt}+y_{klt}=1,y_{ijt}+y_{ilt}=1\}$ as before. Therefore, we have the same function for maximization as in the case of model (6.3).

Alternatively for convenience, symmetric conditional likelihood functions can be derived depending on the data structure available. We will summarize later what kind of data one should have in order to apply these estimation techniques. Since there are two ways for eliminating each of the fixed effects in this model (6.5), one might consider removing the fixed effects by conditioning on the sum of the outcome variable across the other free dimension. For example, eliminating α_i is possible not just by conditioning on the sum of the outcome over individual-j, but by conditioning for the sum of the outcome over time t. However, we end up having eliminated two sorts of fixed effects in the same step, just like before. Proceeding further, the function to be maximized can be derived similarly as above.

The derivation of the Conditional Maximum Likelihood Estimator for model (6.6) is based on the same logic as seen before. Since the model contains multiple fixed effects, we proceed with their elimination one by one. The sufficient statistic for the time varying individual-i fixed effect α_{it} is the sum of the dependent variable over the individual-j dimension. We condition for it and get

$$\begin{aligned} P(y_{kjs} = 1 \mid \{x_{kjs}\}_{kjs}, \{\alpha\}_{ks}, \{\alpha\}_{kj}, y_{kjs} + y_{kls} = 1) \\ = \Lambda\left[(\alpha^*_{js} - \alpha^*_{ls}) + (\gamma_{kj} - \gamma_{kl}) + (x_{kjs} - x_{kls})'\beta\right], \end{aligned}$$

which is independent of α_{ks} for all $k = 1, \ldots, N_1$ and for all $s = 1, \ldots, T$. Similarly, the conditional probability also holds for i

$$\begin{aligned} P(y_{ijs} = 1 \mid \{x_{ijs}\}_{ijs}, \{\alpha\}_{is}, \{\alpha\}_{ij}, y_{ijs} + y_{ils} = 1) \\ = \Lambda\left[(\alpha^*_{js} - \alpha^*_{ls}) + (\gamma_{ij} - \gamma_{il}) + (x_{ijs} - x_{ils})'\beta\right]. \end{aligned}$$

Let C_s denote a similar conditioning set $C_s = \{y_{kjs} + y_{kls} = 1, y_{ijs} + y_{ils} = 1\}$ as before. In the second step, we condition on the sufficient statistic for the fixed effects α_{js} to eliminate them

$$\begin{aligned} P(y_{kjs} = 1 \mid \{x_{kjs}\}_{kjs}, \{\alpha\}_{ks}, \{\alpha^*_{js}\}_{js}, \{\gamma_{kj}\}_{kj}, C_s, y_{ijs} + y_{kjs} = 1) \\ = \Lambda\left[(\gamma_{kj} - \gamma_{kl}) + (\gamma_{ij} - \gamma_{il}) + \left[(x_{kjs} - x_{kls}) - (x_{ijs} - x_{ils})\right]'\beta\right], \end{aligned}$$

which holds for every $s = 1, \ldots, T$, in particular, for t as well

$$\begin{aligned} P(y_{kjt} = 1 \mid \{x_{kjt}\}_{kjt}, \{\alpha\}_{kt}, \{\alpha^*_{jt}\}_{jt}, \{\gamma_{kj}\}_{kj}, C_t, y_{ijt} + y_{kjt} = 1) \\ = \Lambda\left[(\gamma_{kj} - \gamma_{kl}) + (\gamma_{ij} - \gamma_{il}) + \left[(x_{kjt} - x_{klt}) - (x_{ijt} - x_{ilt})\right]'\beta\right]. \end{aligned}$$

Eventually, we are able to make the following conditional probability independent of all the fixed effects

$$\begin{aligned} P(y_{kjs} = 1 \mid \{x_{kjs}\}_{kjs}, \{\alpha_{ks}\}_{ks}, \{\alpha^*_{js}\}_{js}, \{\gamma_{kj}\}_{kj}, C_s, C_t, \\ y_{ijs} + y_{kjs} = 1, y_{ijt} + y_{kjt} = 1, y_{kjt} + y_{kjs} = 1) \\ = \Lambda\left[\left((x_{kjs} - x_{kls}) - (x_{ijs} - x_{ils}) - (x_{kjt} - x_{klt}) - (x_{ijt} - x_{ilt})\right)'\beta\right]. \end{aligned}$$

Therefore, the function to be maximized can be written as

$$\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\sum_{t=1}^{T} \sum_{(k,l,s)\in B_{ijt}} \ln\Lambda\left[\left((x_{kjs}-x_{kls})-(x_{ijs}-x_{ils})-(x_{kjt}-x_{klt})-(x_{ijt}-x_{ilt})\right)'\beta\right],$$

where $B_{ijt} = \{(k,l,s) : y_{kjs}+y_{kls}=1, y_{ijs}+y_{ils}=1, y_{kjt}+y_{klt}=1, y_{ijt}+y_{ilt}=1, y_{ijs}+y_{kjs}=1, y_{ijt}+y_{kjt}=1, y_{kjt}+y_{kjs}=1\}$, $i=1,\ldots,N_1$, $j=1,\ldots,N_2$ and $t=1,\ldots,T$.

As a result of the conditional likelihood approach, from the maximization of the above functions, all the estimation results can be interpreted in relative terms, namely with the log of the odds ratios. The log of odds ratios can easily be derived, for instance, in the case of model (6.1) it is given by

$$\ln\left(\frac{P(y_{(ij)t}=1 \mid \gamma_{(ij)},\beta,x_{(ij)t})}{P(y_{(ij)t}=0 \mid \gamma_{(ij)},\beta,x_{(ij)t})}\Big/\frac{P(y_{(ij)t}=1 \mid \gamma_{(ij)},\beta,x_{(ij)'t})}{P(y_{(ij)t}=0 \mid \gamma_{(ij)},\beta,x_{(ij)'t})}\right)=(x_{(ij)t}-x_{(ij)'t})'\beta,$$

which is due to the functional form of the logistic distribution. The interpretation is that the odds of an individual with characteristics $x_{(ij)t}$ over the odds of another individual with characteristics $x_{(ij)'t}$ is calculated from $\exp\left((x_{(ij)t}-x_{(ij)'t})'\beta\right)$.

6.2.4 Caveats of the Procedure

There are some issues with the conditional likelihood approach that need special attention. These problems have already been briefly noted, but now they can be seen from the functions that enable us to obtain the estimations of the slope parameters. First, one needs to take into account the identifying variation in data. Note that whenever there is no variation in the covariates in the desired dimensions determined by the fixed effects, the numerical maximization would not converge, since the log-likelihood will be $-\infty$ (the difference in characteristics vanishes inside the logarithm). However, a more detailed database (higher dimensional data) might help overcome the problem. In this case, one might use the extra free dimension available for eliminating some fixed effects in the equations to be estimated.

A key problem with the estimation functions derived above, however, is that they are formed over just a set of observations represented by B_{ijt}. Therefore, one loses some information when excluding observations from the procedure, similarly to the linear model case with the orthogonal projections.

Table 6.1 summarizes what variation one needs for both the dependent and independent variables across dimensions to implement the estimation procedure outlined above.

Table 6.1 Variation needed in three-dimensional data to use the conditional likelihood techniques

Model	Fixed effects	Conditioning set B_{ijt}	Required variation in covariates
(6.1)	γ_{ij}	$\{s : y_{ijt} + y_{ijs} = 1\}\ \forall t$	across time t
(6.2)	α_{it}	$\{l : y_{ijt} + y_{ilt} = 1\}\ \forall j$	across individual-j
(6.3)	$\alpha_{it} + \alpha^*_{jt}$	$\{(k,l) : y_{kjt} + y_{klt} = 1, y_{ijt} + y_{ilt} = 1, y_{ijt} + y_{kjt} = 1\}\ \forall i,j$	across individual-i and j
(6.4)	$\gamma_{ij} + \lambda_t$	$\{((ij)',s) : y_{(ij)'t} + y_{(ij)'s} = 1, y_{(ij)t} + y_{(ij)s} = 1, y_{(ij)t} + y_{(ij)s} = 1\}\ \forall i,j,t$	across individual-(ij) and time t
(6.5)	$\alpha_i + \gamma_j + \lambda_t$	$\{(k,l) : y_{kjt} + y_{klt} = 1, y_{ijt} + y_{ilt} = 1, y_{ijt} + y_{kjt} = 1\}\ \forall i,j$ or $\{(k,s) : y_{kjt} + y_{kjs} = 1, y_{ijt} + y_{ijs} = 1, y_{ijt} + y_{kjt} = 1\}\ \forall i,t$ or $\{(l,s) : y_{ilt} + y_{ils} = 1, y_{ijt} + y_{ijs} = 1, y_{ijt} + y_{ilt} = 1\}\ \forall i,t$	(across individual-i and j) or (across individual-i and time t) or (across individual-j and time t)
(6.6)	$\alpha_{it} + \alpha^*_{jt} + \gamma_{ij}$	$B_{ijt} = \{(k,l,s) : y_{kjs} + y_{kls} = 1, y_{ijs} + y_{ils} = 1, y_{kjt} + y_{klt} = 1, y_{ijt} + y_{ilt} = 1, y_{ijs} + y_{kjs} = 1, y_{ijt} + y_{kjt} = 1, y_{kjt} + y_{kjs} = 1\}\ \forall i,j,t$	across individual-i and j and time t

One additional drawback of the conditional likelihood approach is that neither the partial effect of the pth covariate $\partial P(y_{ijt} = 1 \mid \gamma_{ij}, \beta, x_{ijt})/\partial x^p_{ijt}$ nor the average partial effect of the pth covariate $E(\partial P(y_{ijt} = 1 \mid \gamma_{ij}, \beta, x_{ijt})/\partial x^p_{ijt})$ can be given because the fixed effects have not been estimated and their distribution has not been specified. For simplicity, let us consider the first model (6.1), where the partial effect is given by

$$\frac{\partial P(y_{ijt} = 1 \mid \gamma_{ij}, \beta, x_{ijt})}{\partial x^p_{ijt}} = \Lambda'(\gamma_{ijt} + x'_{ijt}\beta)\beta^p\,,$$

which contains the fixed effect parameter. However, we are at least able to tell the direction of the effect, since $\Lambda' > 0$, therefore, the sign of the coefficient β^p will indicate it.

One could also worry about the computational feasibility of implementing the conditional likelihood approach. This is because the index set B_{ijt} in the formula of the conditional likelihood functions enumerates all the coordinates (index triplets) that are relevant to the estimation procedure. It can be quite large and grows with the size of the dimension for which the sufficient statistics are calculated. However, Pforr (2014) used a recursive algorithm to overcome the problem in a time-efficient way in two-dimensional panel models, which can be generalized to higher dimensions.

6.2.5 Unbalanced Panels

The conditional likelihood approach is a very convenient tool to deal with unbalanced panels. This is due to the form of the function to be maximized as it compares the characteristics of an individual in certain dimensions, and whenever some observations are missing (at random), they are simply disregarded during the procedure. They just restrict the sample for which the estimation is carried out. Therefore, the researcher needs not to worry about the data at hand being unbalanced.

6.3 Selection Bias

Let us turn now to a useful application of the binary choice models in empirical problems. One is often faced with data affected by non-random selectivity (for a survey, see Vella, 1998). This phenomenon leads to the well-known selection bias problem, which causes unreliable and inconsistent estimators if left untreated. Heckman (1979) came up with a remedy for this issue in cross-sectional data and showed how to correct for selection bias. Although sample selection bias is frequently encountered in cross sectional models, controlling for unobserved heterogeneity in panel models may not eliminate all the selection bias (like the Within transformation for the fixed effects in higher-dimensional panel models discussed in Chap. 1), see Verbeek and Nijman (1996) and Honore et al. (2008). Attrition may also result in selectivity, as seen in many empirical instances.

When estimating such models the main problem is the selection mechanism. The selection comes up as an unknown non-linear function of the observed and unobserved time-varying regressors in the model one would like to estimate. As the selection effect is time-varying, it would not disappear by simply taking the first differences or using any similar trick.

The general setup is the following in the spirit of bilateral individual fixed effects: One would like to estimate a model (called the primary or structural equation) taking into account the selection mechanism described by a different equation (called the selection or control equation). We assume a linear relationship between the dependent variable and the explanatory variables in the structural equation, just like we did in Chap. 1, and include fixed effect parameters $\tilde{\gamma}_{ij}$ controlling for the heterogeneity. The sample selection rule is assumed to follow a binary response, in which these effects are taken into account. The disturbance terms $\tilde{\varepsilon}_{ijt}$ and η_{ijt} of these equations are likely to be dependent of each other ($\mathrm{E}(\tilde{\varepsilon}_{ijt} \mid \eta_{ijt}) \neq 0$). Thus, the structural equation and the selection equation can be given as

$$\begin{aligned} y_{ijt} &= (\tilde{\gamma}_{ij} + \tilde{x}'_{ijt}\beta + \tilde{\varepsilon}_{ijt}) \cdot d_{ijt} && \text{structural equation} \\ d_{ijt} &= \mathbf{1}\{\gamma^s_{ij} + w'_{ijt}\theta + \eta_{ijt} \geq 0\} && \text{selection equation,} \end{aligned}$$

where we might have common explanatory variables in $\tilde{x}_{ijt}$ and w_{ijt}. Variables with tilde denote the latent variable counterpart of the observables. $\tilde{\gamma}_{ij}$ and γ^s_{ij} are bilat-

eral individual effects. The dependent variable y_{ijt} is observable to the researcher only if the binary indicator variable takes 1, $d_{ijt} = 1$. The researcher would like to obtain parameter estimates for β and θ based on the explanatory variables $\xi_{ij} = \{\{w_{ijt}\}_t, \{\tilde{x}_{ijt}\}_t, \tilde{\gamma}_{ij}, \gamma_{ij}\}$ treating the individual effects as nuisance parameters. The basic idea is that, in this form of the sample selection mechanism, we are able to control for both the bilateral individual and time-varying selection effects through a control function estimated from the selection equation. To see the time-varying nature of the selection, one could rewrite the structural equation in the form of

$$y_{ijt} = (\tilde{\gamma}_{ij} + \tilde{x}'_{ijt}\beta + \tilde{\varepsilon}_{ijt}) \cdot d_{ijt} = \gamma_{ijt} + x'_{ijt}\beta + \varepsilon_{ijt}.$$

When one is concerned about omitting some other form of unobserved heterogeneity as proposed in Chap. 1 and Sect. 6.2, the procedures described below can be carried out in a similar way, though they may be complicated depending on the structure of the fixed effects.

We further assume that the disturbance terms in the selection equation follow a logistic distribution in accordance with the rest of this chapter for the fixed effect estimation procedure derived above to be implementable. Thereafter, the parameters of the structural equation are estimated by a fixed effect approach discussed in Chap. 1. However, this naive estimator of β ignores sample selectivity, therefore it is inconsistent as well (see Jensen et al., 2002).

The sample selection model above might be estimated by the full information Maximum Likelihood method. The general form of the likelihood function is given by

$$L = \prod_{(ij)} \prod_t \left(\int_{-\infty}^{-\gamma_{ij}^s - w'_{ijt}\theta} f_\eta(\eta)\,\mathrm{d}\eta \right)^{1-d_{(ij)t}} \left(\int_{-\gamma_{ij}^s - w'_{ijt}\theta}^{\infty} f_{\tilde{\varepsilon},\eta}(\tilde{\varepsilon},\eta)\,\mathrm{d}\tilde{\varepsilon}\mathrm{d}\eta \right)^{d_{(ij)t}},$$

where f_η denotes the (arbitrary) probability distribution function of the disturbance terms in the selection equation η and $f_{\tilde{\varepsilon},\eta}$ stands for the (arbitrary) joint probability distribution function of the two disturbance terms. To implement this approach, the researcher needs to specify the distribution functions. In fact, if we were to allow for joint normality, we would get back Heckman's standard estimator. However, the bivariate normal assumption often proves to be a very restrictive and violated in empirical applications, which leads to serious inconsistency problems. Hasebe and Vijverberg (2012) studied several practical settings where the joint normality assumption was not satisfied.

The estimation procedure we are going to discuss is in contrast with previous works in the fixed effect framework carried out on the basis of the joint normality assumption (see, for instance, Wooldridge, 1995; Verbeek and Nijman, 1992; Honore et al., 2008).

To overcome the shortcomings of the restrictive joint normality assumption, we discuss a parametric and a semi-parametric approach to deal with selection issues with alternative joint distributions and briefly mention a non-parametric method at the end of the section.

6.3.1 Parametric Approach

The bivariate normality assumption of the disturbance terms $\tilde{\varepsilon}_{ijt}$ and η_{ijt} significantly simplifies the treatment of selection bias. This is because of a special property of the joint distribution: its marginal distributions are also normal as usually assumed in the structural and selection equations. As a departure from the bivariate normality assumption, in contrast, when one would like to relax this assumption and has a strong view about what the marginal distribution of the disturbance terms could be, the researcher often has to deal with copula functions. In our specific case, the disturbance term η_{ijt} in the selection equation follows a logistic, while the disturbance term $\tilde{\varepsilon}_{ijt}$ in the structural equation follows a normal distribution (marginally).

We assume that

- both d_{ijt} and w_{ijt} are observable – this is the sample selection;
- $(\tilde{\varepsilon}_{ijt}, \eta_{ijt}) \perp \tilde{x}_{ijt}$ with zero means – in other words, that the characteristics $\tilde{x}_{ijt}$ are exogenously given in the population;
- ε_{ijt} follows a standard normal distribution, while η_{ijt} follows a standard logistic distribution;
- γ_{ij} and γ^s_{ij} are nuisance parameters;
- the conditional expectation $\mathrm{E}(\tilde{\varepsilon}_{ijt} \mid \eta_{ijt})$ can be given with a copula representation.

We are interested in what the conditional expectation of the outcome variable is

$$\mathrm{E}(y_{ijt} \mid \tilde{x}_{ijt}, d_{ijt} = 1).$$

Let us begin with

$$\begin{aligned}\mathrm{E}(y_{ijt} \mid w_{ijt}, \gamma^s_{ij}, \eta_{ijt}, \theta) &= \tilde{\gamma}_{ij} + \tilde{x}'_{ijt}\beta + \mathrm{E}(\tilde{\varepsilon}_{ijt} \mid w_{ijt}, \gamma^s_{ij}, \eta_{ijt}, \theta) \\ &= \tilde{\gamma}_{ij} + \tilde{x}'_{ijt}\beta + \mathrm{E}(\tilde{\varepsilon}_{ijt} \mid \eta_{ijt}).\end{aligned}$$

If the last term is zero, we do not need to worry about sample selection and can proceed to the fixed effect estimation procedures as discussed in Chap. 1. However, if the last term is non-zero, the researcher faces non-random sample selection. Applying the law of iterated expectation to the equation of interest

$$\begin{aligned}\mathrm{E}(y_{ijt} \mid \tilde{x}_{ijt}, d_{ijt} = 1) &= \mathrm{E}\left[\mathrm{E}(y_{ijt} \mid \tilde{x}_{ijt}, \eta_{ijt}) \mid \tilde{x}_{ijt}, d_{ijt} = 1\right] \\ &= \mathrm{E}[\tilde{\gamma}_{ij} + \tilde{x}'_{ijt}\beta + \mathrm{E}(\tilde{\varepsilon}_{ijt} \mid \eta_{ijt}) \mid \tilde{x}_{ijt}, d_{ijt} = 1] \\ &= \tilde{\gamma}_{ij} + \tilde{x}'_{ijt}\beta + \mathrm{E}[\tilde{\varepsilon}_{ijt} \mid \eta_{ijt}, \tilde{x}_{ijt}, d_{ijt} = 1],\end{aligned}$$

where the last term represents the sample selectivity. If we assume bivariate normality and linear dependency $\mathrm{E}(\tilde{\varepsilon}_{ijt} \mid \eta_{ijt}) = \delta\eta_{ijt}$ with the measure of co-movement δ between the disturbance terms, then the previous expression would take the following form

$$\mathrm{E}(y_{ijt} \mid \tilde{x}_{ijt}, d_{ijt} = 1) = \tilde{\gamma}_{ijt} + \tilde{x}'_{ijt}\beta + \delta h(x_{ijt}),$$

where $\mathrm{E}(\eta_{ijt} \mid x_{ijt}) = h(x_{ijt})$ is the Heckman correction factor, in fact the inverse Mills ratio. Under bivariate normality, this equation is estimated in two steps: first, the Heckman correction factor is calculated, and then the primary equation, complemented with the Heckman correction factor, is estimated by an appropriate fixed effect estimator as discussed in Chap. 1.

In the case of our logistic and normal marginal distributions, the joint distribution and the dependence structure of the disturbance terms is no longer that simple:

$$\begin{aligned}\mathrm{E}(\tilde{\varepsilon} \mid \eta) &= \int_{-\infty}^{\infty} \tilde{\varepsilon}\, \frac{f(\tilde{\varepsilon}, \eta)}{\lambda(\eta)}\, \mathrm{d}\tilde{\varepsilon} \\ &= \int_{-\infty}^{\infty} \tilde{\varepsilon}\, c_{\upsilon}(\Phi(\tilde{\varepsilon}), \Lambda(\eta))\, \phi(\tilde{\varepsilon})\, \mathrm{d}\tilde{\varepsilon},\end{aligned}$$

where $f(\tilde{\varepsilon}, \eta)$ is the joint probability density function (pdf), $\lambda(\eta)$ is the pdf corresponding to the logistic distribution, $\Phi(\tilde{\varepsilon})$ is the cumulative distribution function (cdf) of the standard normal distribution, $\Lambda(\eta)$ is the cdf of the standard logistic distribution, and $\phi(\tilde{\varepsilon})$ is the pdf of the standard normal distribution. c_{υ} denotes the density of the copula C_{υ}. Since the joint distribution can be represented as

$$F(\tilde{\varepsilon}, \eta) = C_{\upsilon}(\Phi(\tilde{\varepsilon}), \Lambda(\eta))\,,$$

the joint pdf is

$$f(\tilde{\varepsilon}, \eta) = c_{\upsilon}(\Phi(\tilde{\varepsilon}), \Lambda(\eta))\, \phi(\tilde{\varepsilon})\, \lambda(\eta),$$

where $c_{\upsilon}(\Phi(\tilde{\varepsilon}), \Lambda(\eta)) = \frac{\partial^2 C_{\upsilon}(\Phi(\tilde{\varepsilon}), \Lambda(\eta))}{\partial \Phi(\tilde{\varepsilon})\, \partial \Lambda(\eta)}$. The parameter υ governs the dependency between the two disturbance terms, which should be estimated similarly to the original Heckman approach. The Appendix of this chapter briefly overviews the basic concepts with copulas.

6.3.2 Semi-Parametric Approach

Next, we propose an extension of the modeling framework analyzed by Ahn and Powell (1993) and Kyriazidou (1997) to higher dimensional panel data models while continuing to assume the model introduced at the beginning of this section. The estimation procedure involves two steps. First, one consistently estimates the unknown coefficients of the selection equation. In the second step, these are used to estimate the structural equation with weighted least squares. These weights are dependent on the magnitude of the sample selection bias, observations with less selection bias get larger weights, while observations more affected by selection bias get smaller weights. Parameter estimates from the first step help construct these weights. For tractability, we assume a simple structure of the fixed effects γ_{ij} and let $T = 2$ for now. Then the sample selection can be expressed as

$$\begin{aligned}\lambda_{ijt} &= E(\tilde{\varepsilon}_{ijt} \mid d_{ij1}=1, d_{ij2}=1, \xi_{ij}) \\ &= E(\tilde{\varepsilon}_{ijt} \mid \eta_{ij1} \le \gamma_{ij}^s + w_{ij1}'\theta, \eta_{ij2} \le \gamma_{ij}^s + w_{ij2}'\theta, \xi_{ij}) \\ &= l\left(\gamma_{ij}^s + w_{ij1}'\theta, \gamma_{ij}^s + w_{ij2}'\theta; F_{ijt}(\tilde{\varepsilon}_{ijt}, \eta_{ij1}, \eta_{ij2} \mid \xi_{ij})\right) \\ &= l_{ijt}(\gamma_{ij}^s + w_{ij1}'\theta, \gamma_{ij}^s + w_{ij2}'\theta, \xi_{ij}).\end{aligned}$$

In other words, under weak distributional assumptions, we allow the selection effect l_{ijt} to vary across bilateral individuals; so we require neither the disturbance terms $(\tilde{\varepsilon}_{ijt}, \eta_{ijt})$ be i.i.d. nor independent of the explanatory variables ξ_{ij}.

Therefore, the structural equation can be written as a partially linear regression

$$y_{ijt} = \tilde{\gamma}_{ijt} + x_{ijt}'\beta + \lambda_{ijt} + v_{ijt},$$

where $v_{ijt} = \varepsilon_{ijt} - \lambda_{ijt}$ is a new disturbance term by construction satisfying $\mathrm{E}(v_{ijt} \mid d_{ij1}=1, d_{ij2}=1, \xi_{ij}) = 0$. The idea is to get rid of the nuisance terms $\tilde{\gamma}_{ijt}$ and λ_{ijt} by differencing them out. (The appropriate Within transformation should be used for models with more sophisticated fixed effects structures.)

Under the conditional exchangeability assumption

$$F(\tilde{\varepsilon}_{ij1}, \tilde{\varepsilon}_{ij2}, \eta_{ij1}, \eta_{ij2} \mid \xi_{ij}) = F(\tilde{\varepsilon}_{ij2}, \tilde{\varepsilon}_{ij1}, \eta_{ij2}, \eta_{ij1} \mid \xi_{ij}),$$

the sample selection effect is the same across the free dimension of the fixed effect parameter γ_{ij}, $\lambda_{ij1} = \lambda_{ij2}$ for those bilateral individuals who have $w_{ij1}'\theta = w_{ij2}'\theta$

$$\begin{aligned}\lambda_{ij1} &= \mathrm{E}(\tilde{\varepsilon}_{ij1} \mid \eta_{ij1} \le \gamma_{ij}^s + w_{ij1}'\theta, \eta_{ij2} \le \gamma_{ij}^s + w_{ij2}'\theta, \xi_{ij}) \\ &= \mathrm{E}(\tilde{\varepsilon}_{ij2} \mid \eta_{ij2} \le \gamma_{ij}^s + w_{ij1}'\theta, \eta_{ij1} \le \gamma_{ij}^s + w_{ij2}'\theta, \xi_{ij}) \\ &= \lambda_{ij2}.\end{aligned}$$

Now we have two components missing from the procedure. *(i)* θ should be found to assure *(ii)* selection on observations for which the rest of the estimation is carried out. The first step, finding the estimate for θ in the binary choice equation can be resolved by, for example, the conditional likelihood approach discussed earlier in this chapter. For the second step, Kyriazidou (1997) proposes that one should weight those individuals which are relatively close to each other in terms of $w_{ijt}'\hat{\theta}$ and $w_{ijs}'\hat{\theta}$ and thus obtain the (infeasible) estimator for general T

$$\begin{aligned}\hat{\beta} = &\left[\sum_{(ij)} \frac{1}{T_{ij}-1} \sum_{s<t} \hat{\psi}_{ij} \Delta_{ts}'(x_{ij}') \Delta_{ts}(x_{ij}') d_{ijt} d_{ijs}\right]^{-1} \\ &\times \left[\sum_{(ij)} \frac{1}{T_{ij}-1} \sum_{s<t} \hat{\psi}_{ij} \Delta_{ts}'(x_{ij}') \Delta_{ts}(y_{ij}) d_{ijt} d_{ijs}\right]\end{aligned}$$

with a kernel weight $\hat{\psi}_{ij}$ decreasing to zero as $|w_{ijt}'\theta - w_{ijs}'\theta|$ increases and where Δ_{ts} denotes the difference operator between the tth and sth observation of the given variable (based on the free dimension of the fixed effect γ_{ij}). This operation in fact

corresponds to the ordinal first-differencing in a linear equation. The kernel weights are given by

$$\hat{\psi}_{ijts} = \frac{1}{h} K\left(\frac{\Delta_{ts}(w'_{ij})\hat{\theta}}{h}\right),$$

where K is a kernel density and h is a bandwidth parameter, both of them being the researcher's choice and therefore the parameters of the approach.

Kyriazidou (1997) also showed how to obtain a consistent estimator $\hat{\hat{\beta}}$. Letting $\hat{\beta}$ be an estimator with bandwidth $h = \#(ij)^{-1/(2(r+1)+1)}$, the $-1/(2(r+1)+1)$th order of the number of bilateral individuals $\#(ij)$ and $\hat{\beta}_\delta$ another estimator with bandwidth $h_\delta = \#(ij)^{-\delta/(2(r+1)+1)}$ and $\delta \in (0,1)$, we define

$$\hat{\hat{\beta}} = \frac{\hat{\beta} - \#(ij)^{-(1-\delta)(r+1)/(2(r+1)+1)}\hat{\beta}_\delta}{1 - \#(ij)^{-(1-\delta)(r+1)/(2(r+1)+1)}},$$

where r stands for an r-times continuously differentiability condition on the density of the index function $W = \Delta w'\theta$. This estimator is consistent in the number of bilateral individuals (and, clearly, in T as well) and asymptotically normally distributed.

6.3.3 Non-Parametric Approach

Fernandez-Val and Francis (2011) propose a two-step analytical large-T bias correction procedure for the sample selection model that can be written in higher dimensional panels as

$$\begin{aligned} y_{ijt} &= f(d_{ijt}, \gamma_{ij}, x_{ijt}, \lambda_{ijt}; \beta) + \varepsilon_{ijt} && \text{structural equation} \\ d_{ijt} &= g(\gamma^s_{ij}, w_{ijt}; \theta) + \eta_{ijt} && \text{selection equation,} \end{aligned}$$

where f and g are known functions up to the finite dimensional parameters β and θ, λ_{ijt} represents the control variable underlying the selection of d_{ijt} in the structural equation. The estimation procedure consists of two steps. First the selection equation should be estimated non-parametrically, from which the researcher can construct the control function λ_{ijt} in the same manner as Heckman (1979) suggests. We then account for the selection in the structural equation by including the control function in the equation. For more details, the reader is redirected to the paper cited. The authors show that the two-step bias correction method leads to reasonably low bias even in relatively short panels with $T = 6$.

6.4 Fixed Effects Multinomial Choice Models

Next, let us briefly show how to deal with multinomial choice models following Chamberlain (1980). In a multinomial choice setup, individuals are given a set of options denoted by $0,1,2,\ldots,C$ from which they can form their decision. Choices are not assumed to follow any specific order. Let us remark here that the choice set may potentially be quite large in some instances, as in the case of menu-choices, etc. Just like in binary choice models, for notational convenience, we denote the first option with 0. As we have seen in Sect. 6.2, one of the choices should be treated as a reference point.

As in the binary choice case, we model the choice probabilities where one would like to control for heterogeneity characterized by

$$P(y_{ijt}=c \mid \{\mathrm{FE}_{ijtc}\}_{ijtc},\beta,\{x_{ijtc}\}_{ijtc}) = F(\mathrm{FE}_{ijtc}+x'_{ijtc}\beta), \qquad c=0,1,\ldots,C,$$

which is in fact the *strict exogeneity assumption*, where the multiple choice counterparts of the fixed effects (6.1)–(6.6) are described as

$$\mathrm{FE}_{ijtc} = \gamma_{ijc} \tag{6.7}$$
$$\mathrm{FE}_{ijtc} = \alpha_{itc} \tag{6.8}$$
$$\mathrm{FE}_{ijtc} = \alpha_{itc} + \alpha^*_{jtc} \tag{6.9}$$
$$\mathrm{FE}_{ijtc} = \gamma_{ijc} + \lambda_{tc} \tag{6.10}$$
$$\mathrm{FE}_{ijtc} = \alpha_{ic} + \gamma_{jc} + \lambda_{tc} \tag{6.11}$$
$$\mathrm{FE}_{ijtc} = \alpha_{itc} + \alpha^*_{jtc} + \gamma_{ijc} \tag{6.12}$$

for $c=0,1,\ldots,C$ and $F(u)=\Lambda(u)$.

For the corresponding models, we require exactly the same *conditional independence assumptions* as in the case of the binary choice models.

Each of the multinomial choice problems can be broken down into a set of binary choice problems. Each possible combination of choices should be considered over the free dimension of the fixed effect parameters over which we form the sufficient statistic that eliminates them. This simply adds an additional layer (an additional dimension in fact) to the function from which we calculate the estimation of β. Let us introduce a new variable w_{ijtc} for individual-i, individual-j in period t choosing option c, in other words $w_{ijtc}=1$ if $y_{ijt}=c$ and $w_{ijtc}=0$ otherwise.

For a moment, let us consider the $T=2$ case. Then the conditional probability of exactly observing options a and b over time from the choice set $\{0,1,\ldots,C\}$ in the case of multinomial choice model (6.7) conditional on the number of occurrences is

$$P\left((w_{ij1a},w_{ij2b}) \mid (w_{ij1a},w_{ij2b}) \text{ or } (w_{ij1b},w_{ij2a}),\{\gamma_{ijc}\}_{ijc},\beta,\{x_{ijtc}\}_{ijtc}\right)$$
$$= \Lambda\left(\left[(x_{ij2b}-x_{ij2a})-(x_{ij1b}-x_{ij1a})\right]'\beta\right).$$

Therefore, we have a binary choice logit model with (w_{ij1a}, w_{ij2b}) and (w_{ij1b}, w_{ij2a}) as two alternatives and $(x_{ij2b} - x_{ij2a}) - (x_{ij1b} - x_{ij1a})$ as the explanatory variable. Similar conditional probabilities can be written for any two elements in the choice set. From this formula one could write up the Conditional Maximum Likelihood function.

The probability of observing a choice c conditional on the sufficient statistic is given by

$$\begin{aligned} &P(y_{ijt} = c \mid \{\gamma_{ijc}\}_{ijc}, \beta, \{x_{ijtc}\}_{ijtc}, w_{ijtc} + w_{ijsc} = 1) \\ &= P(w_{ijtc} = 1 \mid \{\gamma_{ijc}\}_{ijc}, \beta, \{x_{ijtc}\}_{ijtc}, w_{ijtc} + w_{ijsc} = 1) \\ &= \Lambda\left[(x_{ijtc} - x_{ijsc})'\beta\right]. \end{aligned}$$

It follows that for general T, the function to be maximized composed of cell probabilities is written as

$$\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\sum_{t=1}^{T}\sum_{c=0}^{C}\sum_{s\in B_{ijtc}} \ln\left(\frac{\exp\left[(x_{ijtc} - x_{ijsc})'\beta\right]}{1+\exp\left[(x_{ijtc} - x_{ijsc})'\beta\right]}\right),$$

where $B_{ijtc} = \{s : w_{ijtc} + w_{ijsc} = 1\}$ for each $t = 1, \ldots, T$ and $c = 0, 1, \ldots, C$.

The Conditional Likelihood function for the other models can be given similarly.

For the multinomial choice model (6.8), the function to be maximized is as follows

$$\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\sum_{t=1}^{T}\sum_{c=0}^{C}\sum_{l\in B_{ijtc}} \ln\left(\frac{\exp\left[(x_{ijtc} - x_{iltc})'\beta\right]}{1+\exp\left[(x_{ijtc} - x_{iltc})'\beta\right]}\right),$$

where $B_{ijtc} = \{l : w_{ijtc} + w_{iltc} = 1\}$ for each $j = 1, \ldots, N_2$ and $c = 0, 1, \ldots, C$. This is derived from the probabilities conditional on the number of a certain choice c made on the individual-j dimension.

Model (6.9) requires conditioning along two dimensions of the data. Once we eliminate the bilateral fixed effect α_{it} by conditioning on the number of occurrences of different choices c on the individual-j dimension, then we eliminate the other bilateral fixed effect α^*_{jt} by conditioning on the number of occurrences when choice c was chosen along the individual-i dimension. Therefore, the function to be maximized is

$$\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\sum_{t=1}^{T}\sum_{c=0}^{C}\sum_{(k,l)\in B_{ijtc}} \ln\left(\frac{\exp\left[(x_{kjtc} - x_{kltc}) - (x_{ijtc} - x_{iltc})\right]'\beta}{1+\exp\left[(x_{kjtc} - x_{kltc}) - (x_{ijtc} - x_{iltc})\right]'\beta}\right),$$

where $B_{ijtc} = \{(k,l) : w_{kjtc} + w_{kltc} = 1, w_{ijtc} + w_{iltc} = 1, w_{ijtc} + w_{kjtc} = 1\}$ for each $i = 1, \ldots, N_1$, $j = 1, \ldots, N_2$ and $c = 0, 1, \ldots, C$.

Similarly to the binary choice model (6.4), after calculating the corresponding conditional probabilities, the function in the case of the multinomial choice model (6.10) can be written as

$$\sum_{(ij)}\sum_{t=1}^{T}\sum_{c=0}^{C}\sum_{((ij)',s)\in B_{(ij)tc}} \ln\left(\frac{\exp\left(\left[(x_{(ij)'tc}-x_{(ij)'sc})-(x_{(ij)tc}-x_{(ij)sc})\right]'\beta\right)}{1+\exp\left(\left[(x_{(ij)'tc}-x_{(ij)'sc})-(x_{(ij)tc}-x_{(ij)sc})\right]'\beta\right)}\right),$$

where $B_{(ij)tc}=\{((ij)',s): w_{(ij)'tc}+w_{(ij)'sc}=1, w_{(ij)tc}+w_{(ij)sc}=1, w_{(ij)tc}+w_{(ij)sc}=1\}$, (ij) runs through the pairs, $t=1,\ldots,T$ and $c=0,1,\ldots,C$.

Just like in the binary choice case, we cannot exploit the structure of the fixed effects in model (6.11). As a result, we obtain the same function for maximization as in the case of model (6.9).

Turning to the last specification of the fixed effects (6.12), due to the repeated elimination of the fixed effects by conditioning on the number of observed choices along the corresponding free dimensions, we obtain the following function to be maximized

$$\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\sum_{t=1}^{T}\sum_{c=0}^{C}$$
$$\sum_{(k,l,s)\in B_{ijtc}} \ln\Lambda\left[\left((x_{kjsc}-x_{klsc})-(x_{ijsc}-x_{ilsc})-(x_{kjtc}-x_{kltc})-(x_{ijtc}-x_{iltc})\right)'\beta\right],$$

where $B_{ijtc}=\{(k,l,s): w_{kjsc}+w_{klsc}=1, w_{ijsc}+w_{ilsc}=1, w_{kjtc}+w_{kltc}=1, w_{ijtc}+w_{iltc}=1, w_{ijsc}+w_{kjsc}=1, w_{ijtc}+w_{kjtc}=1, w_{kjtc}+w_{kjsc}=1\}$, $i=1,\ldots,N_1$, $j=1,\ldots,N_2$, $t=1,\ldots,T$ and $c=0,1,\ldots,C$.

A stark issue arising from the Conditional Maximum Likelihood approach stems from its assumptions crucial in multinomial choice setting. We assumed the independence of the outcomes between and across groups. This is equivalent to the independence of irrelevant alternatives assumption, which often fails in empirical applications. The researcher needs to pay special attention to this assumption, since if it is not satisfied, it invalidates the estimation results.

References

Ahn, H. and Powell, J. L. (1993). Semiparametric estimation of censored selection models with a nonparametric selection mechanism. *Journal of Econometrics*, 58(1–2):3–29.

Amemiya, T. (1985). *Advanced Econometrics*. Harvard University Press, Cambridge, Massachusetts.

Andersen, E. B. (1970). Asymptotic Properties of Conditional Maximum-Likelihood Estimators. *Journal of the Royal Statistical Society. Series B*, 32(2):283–301.

Arellano, M. and Bonhomme, S. (2009). Robust priors in nonlinear panel data models. *Econometrica*, 77(2):489–536.

Arellano, M. and Honore, B. (2001). Panel data models: some recent developments. In Heckman, J. and Leamer, E., editors, *Handbook of Econometrics*, volume 5, chapter 53, pages 3229–3296. Elsevier.

Bonhomme, S. (2012). Functional differencing. *Econometrica*, 80:1337–1385.

Chamberlain, G. (1980). Analysis of Covariance with Qualitative Data. *Econometrica*, 47(1):225–238.

Chamberlain, G. (1984). Panel data. In Griliches, Z. and Intriligator, M. D., editors, *Handbook of Econometrics*, volume 2, chapter 22, pages 1247–1318. Elsevier.

Charbonneau, K. (2012). Multiple Fixed Effects in Nonlinear Panel Data Models. *Unpublished Manuscipt*.

Charbonneau, K. (2014). Multiple fixed effects in binary response panel data models. *Bank of Canada Working Paper 2014-17*.

Fernandez-Val, I. and Francis, V. (2011). Bias corrections for two-step fixed effects panel data estimators. *Journal of Econometrics*, 163:144–162.

Fernandez-Val, I. and Weidner, M. (2016). Individual and time effects in nonlinear panel data models with large N, T. *Journal of Econometrics*, 192(1):291–312.

Fisher, R. A. (1922). On the Mathematical Foundations of Theoretical Statistics. *Philosophical Transactions of the Royal Society, Series A*, 222:309–368.

Hahn, J. and Newey, W. (2004). Jackknife and analytical bias reduction for nonlinear panel models. *Econometrica*, 72(4):1295–1319.

Hasebe, T. and Vijverberg, W. (2012). A flexible sample selection model: A GLT-copula approach. IZA Discussion Papers 7003, Institute for the Study of Labor (IZA).

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, 47:153–161.

Honore, B., Francis, V., and Marno, V. (2008). Attrition, Selection Bias and Censored Regressions. In Matyas, L. and Sevestre, P., editors, *The Econometrics of Panel Data*, Advanced Studies in Theoretical and Applied Econometrics, chapter 7, pages 385–418. Springer, 3rd edition.

Jensen, P., Rosholm, M., and Verner, M. (2002). A comparison of different estimators for panel data sample selection models. Working Papers 02-1, University of Aarhus, Aarhus School of Business, Department of Economics.

Joe, H. (1993). Parametric families of multivariate distributions with given margins. *Journal of Multivariate Analysis*, 46(2):262–282.

Joe, H. (1997). *Multivariate Models and Dependence Concepts*. Chapman and Hall, London.

Kyriazidou, E. (1997). Estimation of a panel data sample selection model. *Econometrica*, 65:1335–1364.

Lechner, M., Lollivier, S., and Magnac, T. (2008). Parametric Binary Choice Models. In Matyas, L. and Sevestre, P., editors, *The Econometrics of Panel Data*, Advanced Studies in Theoretical and Applied Econometrics, chapter 7, pages 215–243. Springer, 3rd edition.

Lee, L.-F. (1982). Some approaches to the correlation of selectivity bias. *The Review of Econometric Studies*, 49(157):355–372.

Lee, L.-F. (1983). Generalized econometric models with selectivity. *Econometrica*, 51(2):507–512.

Manski, C. F. (1975). Maximum score estimation of the stochastic utility model of choice. *Journal of Econometrics*, 3:205–228.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behaviour. In Zarembka, P., editor, *Frontiers in Econometrics*, pages 105–142. Academic Press New York, New York, NY, USA.

Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica*, 46:69–85.

Nelsen, R. B. (2006). *An Introduction to Copulas*. Springer, New York, 2nd edition edition.

Neyman, J. and Scott, E. L. (1948). Consistent Estimates Based on Partially Consistent Observations. *Econometrica*, 16(1):1–32.

Pakes, A., Porter, J., Ho, K., and Ishii, J. (2015). Moment inequalities and their application. *Econometrica*, 83:315–334.

Pforr, K. (2014). femlogit—implementation of the multinomial logit model with fixed effects. *The Stata Journal*, 14(4):847–862.

Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests*. Danish National Institute for Educational Research, Copenhagen.

Rasch, G. (1961). On General Laws and the Meaning of Measurement in Psychology. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, 4:321–333.

Sklar, A. (1959). Fonctions de répartition à *n* dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris*, 8:229–231.

Smith, M. D. (2003). Modelling sample selection using archimedean copulas. *Econometrics Journal*, 6(1):99–123.

Trivedi, P. K. and Zimmer, D. M. (2006). Using trivariate copulas to model sample selection and treatment effects: Application to family health care demand. *Journal of Business & Economic Statistics*, 24(1):63–76.

Trivedi, P. K. and Zimmer, D. M. (2007). *Copula Modeling: An Introduction for Practitioners*. New Publishers Inc., Delft, The Netherlands.

Vella, F. (1998). Selection corrections for panel data models under conditional mean independence assumptions. *University of Wisconsin Press*, 33(1):127–169.

Verbeek, M. and Nijman, T. (1992). Testing for selectivity bias in panel data models. *International Economic Review*, 33(3):681–703.

Verbeek, M. and Nijman, T. E. (1996). Incomplete panels and selection bias. In Matyas, L. and Sevestre, P., editors, *The Econometrics of Panel Data. Handbook of the Theory with Applications*, pages 442–490. Kluwer Academoic Publishers, Dordrecht.

Wooldridge, J. M. (1995). Selection corrections for panel data models under conditional mean independence assumptions. *Journal of Econometrics*, 68(1):115–132.

Appendix

Here, we briefly introduce the essential properties of copulas, for a deeper treatment the reader is referred to Nelsen (2006), Joe (1993, 1997) and Trivedi and Zimmer (2007). Developing estimation procedures using generalized distributional assumptions has been initiated by Lee (1982, 1983), further analyzed by Smith (2003) and Trivedi and Zimmer (2006). Lee relaxes the joint normality assumption, while Smith introduces the copula approach to selection mechanisms, which is in fact the generalized case of Lee's solution allowing for more flexible model specification.

The copula approach is a modelling tool and provides a successful method when one would like to model non-linear dependencies with alternative marginal distributions. In this procedure, the researcher needs to specify the desired marginal distribution functions and a function that binds them together, called a copula. The copula function describes the complex dependency structure between the random variables. The key property of this approach is that it does not just encapsulates the multivariate normal distribution, but can be applied to any specific distributions. The theoretical foundation is given by Sklar's theorem (see Sklar, 1959; Nelsen, 2006).

Definition 2. Let $X = (X_1, \ldots, X_n)$ be a random vector with distribution function F and with marginal distribution functions F_i, $X_i \sim F_i$, $i = 1, \ldots, n$. A distribution function C with marginals on $[0,1]$ is called a copula of X if

$$F = C(F_1, \ldots, F_n).$$

Theorem 2. *Let $F \in \mathscr{F}(F_1, \ldots, F_n)$ be an n-dimensional distribution function with marginals $F_1, \ldots, F_n$. Then there exists a copula C such that the n-dimensional joint distribution function can be represented as*

$$F(x_1, \ldots, x_n) = C(F_1(x_1), \ldots, F_n(x_n)).$$

Moreover, the representation is unique whenever the marginal distributions are continuous. Intuitively, Sklar's theorem above states that there exists a copula function C which represents the joint cumulative distribution function F of the random vector X in terms of their underlying marginal distributions F_i, $i = 1, \ldots, n$, which is given exogenously by the researcher. One might, for example, write up the copula representation for uniform random variables U_i, $i = 1, \ldots, n$, using the probability integral transformation

$$\begin{aligned} F(x_1, \ldots, x_n) &= P(X_1 \leq x_1, \ldots, X_n \leq x_n) \\ &= P(U_1 \leq F_1(x_1), \ldots, U_n \leq F_n(x_n)) \\ &= C(u_1 = F_1(x_1), \ldots, u_n = F_n(x_n)). \end{aligned}$$

Hence, copulas can be easily generated by the inversion method, assuming that the joint and marginal distributions are given

$$C(u_1, \ldots, u_n) = F(x_1 = F_1^{-1}(u_1), \ldots, x_n = F_n^{-1}(u_n)).$$

Note that, implicitly in the representation equation in Sklar's theorem for the bivariate case, $C(u,v) = 0$ if either or both u and v are zero, and $C(1,v) = v$ as well as $C(u,1) = u$, where $(u,v) \in [0,1]^2$. In our case with a structural and a selection equation, the copula function is a two-dimensional object.

For statistical purposes, it is reasonable to parametrize a copula function in a way that it captures the level of association between the underlying random variables of interest. Let these parameters be denoted by υ for the bivariate distribution. In such a case, we can write a copula in the following form

$$C_\upsilon(u,v).$$

This notation represents a family of copulas. Provided that the marginal distributions F_1 and F_2 do not depend on υ, the representation in the theorem holds for all members of the family.

We have three copulas of special importance, which are

$$\begin{aligned} \Pi &= uv, && \text{product copula} \\ W &= \max\{u+v-1,0\}, && \text{Fréchet lower bound} \\ M &= \min\{u,v\}, && \text{Fréchet upper bound} \end{aligned}$$

where $(u,v) \in [0,1]^2$. The product copula corresponds to the stochastic independence, that is if two random variables X and Y are independent, then Π is the copula of their joint distribution, $\Pi = uv = F_X(x)F_Y(y) = F_{X,Y}(x,y)$. The closed interval $[W,M]$ contains all the bivariate copulas, so for all C on $[0,1]^2$ we have

$$W \leq C \leq M.$$

Fréchet bounds determine the coverage of a given family of copulas. A family of copulas is said to be comprehensive if that family includes all three special copulas amongst its members at least in the limiting case.

It is worth mentioning that statisticians characterize dependence with another measure than the traditional Pearson's product moment correlation coefficient $\rho \in [-1,1]$. This is because of its limitation, $\rho = 0$ does not necessarily imply independence between two random variables. Therefore, we rely on the notion of concordance (discordance), which means that large values of one random variable are associated with large (small) values of the other random variable and small values of one random variable are associated with small (large) values of the other. This concept leads to two measures of dependence generally used in the literature, Kendall's τ and Spearman's ρ_S. For independent pairs of (X_i,Y_i), $i = 1,2,3$, which are copies of (X,Y), they are defined as

$$\tau = P\left((X_1 - X_2)(Y_1 - Y_2) > 0\right) - P\left((X_1 - X_2)(Y_1 - Y_2) < 0\right),$$

intuitively, the probability of concordance minus the probability of discordance, while

$$\rho_S = 3\left[P\left((X_1 - X_2)(Y_1 - Y_3) > 0\right) - P\left((X_1 - X_2)(Y_1 - Y_3) < 0\right)\right].$$

Both τ and ρ_S are bounded between $[-1,1]$ and both are equal to -1 at W, 0 at Π (hence indicates independence) and 1 at M. If (X,Y) are a pair of continuous random variables, that is $F(x,y) = C_\upsilon(u = F_X(x), v = F_Y(y))$, these concordance measures simplify to

$$\tau = 4 \iint_{[0,1]^2} C(u,v)\,\mathrm{d}C(u,v) - 1 = 4E(C(U,V)) - 1$$
$$\rho_S = 12 \iint_{[0,1]^2} uv\,\mathrm{d}C(u,v) - 3 = 12E(UV) - 3,$$

where $U \sim U[0,1]$ and $V \sim U[0,1]$ random variables with joint cdf C.

With these notions in mind, we turn back to the econometric estimation of the structural equation affected by selectivity. Let the joint distribution of the latent variables $\tilde{y}_{ijt} = \tilde{\gamma}_{ij} + \tilde{x}'_{ijt}\beta + \tilde{\varepsilon}_{ijt}$ and $\tilde{d}_{ijt} = \gamma^s_{ij} + w'_{ijt}\theta + \eta_{ijt}$ be denoted by $F(\tilde{d}_{ijt}, \tilde{y}_{ijt})$, and the marginals be denoted by $F_{\tilde{y}}(\tilde{y})$ and $F_{\tilde{d}}(\tilde{d})$. Amemiya (1985) showed that the likelihood for the selection model is given by

$$L = \prod_{ijt:d_{ijt}=0} P(\tilde{d}_{ijt}) \prod_{ijt:d_{ijt}=1} f_{\tilde{y}|\tilde{d}}(y_{ijt} \mid \tilde{d}_{ijt} \geq 0) P(\tilde{d}_{ijt} \geq 0),$$

where $f_{\tilde{y}|\tilde{d}}$ denotes the probability density function of $\tilde{y}_{ijt}$ given its observability, that is $\tilde{d}_{ijt} \geq 0$. Dropping the indices for convenience, this conditional density in the formula can be rewritten as

$$\begin{aligned} f_{\tilde{y}|\tilde{d}}(y \mid \tilde{d} \geq 0) &= \frac{1}{1 - F_{\tilde{d}}(0)} \frac{\partial}{\partial y}\left(F_{\tilde{y}}(y) - F(0,y)\right) \\ &= \frac{1}{1 - F_{\tilde{d}}(0)} \left(f_{\tilde{y}}(y) - \frac{\partial}{\partial y} F(0,y)\right), \end{aligned}$$

where $F_{\tilde{d}}(0) = P(\tilde{d} < 0) = P(d = 0)$. After substituting back into the likelihood function, we get

$$L = \prod_{d=0} F_{\tilde{d}}(0) \prod_{d=1} \left(f_{\tilde{y}}(y) - \frac{\partial}{\partial y} F(0,y)\right). \tag{6.13}$$

Note that, if $\tilde{y}$ and $\tilde{d}$ were independent of each other, then we would obtain $\frac{\partial}{\partial y} F(0,y) = F_{\tilde{d}}(0) f_{\tilde{y}}(y)$, and thus the likelihood could be separated as

$$L = \left(\prod_{d=0} F_{\tilde{d}}(0) \prod_{d=1} (1 - F_{\tilde{d}}(0))\right) \times \left(\prod_{d=1} f_{\tilde{y}}(y)\right)$$

However, the most difficult part in the likelihood function (6.13) is the partial derivative of the joint distribution $\frac{\partial}{\partial y} F(0,y)$ that contains the complicated relationship be-

tween the underlying variables $\tilde{y}$ and $\tilde{d}$, which leads to selection bias if the researcher is unaware of it. We give a solution to this issue below.

Here we display a non-comprehensive list of popular copulas employed in empirical studies along with their basic properties. For an extensive list of available copulas and their defining properties, see Nelsen (2006).

Table 6.2 Dependence parameter υ and Kendall's τ of alternative copulas

Copula name	Range of υ	Kendall's $\tau(\upsilon)$	Range of τ
Gaussian	$-1 \leq \upsilon \leq 1$	$2\sin^{-1}(\upsilon)/\pi$	$-1 \leq \tau \leq 1$
FGM	$-1 \leq \upsilon \leq 1$	$2\upsilon/9$	$-2/9 \leq \tau \leq 2/9$
Archimedean class			
Frank	$-\infty < \upsilon < \infty$	$1-4[1-D_1(\upsilon)]/\upsilon$	$-1 < \tau < 1$
Clayton	$0 \leq \upsilon < \infty$	$\upsilon/(\upsilon+2)$	$0 \leq \tau < 1$
Gumbel	$1 \leq \upsilon < \infty$	$(\upsilon-1)/\upsilon$	$0 \leq \tau < 1$
Joe	$1 \leq \upsilon < \infty$		$0 \leq \tau < 1$

Note that $D_1(\upsilon)$ is the Debye function, $D_1(\upsilon) = \frac{1}{\upsilon}\int_0^{\upsilon} \frac{t}{e^t-1}\,\mathrm{d}t$, for Joe, there is no closed form of Kendall's τ.

We highlight some additional properties for the copulas included in the Table 6.2. The Gaussian family assumes radial symmetry and asymptotic independence between the random variables, because tail events are rare. FGM assumes the same as the Gaussian family, though its formulation is simpler, but can capture only a limited concordance between the random variables, Kendall's τ cannot be bigger than $2/9$ in absolute terms. Therefore, it is not a comprehensive family of copulas either. Frank copula in turn offers a full range of concordance, though it is useful if the researcher would like to capture the central dependency of random variables in a radially symmetric way, as on the tails, the dependency is low (lower than in the case of the Gaussian family). Clayton, Gumbel and Joe families imply radial asymmetry and fat tailed distributions, hence are not comprehensive. Clayton has a strong left tail and weak right tail dependence as opposed to Gumbel and Joe, which have a weak left tail and strong right tail dependence (stronger right tail dependence in the case of Joe). The researcher should have a firm intuition about the sign of the concordance (negativity of the Kendall's τ measure is easily achieved by reparametrization of the disturbance term $\varepsilon_{ijt} = -\varepsilon^*_{ijt}$). Therefore, the researcher can select from a wide range of distributional properties (dependency structure, comprehensiveness, radial (a)symmetry, asymptotic tail (in)dependence) to be captured by the model in a flexible way. Hasebe and Vijverberg (2012) propose the GLT-copula, that is also flexible.

Based on these properties, the researcher can select the desired copula formulation, but there is a particularly elegant treatment of the joint probability distribution addressed by the Archimedean class of copulas. Common to all the members of this class is that there is an additive generator function $\varphi : [0,1] \to [0,\infty]$ (the range of φ is the extended version of the non-negative part of the real line), which is a continuous, convex and decreasing function, with terminal $\varphi(1) = 0$. The copula C is

assumed to be generated according to the following relationship:

$$\varphi(C(u,v)) = \varphi(u) + \varphi(v).$$

From this characterization, it can be seen that the class is particularly useful in the case of limited dependent variable models, because of the ability to reduce the dimensionality of the econometric problem at hand. Generator functions are summarized in table 6.3.

Whenever the terminal $\varphi(0) = \infty$, we say that the generator φ is termed strict, and its inverse φ^{-1} exists, so

$$C(u,v) = \varphi^{-1}(\varphi(u) + \varphi(v)),$$

ora pseudo-inverse function should be introduced. A relevant result for Archimedean copulas in our modeling framework is that its derivative can be given by

$$\frac{\partial}{\partial v} C_\upsilon(u,v) = \frac{\varphi'(v)}{\varphi'(C_\upsilon(u,v))},$$

which makes this class so popular in empirical work. This partial derivative can be found in Table 6.3 for various Archimedean copulas.

For Archimedean copulas, the partial derivative in the likelihood function (6.13) can be derived as follows

$$\begin{aligned}\frac{\partial}{\partial y} F(0,y) &= \frac{\partial}{\partial v} C_\upsilon(F_{\tilde{d}}(0), v)\Big|_{v \to F_{\tilde{y}}(y)} \times \frac{\partial F_{\tilde{y}}(y)}{\partial y} \\ &= \frac{\varphi'(F_{\tilde{y}}(y))}{\varphi'(C_\upsilon)} \times f_{\tilde{y}}(y),\end{aligned}$$

where C_υ denotes $C_\upsilon(F_{\tilde{d}}, F_{\tilde{y}}) = C_\upsilon(F_{\tilde{d}}(0), F_{\tilde{y}}(y))$, which is obtained from the inverse of the generator $\varphi^{-1}(\varphi(F_{\tilde{d}}(0)) + \varphi(F_{\tilde{y}}(y)))$. Since the functional form of the derivative of the generator function φ is generally easy to derive, the likelihood estimation can also be implemented, and the likelihood function (6.13) becomes

$$L = \prod_{d=0} F_{\tilde{d}}(0) \prod_{d=1} \left(1 - \frac{\varphi'(F_{\tilde{y}}(y))}{\varphi'(C_\upsilon)}\right) f_{\tilde{y}}(y). \tag{6.14}$$

Finally, this function can be maximized by standard quasi-Newton procedures.

Although the likelihood formulation of the copula estimation procedure (6.14) is quite attractive, there are some drawbacks the researcher needs to face. The computation of the Hessian of this function is troublesome, so the second derivatives should be numerically approximated in order to use the Newton-Raphson algorithm. The researcher should also pay attention to the possibility of multiple maxima. Another issue is also connected to the Hessian: distributional assumptions become hard to test in the absence of the information matrix. As a result, the practitioner can only

Table 6.3 Generator function $\varphi(t)$ and the expression for $\frac{\partial C_\upsilon(u,v)}{\partial v}$ for Archimedean class of copulas

Copula name	Generator $\varphi(t)$	Expression for $\frac{\partial C_\upsilon(u,v)}{\partial v}$
Frank	$-\ln\frac{e^{-\upsilon t}-1}{e^{-\upsilon}-1}$	$[1-e^{\upsilon C_\upsilon(u,v)}](1-e^{\upsilon v})^{-1}$
Clayton	$\frac{1}{\upsilon}(t^{-\upsilon}-1)$	$v^{-(\upsilon+1)}(u^{-\upsilon}+v^{-\upsilon}-1)^{-\frac{1+\upsilon}{\upsilon}}$
Gumbel	$(-\ln t)^{\upsilon}$	$v^{-1}(-\ln v)^{\upsilon-1}C_\upsilon(u,v)[(-\ln u)^{\upsilon}+(-\ln v)^{\upsilon}]^{\frac{1}{\upsilon}-1}$
Joe	$-\ln(1-(1-t)^{\upsilon})$	$(1-v)^{\upsilon-1}(1-(1-u)^{\upsilon})[(1-u)^{\upsilon}+(1-v)^{\upsilon}-(1-u)^{\upsilon}(1-v)^{\upsilon}]^{\frac{1}{\upsilon}-1}$

rely on pairwise comparison of models based, for example, on standard information criteria (AIC, BIC).

Chapter 7
Nonparametric Models with Random Effects

Yiguo Sun, Wei Lin, and Qi Li

Abstract This chapter considers the three-dimensional nonparametric models with random effects, and proposes pooled local linear and two-step estimators for them. We find that the pooled local linear estimator can be inconsistent when the sum of all the error term covariances in absolute values diverges to infinity too quickly. When the pooled local linear estimator is consistent, the optimal convergence rate of the estimator, its corresponding optimal bandwidth and asymptotic variance depend on the number of regressors and the limit of certain sample indices ratio; and we propose an asymptotically more efficient two-step estimator along the line of Su et al. (2013). Some extensions on nonparametric models with fixed effects, mixed effects, and higher dimensions are also discussed.

7.1 Introduction

In past few decades, panel data analysis techniques have become part of classical econometric data analysis skills that every applied economist must grasp due to their promising and powerful capability to handle complex social economic data as compared to pure cross-sectional or time-series data analysis techniques. Sun et al. (2015) have recently reviewed estimation of nonparametric panel models, and this chapter extends Sun et al.'s (2015) work to the estimation of high-dimensional panel data models in a nonparametric regression framework.

Yiguo Sun
University of Guelph, Guelph, Canada

Wei Lin
Capital University of Economics and Business, Beijing, China

Qi Li
Capital University of Economics and Business, Beijing, China and Texas A&M University, Texas, USA

L. Matyas (ed.), *The Econometrics of Multi-dimensional Panels*, Advanced Studies in Theoretical and Applied Econometrics 50, DOI 10.1007/978-3-319-60783-2_7

Specifically, this chapter considers a three-dimensional nonparametric panel data model with random effects

$$y_{ijt} = m(\mathbf{x}_{ijt}) + u_{ijt}, \tag{7.1}$$

for $i = 1,\ldots,N_1$, $j = 1,\ldots,N_2$, and $t = 1,\ldots,T$, where index t refers to time periods, indexes i and j denote cross-sectional units, $\mathbf{x}_{ijt}$ is a $(k\times 1)$ strictly exogenous vector of continuous variables, both y_{ijt} and u_{ijt} are scalar, u_{ijt} is a random error with zero mean, and $m(x) = \mathrm{E}(y_{ijt}|\mathbf{x}_{ijt} = x)$ is a smooth unknown function to be estimated. The error terms $\{u_{ijt}\}$ may exhibit weak cross-sectional dependence and weak dependence across time.

To the best of our knowledge, there is no existing literature working on three-dimensional nonparametric panel data models. Therefore, this chapter contains brand new material in nonparametric estimation of three-dimensional nonparametric panel data models with random effects. This is the reason why we cannot cover as many topics as Sun et al. (2015), who provide a survey for nonparametric estimation of two-dimensional nonparametric panel data models.

The rest of the chapter is organized as follows. We propose a pooled local linear estimator of model (7.1) for a general error structure in Sect. 7.2. Then, in Sect. 7.3, we introduce an asymptotically more efficient two-step estimator to take into account possible cross-sectional dependence and serial correlation in error terms. In Sect. 7.4, focusing on a pairwise random-effects error structure, we discuss the consistency condition and the asymptotic properties of the pooled local linear estimator and mixed fixed and random effects modelling. In Sect. 7.5, we give a very brief discussion on possible extensions of our proposed estimation method to panel data models with four or higher dimensions. Section 7.6 concludes. We provide a brief mathematical proof in the Appendix.

We first introduce some notation frequently used throughout this chapter. (i) For a k-dimensional variable x, $g^{(j)}(x)$ denotes the jth order derivative of $g(x)$ with respect to x, for example, $g^{(1)}(x) = \partial g(x)/\partial x$ is a $(k\times 1)$ vector of first derivative functions, $g^{(2)}(x) = \partial g(x)/\partial x\partial x'$ is a $(k\times k)$ matrix of second derivative functions; (ii) M denotes a generic positive constant that may take different values at different places; (iii) we denote

$$\mathbb{N} = N_1N_2T, \quad \mathbb{N}_{\max} = \max(N_1,N_2,T), \quad \sum_{ijt} = \sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\sum_{t=1}^{T},$$

$(ijt)_l = i_l j_l t_l$, and $(ijt)_l' = i_l' j_l' t_l'$ for any positive integer l; (iv) let $\mathbf{I}_n$, ι_n and $\mathbf{0}_n$ denote an $(n\times n)$ identify matrix, an $(n\times 1)$ vector of ones, and an $(n\times 1)$ vector of zeros, respectively; (v) $a_n = O_e(1)$ means that $a_n = O_p(1)$ but a_n is not $o_p(1)$; (vi) $A_n \approx B_n$ means that $A_n = B_n(1+o_p(1))$; (vii) $A_n \asymp B_n$ means that $c_1B_n \le A_n \le c_2B_n$ (or let the probability of this event approaches one as $n\to\infty$) for some constant $0 < c_1 < c_2 < \infty$; (viii) $\|\cdot\|$ refers to the Euclidean norm.

7.2 The Pooled Local Linear Estimator

We denote

$$\mathbf{y}=\left[\mathbf{y}_1',...,\mathbf{y}_{N_1}'\right]',\quad \mathbf{y}_i=\left[\mathbf{y}_{i1}',...,\mathbf{y}_{iN_2}'\right]',\quad \mathbf{y}_{ij}=\left[y_{ij1},...,y_{ijT}\right]',$$
$$\mathbf{X}=\left[\mathbf{x}_1',...,\mathbf{x}_{N_1}'\right]',\quad \mathbf{x}_i=\left[\mathbf{x}_{i1}',...,\mathbf{x}_{iN_2}'\right]',\quad \text{and}\quad \mathbf{x}_{ij}=\left[\mathbf{x}_{ij1},...,\mathbf{x}_{ijT}\right]',$$

where $\mathbf{y}$ is an $(\mathbb{N}\times 1)$ vector, $\mathbf{y}_i$'s are $(N_2T\times 1)$ vectors, $\mathbf{y}_{ij}$'s are $(T\times 1)$ vectors, $\mathbf{X}$ is an $(\mathbb{N}\times k)$ matrix, $\mathbf{x}_i$'s are $(N_2T\times k)$ matrices, and $\mathbf{x}_{ij}$'s are $(T\times k)$ matrices. In addition, $\mathbf{u}$ and $m(\mathbf{X})$ stack up $\{u_{ijt}\}$ and $\{m(\mathbf{x}_{ijt})\}$, respectively, into an $(\mathbb{N}\times 1)$ vector conforming to $\mathbf{y}$. To sum up, the panel data is recorded in the following ascending order; index i first, index j second and index t third. Rewriting model (7.1) in matrix form gives

$$\mathbf{y}=m(\mathbf{X})+\mathbf{u}, \tag{7.2}$$

where $\mathrm{E}(\mathbf{u}|\mathbf{X})=\mathbf{0}_{\mathbb{N}}$, and $\Omega_u=\mathrm{Var}(\mathbf{u}|\mathbf{X})$ is an $(\mathbb{N}\times\mathbb{N})$ variance-covariance matrix in a general form.

We are interested in estimating $m(\mathbf{x}_0)$ at an interior point, $\mathbf{x}_0=\left[x_{1,0},...,x_{k,0}\right]'$. In doing so, we assume that both $m(\cdot)$ and $f_{ijt}(\cdot)$ are twice continuously differentiable in the neighborhood of $\mathbf{x}_0$ for all i, j, and t and that

$$\bar{f}_0(\mathbf{x}_0)=\lim_{\mathbb{N}\to\infty}\mathbb{N}^{-1}\sum_{ijt}f_{ijt}(\mathbf{x}_0)>c_0>0$$

for some finite constant c_0, where $f_{ijt}(\mathbf{x})$ is the probability density function of $\mathbf{x}_{ijt}$ evaluated at $\mathbf{x}_{ijt}=\mathbf{x}$. In addition, denoting

$$\bar{f}(\mathbf{x})=\mathbb{N}^{-1}\sum_{ijt}f_{ijt}(\mathbf{x})\quad\text{and}\quad\bar{f}^{(j)}(\mathbf{x})=\mathbb{N}^{-1}\sum_{ijt}f_{ijt}^{(j)}(\mathbf{x})$$

for $j=1,2$, we assume that $\bar{f}(\mathbf{x})$, $\bar{f}^{(1)}(\mathbf{x})$ and $\bar{f}^{(2)}(\mathbf{x})$ are all bounded in the neighborhood of $\mathbf{x}_0$. Moreover, denoting $f_{(ijt)_1(ijt)_2}(\mathbf{x}_1,\mathbf{x}_2)$ to be the joint probability density function of $\left(\mathbf{x}_{i_1j_1t_1},\mathbf{x}_{i_2j_2t_2}\right)$ evaluated at $\left(\mathbf{x}_{i_1j_1t_1},\mathbf{x}_{i_2j_2t_2}\right)=(\mathbf{x}_1,\mathbf{x}_2)$, we assume that $f_{(ijt)_1(ijt)_2}(\mathbf{x}_1,\mathbf{x}_2)$ is twice continuously differentiable and uniformly bounded in the neighborhood of $(\mathbf{x}_0,\mathbf{x}_0)$ across all $(ijt)_1\neq(ijt)_2$.

Below, we first propose a pooled local linear estimator of $m(\mathbf{x}_0)$ under the "working independence"condition, which ignores cross-sectional dependence and serial correlation in the error terms. For two-dimensional nonparametric panel data models with random effects, the literature has shown that the estimator based on the "working independence" condition is not asymptotically efficient. We therefore propose a two-step estimator that takes into account the non-diagonal structure of Ω_u, and show that the two-step estimator is asymptotically more efficient than the simple pooled local linear estimator in Sect. 7.3.

Applying the Taylor's expansion to $m(\mathbf{x}_{ijt})$ at the interior point $\mathbf{x}_0$ gives

$$m(\mathbf{x}_{ijt}) = m(\mathbf{x}_0) + m^{(1)}(\mathbf{x}_0)'(\mathbf{x}_{ijt} - \mathbf{x}_0) + (\mathbf{x}_{ijt} - \mathbf{x}_0)' m^{(2)}(\bar{\mathbf{x}}_{ijt})(\mathbf{x}_{ijt} - \mathbf{x}_0)/2,$$

where $\bar{\mathbf{x}}_{ijt}$ lies between $\mathbf{x}_{ijt}$ and $\mathbf{x}_0$. We can approximate model (7.2) by

$$\begin{aligned} \mathbf{y} &= m(\mathbf{x}_0)\iota_{\mathbb{N}} + \left(\mathbf{X} - \mathbf{x}_0' \otimes \iota_{\mathbb{N}}\right) m^{(1)}(\mathbf{x}_0) + Re(\mathbf{X}, \mathbf{x}_0) + \mathbf{u} \\ &\equiv \alpha\iota_{\mathbb{N}} + \left(\mathbf{X} - \mathbf{x}_0' \otimes \iota_{\mathbb{N}}\right)\beta + Re(\mathbf{X}, \mathbf{x}_0) + \mathbf{u}, \end{aligned} \tag{7.3}$$

where $Re(\mathbf{X}, \mathbf{x}_0) = m(\mathbf{X}) - m(\mathbf{x}_0)\iota_{\mathbb{N}} - (\mathbf{X} - \mathbf{x}_0' \otimes \iota_{\mathbb{N}})m^{(1)}(\mathbf{x}_0)$. When $\left\|\mathbf{x}_{ijt} - \mathbf{x}_0\right\|$ is close to zero for all i, j, t, one can ignore the term $Re(\mathbf{X}, \mathbf{x}_0)$ and treat model (7.3) as a linear regression model, which is achieved by multiplying (7.3) by a kernel weight function. This (local) linear model provides a good approximation of nonlinear (and nonparametric) model (7.2) for a sufficiently large number of observations. Taking $\iota_{\mathbb{N}}$ and $\mathbf{X} - (\mathbf{x}_0' \otimes \iota_{\mathbb{N}})$ as the regressors, one can estimate $\alpha(\equiv m(\mathbf{x}_0))$ and $\beta(\equiv m^{(1)}(\mathbf{x}_0))$ by weighted least squares method. This is the so-called *local linear regression estimation method* in the nonparametric econometrics literature.

To obtain the consistent estimator of α and β, we introduce a kernel function to select index (i, j, t) satisfying $\left\|\mathbf{x}_{ijt} - \mathbf{x}_0\right\| = o(1)$. Specifically, we define an $(\mathbb{N} \times \mathbb{N})$ diagonal kernel function $\mathbf{K}_{\mathbf{H}}(\mathbf{x}_0)$ with a typical diagonal element equal to

$$K\left(\mathbf{H}^{-1}(\mathbf{x}_{ijt} - \mathbf{x}_0)\right), \quad \text{where} \quad K(u) = \prod_{l=1}^{k} k(u_l)$$

and $k(u)$ is a symmetric probability density function over interval $[-1, 1]$. Thus, $K\left(\mathbf{H}^{-1}(\mathbf{x}_{ijt} - \mathbf{x}_0)\right)$ gives a positive value only if

$$\left\|\mathbf{H}^{-1}(\mathbf{x}_{ijt} - \mathbf{x}_0)\right\|_{max} \leq 1$$

and zero otherwise, where for a $(k \times 1)$ vector

$$\mathbf{a} = [a_1, ..., a_k]', \quad \|\mathbf{a}\|_{max} = \max_{1 \leq i \leq k} |a_k|.$$

In addition, $\mathbf{H} = \text{diag}\{h_1, \ldots, h_k\}$ is a $(k \times k)$ diagonal *bandwidth* matrix used to control the size of the neighborhood of $\mathbf{x}_0$ as we can see that $\left\|\mathbf{x}_{ijt} - \mathbf{x}_0\right\| = o(1)$ for

$$\left\|\mathbf{H}^{-1}(\mathbf{x}_{ijt} - \mathbf{x}_0)\right\|_{\max} \leq 1 \quad \text{if} \quad \|\mathbf{H}\| = o(1), \quad \text{where} \quad \|\mathbf{H}\| = \sqrt{\sum_{l=1}^{k} h_l^2}.$$

Now, we estimate α and β by minimizing the following kernel weighted least squares objective function

$$\hat{\gamma} = \arg\min_{\gamma} (\mathbf{y} - \mathscr{X}\gamma)' \mathbf{K}_{\mathbf{H}}(\mathbf{x}_0)(\mathbf{y} - \mathscr{X}\gamma),$$

where we denote $\gamma = [\alpha, \beta']'$, $\hat{\gamma} = \left[\hat{\alpha}, \hat{\beta}'\right]'$, and

$$\mathscr{X} = \left[\iota_{\mathbb{N}}, \mathbf{X} - \mathbf{x}_0' \otimes \iota_{\mathbb{N}}\right]. \tag{7.4}$$

The *pooled local linear estimator* of γ is given by

$$\hat{\gamma} = \left[\mathscr{X}'\mathbf{K}_{\mathbf{H}}(\mathbf{x}_0)\mathscr{X}\right]^{-1}\mathscr{X}'\mathbf{K}_{\mathbf{H}}(\mathbf{x}_0)\mathbf{y}\,, \tag{7.5}$$

where $\hat{m}(\mathbf{x}_0) = \hat{\alpha}$ and $\hat{m}^{(1)}(\mathbf{x}_0) = \hat{\beta}$ estimate $m(\mathbf{x}_0)$ and $m^{(1)}(\mathbf{x}_0)$, respectively. Let $\Pi(\bar{\mathbf{x}})$ be an $(\mathbb{N}\times 1)$ vector conformable to $\mathbf{y}$ with a typical element equal to

$$\Pi_{ijt} = (\mathbf{x}_{ijt} - \mathbf{x}_0)' m^{(2)}(\bar{\mathbf{x}}_{ijt})(\mathbf{x}_{ijt} - \mathbf{x}_0)\ .$$

With (7.2) and (7.5) we have

$$\hat{\gamma} - \gamma = \left[\mathscr{X}'\mathbf{K}_{\mathbf{H}}(\mathbf{x}_0)\mathscr{X}\right]^{-1}\mathscr{X}'\mathbf{K}_{\mathbf{H}}(\mathbf{x}_0)\left[\Pi(\bar{\mathbf{x}})/2 + \mathbf{u}\right] \equiv \mathscr{A}^{-1}\left(\mathscr{B}/2 + \mathscr{C}\right)\ ,$$

where we denote

$$\mathscr{A} = \mathscr{X}'\mathbf{K}_{\mathbf{H}}(\mathbf{x}_0)\mathscr{X}, \quad \mathscr{B} = \mathscr{X}'\mathbf{K}_{\mathbf{H}}(\mathbf{x}_0)\Pi(\bar{\mathbf{x}}), \quad \text{and} \quad \mathscr{C} = \mathscr{X}'\mathbf{K}_{\mathbf{H}}(\mathbf{x}_0)\mathbf{u}\,.$$

Here, $\mathscr{A}^{-1}\mathscr{B}/2$ and $\mathscr{A}^{-1}\mathscr{C}$ are referred to as the *bias* term and *variance* term of $\hat{\gamma}$, respectively.

Applying simple algebra gives

$$\mathscr{X}'\mathbf{K}_{\mathbf{H}}(\mathbf{x}_0)\mathscr{X} = \begin{bmatrix} \sum_{ijt} K_{ijt} & \sum_{ijt} K_{ijt}(\mathbf{x}_{ijt} - \mathbf{x}_0)' \\ \sum_{ijt} K_{ijt}(\mathbf{x}_{ijt} - \mathbf{x}_0) & \sum_{ijt} K_{ijt}(\mathbf{x}_{ijt} - \mathbf{x}_0)(\mathbf{x}_{ijt} - \mathbf{x}_0)' \end{bmatrix}$$

and

$$\mathscr{X}'\mathbf{K}_{\mathbf{H}}(\mathbf{x}_0)\mathbf{z} = \sum_{ijt} K_{ijt}\mathbf{z}_{ijt}\left[1\ (\mathbf{x}_{ijt} - \mathbf{x}_0)'\right]'$$

for any $(\mathbb{N}\times 1)$ vector $\mathbf{z}$, where we denote $K_{ijt} \equiv K\left(\mathbf{H}^{-1}(\mathbf{x}_{ijt} - \mathbf{x}_0)\right)$ for notation simplicity.

To derive the consistency of the pooled local linear estimator, we assume that $\|\mathbf{H}\| \to 0$ and $\mathbb{N}|\mathbf{H}| \to \infty$ as $\mathbb{N} \to \infty$ in the rest of this section, where $|\mathbf{H}| = h_1 h_2 \ldots h_k$. Then, straightforward calculations lead to

$$\begin{aligned} |\mathbf{H}|^{-1}\mathrm{E}(K_{ijt}) &= f_{ijt}(\mathbf{x}_0) + \kappa_{12}\mathrm{tr}\left\{\mathbf{H}f_{ijt}^{(2)}(\mathbf{x}_0)\mathbf{H}\right\}(1 + o(1)) \\ |\mathbf{H}|^{-1}\mathrm{E}\left(K_{ijt}^2\right) &= \nu_0 f_{ijt}(\mathbf{x}_0) + \kappa_{22}\mathrm{tr}\left\{\mathbf{H}f_{ijt}^{(2)}(\mathbf{x}_0)\mathbf{H}\right\}(1 + o(1)) \end{aligned}$$

uniformly over i, j,t, where we denote $\kappa_{l_1 l_2} = \int k^{l_1}(u)u^{l_2}du$ and $\nu_0 = \int K^2(\mathbf{u})d\mathbf{u}$ and $\mathrm{tr}(A)$ defines the trace of matrix A. Furthermore, we have

$$\mathrm{Var}\left(\sum_{ijt} K_{ijt}\right) = \sum_{ijt}\mathrm{Var}(K_{ijt}) + R_n\ ,$$

where

$$R_n = \sum_{ijt}\sum_{t'\neq t}\mathrm{Cov}\left(K_{ijt},K_{ijt'}\right)+\sum_{ijt}\sum_{j'\neq j}\mathrm{Cov}\left(K_{ijt},K_{ij't}\right)+\sum_{ijt}\sum_{i'\neq i}\mathrm{Cov}\left(K_{ijt},K_{i'jt}\right)$$
$$+\sum_{ijt}\sum_{j'\neq j}\sum_{t'\neq t}\mathrm{Cov}\left(K_{ijt},K_{ij't'}\right)+\sum_{ijt}\sum_{i'\neq i}\sum_{t'\neq t}\mathrm{Cov}\left(K_{ijt},K_{i'jt'}\right)$$
$$+\sum_{ijt}\sum_{i'\neq i}\sum_{j'\neq j}\mathrm{Cov}\left(K_{ijt},K_{i'j't}\right)+\sum_{ijt}\sum_{i'}\sum_{j't'\neq ijt}\mathrm{Cov}\left(K_{ijt},K_{i'j't'}\right), \tag{7.6}$$

i.e., R_n contains all the covariance terms. It is straightforward to obtain

$$\sum_{ijt}\mathrm{Var}\left(K_{ijt}\right)\approx \mathbb{N}|\mathbf{H}|\,\nu_0\bar{f}(\mathbf{x}_0)=O(\mathbb{N}|\mathbf{H}|)\,,$$

where the notation $A_n\approx B_n$ means that $A_n=B_n(1+o(1))$. Assuming that $\left\{\mathbf{x}_{ijt}\right\}$ is weakly dependent across three indices such that R_n is asymptotically negligible relative to $\sum_{ijt}\mathrm{Var}\left(K_{ijt}\right)$, i.e.,

$$R_n = o\left(\sum_{ijt}\mathrm{Var}\left(K_{ijt}\right)\right)\,, \tag{7.7}$$

we obtain

$$\mathrm{Var}\left(\frac{1}{\mathbb{N}|\mathbf{H}|}\sum_{ijt}K_{ijt}\right)=O\left((\mathbb{N}|\mathbf{H}|)^{-1}\right)\,,$$

and then

$$\frac{1}{\mathbb{N}|\mathbf{H}|}\sum_{ijt}K_{ijt}=\bar{f}\left(\mathbf{x}_0\right)+k_{1,2}\mathrm{tr}\left\{\mathbf{H}\bar{f}^{(\mathbf{2})}\left(\mathbf{x_0}\right)\mathbf{H}\right\}+o_p\left(\|\mathbf{H}\|^2\right)+O_p\left((\mathbb{N}|\mathbf{H}|)^{-1/2}\right).$$

Similarly, we obtain

$$\frac{1}{\mathbb{N}|\mathbf{H}|}\sum_{ijt}K_{ijt}\mathbf{H}^{-1}\left(\mathbf{x}_{ijt}-\mathbf{x}_0\right)=\mathbf{H}\kappa_{12}\bar{f}^{(1)}\left(\mathbf{x}_0\right)+o_p\left(\|\mathbf{H}\|^2\right)+O_p\left((\mathbb{N}|\mathbf{H}|)^{-1/2}\right)$$

and

$$\begin{aligned}&\tfrac{1}{\mathbb{N}|\mathbf{H}|}\textstyle\sum_{ijt}K_{ijt}\mathbf{H}^{-1}\left(\mathbf{x}_{ijt}-\mathbf{x}_0\right)\left(\mathbf{x}_{ijt}-\mathbf{x}_0\right)'\mathbf{H}^{-1}\\&\quad=\kappa_{12}\bar{f}\left(\mathbf{x}_0\right)\mathbf{I}_k+O_p\left(\|\mathbf{H}\|^2+\left(\mathbb{N}|\mathbf{H}|\right)^{-1/2}\right)\,.\end{aligned}$$

Hence, denoting

$$D_n=\mathrm{diag}\left\{1,\mathbf{H}\right\}=\mathrm{diag}\left\{1,h_1,...,h_k\right\}\,,$$

a $((k+1)\times(k+1))$ diagonal matrix, we have

$$\frac{D_n^{-1}\mathscr{A}}{\mathbb{N}|\mathbf{H}|}\xrightarrow{p}\bar{f}_0\left(\mathbf{x}_0\right)\begin{bmatrix}1 & \mathbf{0}_k'\\ \mathbf{0}_k & \kappa_{12}\mathbf{I}_k\end{bmatrix} \tag{7.8}$$

and

$$\frac{D_n^{-1}\mathscr{B}}{2\mathbb{N}|\mathbf{H}|}=\frac{\kappa_{12}}{2}\begin{bmatrix}\mathrm{tr}\left\{\mathbf{H}\mathbf{m}^{(2)}(\mathbf{x}_0)\mathbf{H}\right\}\bar{f}(\mathbf{x}_0)\\ \mathbf{0}_k\end{bmatrix}+o_p\left(\|\mathbf{H}\|^2\right). \tag{7.9}$$

Note that condition (7.7) is a high level assumption ensuring the validity of the conventional results (7.8) and (7.9). Evidently, condition (7.7) holds if $\{\mathbf{x}_{ijt}\}$ is uncorrelated across all three indexes. Additionally, (7.7) can also hold with Proposition 2 in El Machkouri et al. (2013) if $\{\mathbf{x}_{ijt}\}$ is a stationary random field satisfying $\mathbf{x}_{ijt}=g\left(\varepsilon_{ijt-i'j't'}\right)$ and 2-stable, where $\{\varepsilon_{ijt}\}$ is a sequence of i.i.d. random variables with $\varepsilon_{ijt-i'j't'}=\varepsilon_{(i-i')(j-j')(t-t')}$, $g(\cdot)$ is a measurable function. (Please refer to definition 2 in El Machkouri et al. (2013) for 2-stable.) For illustration purposes, if $g(\cdot)$ is a linear function or $\mathbf{x}_{ijt}=\sum_{i'j't'}a_{i'j't'}\varepsilon_{ijt-i'j't'}$, $\{\mathbf{x}_{ijt}\}$ is 2-stable if $\sum_{i'j't'}\left|a_{i'j't'}\right|<\infty$. In addition, Tran (1990) shows that condition (7.7) holds if $\{\mathbf{x}_{ijt}\}$ is a sequence of strong mixing random fields with mixing coefficients, $\{\varphi_i\}$, satisfying

$$\sum_{i=1}^{\infty}i^2\varphi_i^a<M<\infty$$

for some $a\in(0,1/2)$, and

$$\left|f_{(ijt),(ijt)'}(\mathbf{x}_1,\mathbf{x}_2)-f_{ijt}(\mathbf{x}_1)f_{i'j't'}(\mathbf{x}_2)\right|\leq M<\infty$$

for all $(\mathbf{x}_1,\mathbf{x}_2)$, i,j,t, and i',j',t'. Recently, Jenish (2012) has verified (7.8) and (7.9) for the cases when $\{\mathbf{x}_{ijt}\}$ is a sequence of near-epoch dependent random fields on a strong mixing input process.

Now, we consider $\mathscr{C}$, for which we have $\mathrm{E}\left[D_n^{-1}\mathscr{X}'\mathbf{K}_{\mathbf{H}}(\mathbf{x}_0)\mathbf{u}\right]=\mathbf{0}_{k+1}$ and

$$\mathrm{Var}\left[D_n^{-1}\mathscr{X}'\mathbf{K}_{\mathbf{H}}(\mathbf{x}_0)\mathbf{u}\right]=D_n^{-1}\mathrm{E}\left[\mathscr{X}'\mathbf{K}_{\mathbf{H}}(\mathbf{x}_0)\Omega_u\mathbf{K}_{\mathbf{H}}(\mathbf{x}_0)\mathscr{X}\right]D_n^{-1}.$$

Before deriving the asymptotic results for $\mathscr{C}$, we assume:

(i) The conditional variance of u_{ijt} satisfies that

$$\mathrm{E}(u_{ijt}^2|\mathbf{X})=\mathrm{E}(u_{ijt}^2|\mathbf{x}_{ijt})\equiv\sigma_{ijt}^2(\mathbf{x}_{ijt}),$$

and the conditional covariance is

$$\mathrm{E}(u_{i_1j_1t_1}u_{i_2j_2t_2}|\mathbf{x}_{(ijt)_1},\mathbf{x}_{(ijt)_2})=\sigma_{(ijt)_1(ijt)_2}(\mathbf{x}_{(ijt)_1},\mathbf{x}_{(ijt)_2})$$

for all $(ijt)_1\neq(ijt)_2$;

(ii) $f_{ijt}(\cdot)$ and $f_{(ijt)_1(ijt)_2}(\cdot,\cdot)$ are twice continuously differentiable in the neighborhood of $\mathbf{x}_0$ and $(\mathbf{x}_0,\mathbf{x}_0)$, respectively;

(iii) $\sigma_{ijt}^2(\mathbf{x})$ and its first- and second-order derivatives are uniformly bounded across all i, j, t in the neighborhood of $\mathbf{x}_0$, and $\sigma_{(ijt)_1(ijt)_2}(\mathbf{x}_{(ijt)_1},\mathbf{x}_{(ijt)_2})$ is uniformly bounded for all $(ijt)_1\neq(ijt)_2$ in the neighborhood of $(\mathbf{x}_0,\mathbf{x}_0)$.

Under these conditions, we obtain

$$
\begin{aligned}
&(\mathbb{N}|\mathbf{H}|)^{-1}\ \mathrm{E}\left[\iota_{\mathbb{N}}'\mathbf{K}_{\mathbf{H}}(\mathbf{x}_0)\Omega_u\mathbf{K}_{\mathbf{H}}(\mathbf{x}_0)\iota_{\mathbb{N}}\right]\\
&\quad=\tfrac{1}{\mathbb{N}|\mathbf{H}|}\textstyle\sum_{i_1j_1t_1}\sum_{i_2j_2t_2}\mathrm{E}\left[\sigma_{(ijt)_1(ijt)_2}(\mathbf{x}_0,\mathbf{x}_0)K_{i_1j_1t_1}K_{i_2j_2t_2}\right]\\
&\quad\approx\nu_0\mathbb{N}^{-1}\textstyle\sum_{ijt}\sigma_{ijt}^2(\mathbf{x}_0)f_{ijt}(\mathbf{x}_0)\\
&\qquad+\mathbb{N}^{-1}|\mathbf{H}|\textstyle\sum_{(ijt)_1}\sum_{(ijt)_2\neq(ijt)_1}\sigma_{(ijt)_1(ijt)_2}(\mathbf{x}_0,\mathbf{x}_0)f_{(ijt)_1(ijt)_2}(\mathbf{x}_0,\mathbf{x}_0)\ ,
\end{aligned}
\tag{7.10}
$$

$$
\begin{aligned}
&(\mathbb{N}|\mathbf{H}|)^{-1}\ \mathrm{E}\left[\mathbf{H}^{-1}(\mathbf{X}-\mathbf{x}_0'\otimes\iota_{\mathbb{N}})'\mathbf{K}_{\mathbf{H}}(\mathbf{x}_0)\Omega_u\mathbf{K}_{\mathbf{H}}(\mathbf{x}_0)\iota_{\mathbb{N}}\right]\\
&\quad=(\mathbb{N}|\mathbf{H}|)^{-1}\textstyle\sum_{i_1j_1t_1}\sum_{i_2j_2t_2}\mathrm{E}\left[\sigma_{(ijt)_1(ijt)_2}(\mathbf{x}_0,\mathbf{x}_0)K_{i_1j_1t_1}K_{i_2j_2t_2}\mathbf{H}^{-1}\times\right.\\
&\qquad\left.\times(\mathbf{x}_{i_2j_2t_2}-\mathbf{x}_0)\right]\\
&\quad\approx\kappa_{22}\mathbb{N}^{-1}\mathbf{H}\textstyle\sum_{ijt}\sigma_{ijt}^2(\mathbf{x}_0)f_{ijt}^{(1)}(\mathbf{x}_0)\\
&\qquad+\kappa_{12}\mathbb{N}^{-1}|\mathbf{H}|\textstyle\sum_{i_1j_1t_1}\sum_{(ijt)_2\neq(ijt)_1}\sigma_{(ijt)_1(ijt)_2}(\mathbf{x}_0,\mathbf{x}_0)\mathbf{H}\frac{\partial f_{(ijt)_1(ijt)_2}(\mathbf{x}_0,\mathbf{x}_0)}{\partial\mathbf{x}_{(ijt)_2}}\ ,
\end{aligned}
\tag{7.11}
$$

and

$$
\begin{aligned}
&(\mathbb{N}|\mathbf{H}|)^{-1}\ \mathbf{H}^{-1}\mathrm{E}\left[(\mathbf{X}-\mathbf{x}_0'\otimes\iota_{\mathbb{N}})'\mathbf{K}_{\mathbf{H}}(\mathbf{x}_0)\Omega_u\mathbf{K}_{\mathbf{H}}(\mathbf{x}_0)(\mathbf{X}-\mathbf{x}_0{}'\otimes\iota_{\mathbb{N}})\right]\mathbf{H}^{-1}\\
&\quad=(\mathbb{N}|\mathbf{H}|)^{-1}\textstyle\sum_{i_1j_1t_1}\sum_{i_2j_2t_2}\mathrm{E}\left[\sigma_{(ijt)_1(ijt)_2}(\mathbf{x}_0,\mathbf{x}_0)K_{i_1j_1t_1}K_{i_2j_2t_2}\right.\\
&\qquad\left.\times\mathbf{H}^{-1}(\mathbf{x}_{i_1j_1t_1}-\mathbf{x}_0)(\mathbf{x}_{i_2j_2t_2}-\mathbf{x}_0)'\mathbf{H}^{-1}\right]\\
&\quad\approx\kappa_{22}\mathbf{I}_k\mathbb{N}^{-1}\textstyle\sum_{ijt}\sigma_{ijt}^2(\mathbf{x}_0)f_{ijt}(\mathbf{x}_0)+\kappa_{12}^2\mathbb{N}^{-1}|\mathbf{H}|\times\\
&\qquad\times\textstyle\sum_{i_1j_1t_1}\sum_{(ijt)_2\neq(ijt)_1}\sigma_{(ijt)_1(ijt)_2}(\mathbf{x}_0,\mathbf{x}_0)\mathbf{H}\frac{\partial^2 f_{(ijt)_1(ijt)_2}(\mathbf{x}_0,\mathbf{x}_0)}{\partial\mathbf{x}_{(ijt)_1}\partial\mathbf{x}_{(ijt)_2}'}\mathbf{H}\ .
\end{aligned}
\tag{7.12}
$$

We then obtain the conventional result

$$
\mathrm{Var}\left(\frac{D_n^{-1}\mathscr{C}}{\sqrt{\mathbb{N}|\mathbf{H}|}}\right)\approx\Sigma(\mathbf{x}_0)\begin{bmatrix}\nu_0 & \mathbf{0}_k'\\ \mathbf{0}_k & \kappa_{22}\mathbf{I}_k\end{bmatrix}\ ,
\tag{7.13}
$$

where we denote

$$
\Sigma(\mathbf{x}_0)=\lim_{\mathbb{N}\to\infty}\mathbb{N}^{-1}\sum_{ijt}\sigma_{ijt}^2(\mathbf{x}_0)f_{ijt}(\mathbf{x}_0)\ ,
\tag{7.14}
$$

if the following assumption holds true,

$$
\mathbb{N}^{-1}|\mathbf{H}|\sum_{i_1j_1t_1}\sum_{(ijt)_2\neq(ijt)_1}\left|\sigma_{(ijt)_1(ijt)_2}(\mathbf{x}_0,\mathbf{x}_0)\right|=o(1),
\tag{7.15}
$$

in the neighborhood of $(\mathbf{x}_0,\mathbf{x}_0)$. Then, taking (7.8), (7.9) and (7.13) together, we obtain

$$
\hat{m}(\mathbf{x}_0)-m(\mathbf{x}_0)=O_p\left(\|\mathbf{H}\|^2+(\mathbb{N}|\mathbf{H}|)^{-1/2}\right),
\tag{7.16}
$$

so that the optimal bandwidth will be $h_l=c_l\mathbb{N}^{-1/(4+k)}$ for $l=1,\ldots,k$, where c_l's are positive constants. It follows that $\hat{m}(\mathbf{x}_0)$ is a consistent estimator of $m(\mathbf{x}_0)$ with an asymptotic bias of order $O_p\left(\|\mathbf{H}\|^2\right)$ and an asymptotic variance of order $O_p\left((\mathbb{N}|\mathbf{H}|)^{-1}\right)$ provided that $\|\mathbf{H}\|\to 0$ and $\mathbb{N}|\mathbf{H}|\to\infty$ as $\mathbb{N}\to\infty$.

The result in (7.16) is well expected from the nonparametric literature. However, this result crucially depends on condition (7.15) — another high level assumption which restricts the order of sample size $\mathbb{N}$, bandwidth $\mathbf{H}$, and the degree of dependence in error terms. This condition holds trivially if $\{u_{ijt}\}$ is neither serial nor cross-sectional dependence in all three indices. For another example, if u_{ijt} is weakly dependent across all three indices in the sense that

$$\sum_{(ijt)' \neq (ijt)} |\sigma_{(ijt)(ijt)'}(\mathbf{x}_0, \mathbf{x}_0)| < M < \infty$$

for all i, j, t, then we have

$$\sum_{i_1 j_1 t_1} \sum_{(ijt)_2 \neq (ijt)_1} |\sigma_{(ijt)_1 (ijt)_2}(\mathbf{x}_0, \mathbf{x}_0)| = O(\mathbb{N}) ,$$

and condition (7.15) still holds. In fact, condition (7.15) allows the sum of absolute covariances approaches to infinity slightly faster than $\mathbb{N}$ but slower than $\mathbb{N}|\mathbf{H}|^{-1}$. However, condition (7.15) can fail to hold in many cases. For the pairwise random-effects error structure (7.24), which is discussed in detail in Sect. 7.4, we show that

$$\mathbb{N}^{-1} |\mathbf{H}| \sum_{i_1 j_1 t_1} \sum_{(ijt)_2 \neq (ijt)_1} \left| \sigma_{(ijt)_1 (ijt)_2}(\mathbf{x}_0, \mathbf{x}_0) \right| = O(\max(N_1, N_2, T)|\mathbf{H}|),$$

which may converge to zero or a positive finite number, or even diverge to infinity. Naturally, (7.16) may not hold when condition (7.15) is violated, and we will delay our discussion on this issue to Sect. 7.4.

For now, we obtain (7.8), (7.9), (7.13), and (7.16). We conjecture that under some regularity conditions, a central limit theorem applies to a weighted average of u_{ijt} so that

$$(\mathbb{N}|\mathbf{H}|)^{-1/2} D_n^{-1} \mathscr{X}' \mathbf{K}_{\mathbf{H}}(\mathbf{x}_0) \mathbf{u} \xrightarrow{d} \mathscr{N}\left(\mathbf{0}_{k+1}, \Sigma(\mathbf{x}_0) \begin{bmatrix} \nu_0 & \mathbf{0}_k' \\ \mathbf{0}_k & \kappa_{22} \mathbf{I}_k \end{bmatrix} \right), \tag{7.17}$$

which gives

$$\begin{aligned} &\sqrt{\mathbb{N}|\mathbf{H}|} D_n \left(\hat{\gamma} - \gamma - \begin{bmatrix} \kappa_{12} \mathrm{tr}\left\{ \mathbf{H} \mathbf{m}^{(2)}(\mathbf{x}_0) \mathbf{H} \right\} /2 \\ \mathbf{0}_k \end{bmatrix} \right) \\ &\xrightarrow{d} \mathscr{N}\left(\mathbf{0}_{k+1}, \frac{\Sigma(\mathbf{x}_0)}{\bar{f}_0^2(\mathbf{x}_0)} \begin{bmatrix} \nu_0 & \mathbf{0}_k' \\ \mathbf{0}_k & \kappa_{12}^{-2} \kappa_{22} \mathbf{I}_k \end{bmatrix} \right) \end{aligned}$$

and

$$\sqrt{\mathbb{N}|\mathbf{H}|} \left(\hat{m}(\mathbf{x}_0) - m(\mathbf{x}_0) - \kappa_{12} \mathrm{tr}\left\{ \mathbf{H} \mathbf{m}^{(2)}(\mathbf{x}_0) \mathbf{H} \right\} /2 \right) \xrightarrow{d} \mathscr{N}\left(0, \nu_0 \bar{f}_0^{-2}(\mathbf{x}_0) \Sigma(\mathbf{x}_0)\right) \tag{7.18}$$

under the assumption that

$$\lim_{\mathbb{N}\to\infty} \mathbb{N}\,|\mathbf{H}|\,\|\mathbf{H}\|^4 = c > 0\,,$$

which is a finite constant. The validity of the asymptotic result in (7.17) depends on the error structure of $\{u_{ijt}\}$. If $\{u_{ijt}\}$ is independent across all three indexes, (7.17) holds by Liapounov's central limit theorem and the Cramér-Wold device if

$$\mathrm{E}\left(\left|u_{ijt}\right|^{2+\delta}\right) < M < \infty \quad \text{for some} \quad \delta > 0\,.$$

If $\{u_{ijt}\}$ exhibits weak dependence over one or more indexes, one has to apply a proper central limit theorem for arrays of random fields here; e.g., Dedecker (1998), Jenish and Prucha (2009) and El Machkouri et al. (2013) for strong mixing random fields, and Jenish (2012) for near-epoch dependent random fields.

7.3 Two-step Local Linear Estimator

In this section, we state that the pooled local linear estimator is asymptotically inefficient if u_{ijt} in model (7.1) is weakly dependent cross sections and/or across times even if the estimator is consistent, because the pooled local linear estimator ignores the non-zero covariances in the error term. For panel data with two dimensions, Lin and Carroll (2000), Ruckstuhl et al. (2000), Henderson and Ullah (2005), Martins-Filho and Yao (2009), Su and Ullah (2007), Su et al. (2013), Yao and Li (2013), among others, have attempted to improve upon the pooled local linear estimator. These papers all consider panel data with a sufficiently large number of cross sectional units and a finite number of time periods. To summarize the existing literature, we divide these papers into three groups in terms of different estimation methodologies used.

(i). A *weighted local linear estimator* that is defined as

$$\tilde{\gamma} = \arg\min_{\gamma} (\mathbf{y} - \mathscr{X}\gamma)' \mathbf{W}_{\mathbf{H}}(\mathbf{x}_0) (\mathbf{y} - \mathscr{X}\gamma)\ ,$$

where $\mathscr{X}$ is defined in (7.4),

$$\mathbf{W}_{\mathbf{H}}(\mathbf{x}_0) = \Omega_u^{-1/2} \mathbf{K}_{\mathbf{H}}(\mathbf{x}_0)\, \Omega_u^{-1/2}$$

in Henderson and Ullah (2005),

$$\mathbf{W}_{\mathbf{H}}(\mathbf{x}_0) = \Omega_u^{-1} \mathbf{K}_{\mathbf{H}}(\mathbf{x}_0) \quad \text{and} \quad \mathbf{W}_{\mathbf{H}}(\mathbf{x}_0) = \sqrt{\mathbf{K}_{\mathbf{H}}(\mathbf{x}_0)}\,\Omega_u^{-1}\sqrt{\mathbf{K}_{\mathbf{H}}(\mathbf{x}_0)}$$

in Lin and Carroll (2000). Lin and Carroll (2000) demonstrate that the weighted local linear estimator is less efficient than the pooled local linear estimator and suggest the "working independence" approach, which ignores the correlation structure within clusters.

(ii). Su and Ullah (2007) propose a two-step estimator via the *pre-whitening* method introduced in Ruckstuhl et al. (2000) for nonparametric two-dimensional panel data models with individual specific random effects, and show that the proposed estimator achieves asymptotic improvement over the pooled local linear estimator. In addition, Su et al. (2013) propose a two-step estimator improved upon Martins-Filho and Yao (2009). The four two-step estimators are shown to have the same asymptotic bias term, but smaller asymptotic variances up to a constant scale than the pooled local linear estimator. According to Su and Ullah (2007), the two-step estimators given in Su et al. (2013) and Martins-Filho and Yao (2009) may be less efficient than that given in Su and Ullah (2007), as the latter uses an optimal scale in transformed data. However, the asymptotic good performance of Su and Ullah's (2007) estimator with the optimal scale over the two-step estimator given in Su et al. (2013) may not be easily realized for data with moderate sample sizes as it requires a three-step estimation methodology with an estimated optimal scale that relies on estimated $m^{(2)}(\mathbf{x}_0)$.

(iii). Yao and Li (2013) apply Cholesky decomposition and profile least squares techniques to estimate the correlation structure of the error term and the unknown regression function simultaneously, assuming the error term is i.i.d. across clusters and correlated within each cluster. Yao and Li (2013) show that their proposed estimator is asymptotically more efficient than Lin and Carroll's (2000) weighted local linear estimators and is as asymptotically efficient as if the true covariance matrix were known. Note that all of the methods listed above require the error covariance matrix to be unknown up to a finite number of unknown parameters.

To the best of our knowledge, there are no research results that introduce modified local linear estimators improving upon the pooled local linear estimator for a general error covariance matrix with an infinite number of unknown parameters. Therefore, we will discuss how to construct a two-step estimator improving upon the pooled local linear estimator for the random-effects model when $\Omega_u = \Omega_u(\xi) = \mathrm{Var}(\mathbf{u}|\mathbf{X})$ is determined by a finite number of unknown parameters ξ. As Yao and Li's (2013) estimation method is not applicable if T is sufficiently large and the estimators given in category (ii) are more efficient than the pooled local linear estimator, below we will propose a two-step estimator built upon Su et al.'s (2013) estimation methodology. We will not follow Su and Ullah's (2007) estimation methodology as the incremental improvement of Su and Ullah's (2007) estimator over Su et al.'s (2013) estimator may not be realized in samples of a relatively moderate size as explained above. Without loss of generality, we assume that $\Omega_u = \mathrm{E}(\mathbf{u}\mathbf{u}'|\mathbf{X}) = \mathrm{E}(\mathbf{u}\mathbf{u}')$. This simplicity is aimed to shorten our equations below.

7.3.1 Weighted Local Linear Estimator

In this section, we will explain why the weighted kernel estimator with the weighting matrix $\mathbf{W}_\mathbf{H}(\mathbf{x}_0) = \Omega_u^{-1/2}\mathbf{K}_\mathbf{H}(\mathbf{x}_0)\,\Omega_u^{-1/2}$ is not preferable. We first transform the data matrix by multiplying $\Omega_u^{-1/2}$, and then apply the local linear regression to the

transformed data. The *weighted local linear estimator* of γ is given by

$$\check{\gamma} = \left[\mathscr{X}'\Omega_u^{-1/2}\mathbf{K_H}(\mathbf{x}_0)\Omega_u^{-1/2}\mathscr{X}\right]^{-1}\mathscr{X}'\Omega_u^{-1/2}\mathbf{K_H}(\mathbf{x}_0)\Omega_u^{-1/2}\mathbf{y} \tag{7.19}$$

so that

$$\begin{aligned}\check{\gamma}-\gamma &= \left[\mathscr{X}'\Omega_u^{-1/2}\mathbf{K_H}(\mathbf{x}_0)\Omega_u^{-1/2}\mathscr{X}\right]^{-1}\mathscr{X}'\Omega_u^{-1/2}\mathbf{K_H}(\mathbf{x}_0)\Omega_u^{-1/2}\left[\Pi\left(\bar{\mathbf{x}}\right)/2+\mathbf{u}\right]\\ &\equiv \mathbf{A}^{-1}\left(\mathbf{B}+\mathbf{C}\right),\end{aligned}$$

where we denote

$$\begin{aligned}\mathbf{A} &= \mathscr{X}'\Omega_u^{-1/2}\mathbf{K_H}\left(\mathbf{x}_0\right)\Omega_u^{-1/2}\mathscr{X},\\ \mathbf{B} &= \mathscr{X}'\Omega_u^{-1/2}\mathbf{K_H}(\mathbf{x}_0)\Omega_u^{-1/2}\Pi(\bar{\mathbf{x}}),\\ \mathbf{C} &= \mathscr{X}'\Omega_u^{-1/2}\mathbf{K_H}(\mathbf{x}_0)\Omega_u^{-1/2}\mathbf{u}.\end{aligned}$$

Denoting the typical element of $\Omega_u^{-1/2}$ by $v_{(ijt)_1(ijt)_2}$ and applying simple algebra, we obtain

$$\begin{aligned}\mathbf{A} = \sum_{(ijt)_1}\sum_{(ijt)_2}\sum_{(ijt)_3} K\left(\mathbf{H}^{-1}\left(\mathbf{x}_{(ijt)_1}-\mathbf{x}_0\right)\right)v_{(ijt)_2(ijt)_1}v_{(ijt)_3,(ijt)_1}\times\\ \times\begin{bmatrix}1 & \left(\mathbf{x}_{(ijt)_2}-\mathbf{x}_0\right)'\\ \mathbf{x}_{(ijt)_3}-\mathbf{x}_0 & \left(\mathbf{x}_{(ijt)_3}-\mathbf{x}_0\right)\left(\mathbf{x}_{(ijt)_2}-\mathbf{x}_0\right)'\end{bmatrix}\end{aligned}$$

and

$$\begin{aligned}\mathbf{B} = \sum_{(ijt)_1}\sum_{(ijt)_2}\sum_{(ijt)_3} K\left(\mathbf{H}^{-1}\left(\mathbf{x}_{(ijt)_1}-\mathbf{x}_0\right)\right)v_{(ijt)_2(ijt)_1}v_{(ijt)_3(ijt)_1}\times\\ \times\begin{bmatrix}1\\ \mathbf{x}_{(ijt)_3}-\mathbf{x}_0\end{bmatrix}\left(\mathbf{x}_{(ijt)_2}-\mathbf{x}_0\right)'m^{(2)}\left(\bar{\mathbf{x}}_{(ijt)_2}\right)\left(\mathbf{x}_{(ijt)_2}-\mathbf{x}_0\right)'/2.\end{aligned}$$

It is evident that this estimation method may be inconsistent due to the non-negligible bias term, $\mathbf{A}^{-1}\mathbf{B}$, that results from the non-zero off-diagonal elements in $\Omega_u^{-1/2}$ when $\left\{u_{ijt}\right\}$ exhibits weak dependence in cross sections and/or time periods. On the other hand, this estimation methodology may work for the local constant estimation method. However, compared with the local linear estimation method, the local constant kernel estimator is not adaptive to the distribution of $\left\{x_{ijt}\right\}$ and suffers a boundary problem. Therefore, we prefer the two-step estimation methods used in Su and Ullah (2007), Su et al. (2013) and Martins-Filho and Yao (2009), which move the off-diagonal terms of $\Omega_u^{-1/2}$ to the left-side of the regression model, so that the terms with summations $\sum_{(ijt)_1}\sum_{(ijt)_2\neq(ijt)_1}$, $\sum_{(ijt)_1}\sum_{(ijt)_3\neq(ijt)_1}$ and $\sum_{(ijt)_1}\sum_{(ijt)_2\neq(ijt)_1}\sum_{(ijt)_3\neq(ijt)_1}$ are removed from $\mathbf{A}$ and $\mathbf{B}$. Consequently, the non-

zero asymptotic center is removed from the estimator. We will give our proposed two-step estimator in the next section.

7.3.2 Two-step Estimator

Now, we explain our proposed two-step estimator. In the first step, we obtain the pooled local linear estimator $\hat{m}(\mathbf{X})$ defined in Sect. 7.2. The second step is described as follows: Let P be a square matrix of size $\mathbb{N}$ satisfying $\Omega_u = PP'$, where $p_{(ijt)_1(ijt)_2}$ and $v_{(ijt)_1(ijt)_2}$ are the typical elements of P and P^{-1}, respectively. Additionally, we denote the diagonal element of matrix P^{-1} by v_{ijt}. Then, we denote

$$\mathbf{Z} \equiv P^{-1}\mathbf{y} + \left(\mathscr{H}^{-1} - P^{-1}\right) m(\mathbf{X}) = \mathscr{H}^{-1} m(\mathbf{X}) + P^{-1}\mathbf{u}\,, \tag{7.20}$$

where $\mathscr{H}^{-1} = \text{diag}\left\{P^{-1}\right\}$ equals the diagonal element of matrix P^{-1}. Note that the off-diagonal effect of multiplying P^{-1} on $m(\mathbf{X})$, i.e., $\left(\mathscr{H}^{-1} - P^{-1}\right) m(\mathbf{X})$ is moved to the left-side of the equation and included in $\mathbf{Z}$, which follows Su and Ullah (2007), Su et al. (2013) and Martins-Filho and Yao (2009). The second-step estimator based on (7.20) is the minimizer of the following objective function

$$\tilde{\gamma} = \arg\min_{\gamma} \left(\hat{\mathbf{Z}} - \mathscr{H}^{-1}\mathscr{X}\gamma\right)' \mathbf{K}_{\mathbf{H}_0}(\mathbf{x}_0) \left(\hat{\mathbf{Z}} - \mathscr{H}^{-1}\mathscr{X}\gamma\right),$$

and can be written as

$$\tilde{\gamma} = \left[\mathscr{X}'\mathscr{H}^{-1}\mathbf{K}_{\mathbf{H}_0}(\mathbf{x}_0)\mathscr{H}^{-1}\mathscr{X}\right]^{-1} \mathscr{H}^{-1}\mathscr{X}'\mathbf{K}_{\mathbf{H}_0}(\mathbf{x}_0)\hat{\mathbf{Z}}, \tag{7.21}$$

where $\hat{\mathbf{Z}} = P^{-1}\mathbf{y} + \left(\mathscr{H}^{-1} - P^{-1}\right)\hat{m}(\mathbf{X})$, and $\hat{m}(\mathbf{X})$ is the pooled local linear estimate of $m(\mathbf{X})$, and the bandwidth matrix $\mathbf{H}_0 = \text{diag}\left\{h_{1,0}, ..., h_{k,0}\right\}$ used in the second step is different from the bandwidth matrix $\mathbf{H}$ used to obtain $\hat{m}(\mathbf{X})$ in the first step.

The weighted local linear estimator $\check{\gamma}$ defined in (7.19) and the two-step estimator $\tilde{\gamma}$ defined in (7.21) are closely related to each other. Both estimators transfer model (7.2) into another model with uncorrelated errors. However, the estimator defined in (7.19) with $\mathbf{W}_{\mathbf{H}}(\mathbf{x}_0) = \Omega_u^{-1/2}\mathbf{K}_{\mathbf{H}}(\mathbf{x}_0)\Omega_u^{-1/2}$ estimates γ from the following transformed model

$$P^{-1}\mathbf{y} = P^{-1}m(\mathbf{X}) + P^{-1}\mathbf{u}\,, \quad P^{-1} = \Omega_u^{-1/2}\,,$$

while the estimator defined in (7.21) estimates γ from a differently transformed model (7.20), where the off-diagonal part of P^{-1} times $m(\mathbf{X})$ is moved to the left-hand side of the regression model. Section 7.3.1 explains that the weighted local linear estimator defined in (7.19) may be inconsistent in the presence of weakly dependent errors, u_{ijt}. Below, we will derive the limiting results for $\tilde{\gamma}$ defined in (7.21).

First, since $\hat{\mathbf{Z}}$ can be written as

$$\hat{\mathbf{Z}} = \mathscr{H}^{-1} m(\mathbf{X}) + \left(\mathscr{H}^{-1} - P^{-1}\right)\left[\hat{m}(\mathbf{X}) - m(\mathbf{X})\right] + P^{-1}\mathbf{u}\,,$$

we have $\tilde{\boldsymbol{\gamma}} - \boldsymbol{\gamma} = \mathbf{A}^{-1}(\mathbf{B}_1 + \mathbf{B}_2 + \mathbf{C})$, where we denote $K_{0ijt} = K\left(\mathbf{H}_0^{-1}(\mathbf{x}_{ijt} - \mathbf{x}_0)\right)$,

$$\begin{aligned}
\mathbf{A} &= \mathscr{X}'\mathscr{H}^{-1}\mathbf{K}_{\mathbf{H}_0}(\mathbf{x}_0)\mathscr{H}^{-1}\mathscr{X} \\
&= \sum_{ijt} K_{0ijt} v_{ijt}^2 \begin{bmatrix} 1 & (\mathbf{x}_{ijt} - \mathbf{x}_0)' \\ \mathbf{x}_{ijt} - \mathbf{x}_0 & (\mathbf{x}_{ijt} - \mathbf{x}_0)(\mathbf{x}_{ijt} - \mathbf{x}_0)' \end{bmatrix}, \\
\mathbf{B}_1 &= \mathscr{H}^{-1}\mathscr{X}'\mathbf{K}_{\mathbf{H}_0}(\mathbf{x}_0)\mathscr{H}^{-1}\Pi(\bar{\mathbf{x}})/2 \\
&= 2^{-1}\sum_{ijt} K_{0ijt} v_{ijt}^2 (\mathbf{x}_{ijt} - \mathbf{x}_0)' m^{(2)}(\bar{\mathbf{x}}_{ijt})(\mathbf{x}_{ijt} - \mathbf{x}_0)', \\
\mathbf{B}_2 &= \mathscr{X}'\mathscr{H}^{-1}\mathbf{K}_{\mathbf{H}_0}(\mathbf{x}_0)\left(\mathscr{H}^{-1} - P^{-1}\right)\left[\hat{m}(\mathbf{X}) - m(\mathbf{X})\right] \\
&= -\sum_{i_1 j_1 t_1} K_{0 i_1 j_1 t_1} v_{(ijt)_1} \begin{bmatrix} 1 \\ \mathbf{x}_{ijt} - \mathbf{x}_0 \end{bmatrix}' \sum_{(ijt)_2 \neq (ijt)_1} v_{(ijt)_1 (ijt)_2} e_1' \mathscr{A}^{-1}\left(\mathbf{x}_{(ijt)_2}\right) \times \\
&\quad \times \left[\mathscr{B}\left(\mathbf{x}_{(ijt)_2}\right) + \mathscr{C}\left(\mathbf{x}_{(ijt)_2}\right)\right],
\end{aligned}$$

where $\mathscr{A}(\mathbf{x}) = \mathscr{X}'\mathbf{K}_{\mathbf{H}}(\mathbf{x})\mathscr{X}$, $\mathscr{B}(\mathbf{x}) = \mathscr{X}'\mathbf{K}_{\mathbf{H}}(\mathbf{x})\Pi(\bar{\mathbf{x}})/2$, and $\mathscr{C}(\mathbf{x}) = \mathscr{X}'\mathbf{K}_{\mathbf{H}}(\mathbf{x})\mathbf{u}$, and e_1 is the first column of the identify matrix $\mathbf{I}_{k+1}$, and

$$\mathbf{C} = \mathscr{X}'\mathscr{H}^{-1}\mathbf{K}_{\mathbf{H}_0}(\mathbf{x}_0)P^{-1}\mathbf{u}.$$

Denote $D_{n0} = \operatorname{diag}\{1, \mathbf{H}_0\}$. Again, we derive the limit results, assuming that $\|\mathbf{H}_0\| \to 0$, $\mathbb{N}_c|\mathbf{H}_0| \to \infty$, and $\lim_{\mathbb{N}_c\to\infty} \mathbb{N}_c|\mathbf{H}_0|\,\|\mathbf{H}_0\|^4 = M > 0$ as $\mathbb{N}_c \to \infty$, where $\mathbb{N}_c$, for $c = 1, 2, \ldots, 7$ is defined in (7.27). It is readily seen that

$$\begin{aligned}
&\mathrm{E}\left((\mathbb{N}_c|\mathbf{H}_0|)^{-1} D_{n0}^{-1}\mathbf{A}\right) \\
&= \frac{1}{\mathbb{N}_c|\mathbf{H}_0|}\sum_{ijt}\mathrm{E}\left\{K_{0ijt}v_{ijt}^2\begin{bmatrix} 1 & (\mathbf{x}_{ijt} - \mathbf{x}_0)'\mathbf{H}_0^{-1} \\ \mathbf{H}_0^{-1}(\mathbf{x}_{ijt} - \mathbf{x}_0) & \mathbf{H}_0^{-1}(\mathbf{x}_{ijt} - \mathbf{x}_0)(\mathbf{x}_{ijt} - \mathbf{x}_0)'\mathbf{H}_0^{-1}\end{bmatrix}\right\} \\
&= \mathbb{N}_c^{-1}\sum_{ijt} v_{ijt}^2 f_{ijt}(\mathbf{x}_0)\begin{bmatrix} 1 & \mathbf{0}_k' \\ \mathbf{0}_k & \kappa_{12}\mathbf{I}_k\end{bmatrix}(1 + o(1)),
\end{aligned}$$

and

$$\begin{aligned}
&\mathrm{Var}\left(\frac{D_{n0}^{-1}\mathbf{A}}{\mathbb{N}_c|\mathbf{H}_0|}\right) = \frac{1}{(\mathbb{N}_c|\mathbf{H}_0|)^2} \times \\
&\times \mathrm{Var}\left\{\sum_{ijt} K_{0ijt}v_{ijt}^2\begin{bmatrix} 1 & (\mathbf{x}_{ijt} - \mathbf{x}_0)'\mathbf{H}_0^{-1} \\ \mathbf{H}_0^{-1}(\mathbf{x}_{ijt} - \mathbf{x}_0) & \mathbf{H}_0^{-1}(\mathbf{x}_{ijt} - \mathbf{x}_0)(\mathbf{x}_{ijt} - \mathbf{x}_0)'\mathbf{H}_0^{-1}\end{bmatrix}\right\}.
\end{aligned}$$

To save space, we illustrate the calculation of $\mathrm{Var}\left((\mathbb{N}_c|\mathbf{H}_0|)^{-1} D_{n0}^{-1}\mathbf{A}\right)$ using its $(1,1)$th element. That is, we have

$$\operatorname{Var}\left\{\sum_{ijt} K_{0ijt} v_{ijt}^2\right\} = \sum_{ijt} \operatorname{Var}\left(K_{0ijt} v_{ijt}^2\right) + R_n \,,$$

where

$$\begin{aligned}
R_n = & \sum_{ijt}\sum_{t'\neq t} \operatorname{Cov}\left(K_{0ijt} v_{ijt}^2, K_{0ijt'} v_{ijt'}^2\right) \\
& + \sum_{ijt}\sum_{j'\neq j} \operatorname{Cov}\left(K_{0ijt} v_{ijt}^2, K_{0ij't} v_{ij't}^2\right) + \sum_{ijt}\sum_{i'\neq i} \operatorname{Cov}\left(K_{0ijt} v_{ijt}^2, K_{0i'jt} v_{i'jt}^2\right) \\
& + \sum_{ijt}\sum_{j'\neq j}\sum_{t'\neq t} \operatorname{Cov}\left(K_{0ijt} v_{ijt}^2, K_{0ij't'} v_{ij't'}^2\right) + \sum_{ijt}\sum_{i'\neq i}\sum_{t'\neq t} \operatorname{Cov}\left(K_{0ijt} v_{ijt}^2, K_{0i'jt'} v_{i'jt'}^2\right) \\
& + \sum_{ijt}\sum_{i'\neq i}\sum_{j'\neq j} \operatorname{Cov}\left(K_{0ijt} v_{ijt}^2, K_{0i'j't} v_{i'j't}^2\right) + \sum_{ijt}\sum_{i'j't'\neq ijt} \operatorname{Cov}\left(K_{0ijt} v_{ijt}^2, K_{0i'j't'} v_{i'j't'}^2\right)
\end{aligned}$$

i.e., R_n contains all the covariance terms. It is straightforward to obtain

$$\sum_{ijt} \operatorname{Var}\left(K_{0ijt} v_{ijt}^2\right) \approx \nu_0 |\mathbf{H}_0| \sum_{ijt} v_{ijt}^4 f_{ijt}(\mathbf{x}_0)$$

and

$$\begin{aligned}
|\mathbf{H}_0|^{-2} R_n \approx & \sum_{ijt}\sum_{t'\neq t} v_{ijt}^2 v_{ijt'}^2 \left[f_{(ijt)(ijt')}(\mathbf{x}_0,\mathbf{x}_0) - f_{ijt}(\mathbf{x}_0) f_{ijt'}(\mathbf{x}_0)\right] \\
& + \sum_{ijt}\sum_{j'\neq j} v_{ijt}^2 v_{ij't}^2 \left[f_{(ijt)(ij't)}(\mathbf{x}_0,\mathbf{x}_0) - f_{ijt}(\mathbf{x}_0) f_{ij't}(\mathbf{x}_0)\right] \\
& + \sum_{ijt}\sum_{i'\neq i} v_{ijt}^2 v_{i'jt}^2 \left[f_{(ijt)(ijt')}(\mathbf{x}_0,\mathbf{x}_0) - f_{ijt}(\mathbf{x}_0) f_{ijt'}(\mathbf{x}_0)\right] \\
& + \sum_{ijt}\sum_{j'\neq j}\sum_{t'\neq t} v_{ijt}^2 v_{ij't'}^2 \left[f_{(ijt)(ij't')}(\mathbf{x}_0,\mathbf{x}_0) - f_{ijt}(\mathbf{x}_0) f_{ij't'}(\mathbf{x}_0)\right] \\
& + \sum_{ijt}\sum_{i'\neq i}\sum_{t'\neq t} v_{ijt}^2 v_{i'jt'}^2 \left[f_{(ijt)(i'jt')}(\mathbf{x}_0,\mathbf{x}_0) - f_{ijt}(\mathbf{x}_0) f_{i'jt'}(\mathbf{x}_0)\right] \\
& + \sum_{ijt}\sum_{i'\neq i}\sum_{j'\neq j} v_{ijt}^2 v_{i'j't}^2 \left[f_{(ijt)(i'j't)}(\mathbf{x}_0,\mathbf{x}_0) - f_{ijt}(\mathbf{x}_0) f_{i'j't}(\mathbf{x}_0)\right] \\
& + \sum_{ijt}\sum_{i'j't'\neq ijt} v_{ijt}^2 v_{i'j't'}^2 \left[f_{(ijt)(ijt)'}(\mathbf{x}_0,\mathbf{x}_0) - f_{ijt}(\mathbf{x}_0) f_{i'j't'}(\mathbf{x}_0)\right].
\end{aligned}$$

Under the same weak dependent condition imposed on $\{\mathbf{x}_{ijt}\}$ as in Sect. 7.2 and assuming $\max_{i,j,t} v_{ijt}^2 \leq M < \infty$, we show that $|\mathbf{H}_0|^{-2} R_n$ is asymptotically negligible relative to $\sum_{ijt} \operatorname{Var}\left(K_{0ijt} v_{ijt}^2\right)$. Therefore, we have

$$
\begin{aligned}
&\operatorname{Var}\left(\left(\mathbb{N}_c\left|\mathbf{H}_0\right|\right)^{-1} D_{n0}^{-1}\mathbf{A}\right) \\
&\approx \frac{1}{\mathbb{N}_c^2\left|\mathbf{H}_0\right|}\sum_{ijt} v_{(i,j,t),(i,j,t)}^4 f_{i,j,t}\left(\mathbf{x}_0\right)\begin{bmatrix} \nu_0 & \mathbf{0}_k' \\ \mathbf{0}_k & \int K^2(u)\,u'\mathbf{H}_0^2 u\mathbf{H}_0\mathbf{u}\mathbf{u}'\mathbf{H}du \end{bmatrix} \\
&= O\left(\frac{1}{\mathbb{N}_c\left|\mathbf{H}_0\right|}\right).
\end{aligned}
$$

Taking together the results above, we have

$$
\frac{D_{n0}^{-1}\mathbf{A}}{\mathbb{N}_c\left|\mathbf{H}_0\right|} = \mathbb{N}_c^{-1}\sum_{ijt} v_{ijt}^2 f_{ijt}\left(\mathbf{x}_0\right)\begin{bmatrix} 1 & \mathbf{0}_k' \\ \mathbf{0}_k' & \kappa_{12}\mathbf{I}_k \end{bmatrix} + o_p(1).
$$

Similarly, we obtain

$$
\frac{D_{n0}^{-1}\mathbf{B}_1}{\mathbb{N}_c\left|\mathbf{H}_0\right|} = \frac{\kappa_{12}}{2\mathbb{N}_c}\sum_{ijt} v_{ijt}^2 f_{ijt}(\mathbf{x}_0)\begin{bmatrix} \operatorname{tr}\left\{\mathbf{H}_0 m^{(2)}(\mathbf{x}_0)\mathbf{H}_0\right\} \\ \mathbf{0}_k \end{bmatrix}(1+o_p(1))
$$

and

$$
\begin{aligned}
&\operatorname{Var}\left(\left(\mathbb{N}_c\left|\mathbf{H}_0\right|\right)^{-1/2} D_{n0}^{-1}\mathscr{C}\right) \\
&= \frac{1}{\mathbb{N}_c\left|\mathbf{H}_0\right|}\mathrm{E}\left[D_n^{-1}\mathscr{X}'\mathscr{H}^{-1}\mathbf{K}_{\mathbf{H}}^2(\mathbf{x}_0)\mathscr{H}^{-1}\mathscr{X}D_n^{-1}\right] \\
&= \frac{1}{\mathbb{N}_c}\sum_{ijt} v_{ijt}^2 f_{ijt}\left(\mathbf{x}_0\right)\begin{bmatrix} \nu_0 & \mathbf{0}_k' \\ \mathbf{0}_k' & \kappa_{22}\mathbf{I}_k \end{bmatrix}(1+o(1)).
\end{aligned}
$$

Assuming that the random variable, $\mathbf{x}_{ijt}$, takes value from a compact subset, $S_x \subset R^k$, and the kernel function satisfies $|k(u)-k(v)| \le M|u-v|$ for any $u, v \in R$, closely following the proof given in Masry (1996), we have

$$
\begin{aligned}
&\sup_{\mathbf{x}_0\in S_x}\left|\frac{D_n^{-1}\mathscr{A}\left(\mathbf{x}_0\right)}{\mathbb{N}_c\left|\mathbf{H}\right|} - \bar{f}_c\left(\mathbf{x}_0\right)\begin{bmatrix} 1 & \mathbf{0}_k' \\ \mathbf{0}_k & \kappa_{12}\mathbf{I}_k \end{bmatrix}\right| = O_p\left(\|\mathbf{H}\|^2 + \sqrt{\ln\mathbb{N}_c/\left(\mathbb{N}_c\left|\mathbf{H}\right|\right)}\right), \\
&\sup_{\mathbf{x}_0\in S_x}\left|\frac{D_n^{-1}\mathscr{B}\left(\mathbf{x}_0\right)}{2\mathbb{N}_c\left|\mathbf{H}\right|} - \kappa_{12}\begin{bmatrix} \operatorname{tr}\left\{\mathbf{H}\mathbf{m}^{(2)}\left(\mathbf{x}_0\right)\mathbf{H}\right\}\bar{f}\left(\mathbf{x}_0\right)/2 \\ \mathbf{0}_k \end{bmatrix}\right| \\
&= O_p\left(\|\mathbf{H}\|^4 + \|\mathbf{H}\|^2\sqrt{\ln\mathbb{N}_c/\left(\mathbb{N}_c\left|\mathbf{H}\right|\right)}\right),
\end{aligned}
$$

where we denote $\bar{f}_c(\mathbf{x}_0) = \lim_{\mathbb{N}_c\to\infty}\mathbb{N}_c^{-1}\sum_{ijt} v_{ijt}^2 f_{ijt}(\mathbf{x}_0)$. It then follows that if the bandwidth matrices in two steps satisfy that

$$
\|\mathbf{H}\| = o\left(\|\mathbf{H}_0\|\right), \quad \|\mathbf{H}\|^4 / \left(\mathbb{N}_c\left|\mathbf{H}_0\right|\|\mathbf{H}_0\|^4\right) = o(1), \quad \mathbb{N}_c\left|\mathbf{H}\right|\|\mathbf{H}_0\|^2 \to \infty,
$$

and $\mathbb{N}_c\|\mathbf{H}_0\|^4 \to \infty$, as $\mathbb{N}_c \to \infty$, we have

$$\frac{D_{n0}^{-1}\mathbf{B}_2}{\mathbb{N}_c|\mathbf{H}|} \approx -\Delta_1 - \Delta_2 = O_p\left(\|\mathbf{H}\|^2\right) + O_p\left(\frac{\|\mathbf{H}\|^2}{\sqrt{\mathbb{N}_c|\mathbf{H}|}}\right) + O_p\left(\frac{1}{\mathbb{N}_c|\mathbf{H}|} + \frac{1}{\sqrt{\mathbb{N}_c}}\right)$$
$$= o_p\left(\|\mathbf{H}_0\|^2\right),$$

where

$$\begin{aligned}
\Delta_1 &= \frac{\kappa_{12}}{2\mathbb{N}_c|\mathbf{H}_0|}\sum_{i_1j_1t_1}\sum_{(ijt)_2\neq(ijt)_1} v_{(ijt)_1}v_{(ijt)_1(ijt)_2}\mathrm{tr}\left\{\mathbf{H}\mathbf{m}^{(2)}\left(\mathbf{x}_{(ijt)_2}\right)\mathbf{H}\right\} \\
&\quad\times K_{0i_1j_1t_1}\left[1\ \left(\mathbf{x}_{(ijt)_1}-\mathbf{x}_0\right)'\mathbf{H}_0^{-1}\right] \\
&= \frac{\kappa_{12}}{2\mathbb{N}_c}\sum_{i_1j_1t_1}\sum_{(ijt)_2\neq(ijt)_1} v_{(ijt)_1}v_{(ijt)_1(ijt)_2}f_{(ijt)_1}(\mathbf{x}_0) \\
&\quad\times\mathrm{E}\left[\mathrm{tr}\left\{\mathbf{H}\mathbf{m}^{(2)}\left(\mathbf{x}_{(ijt)_2}\right)\mathbf{H}\right\}|\mathbf{x}_{(ijt)_1}=\mathbf{x}_0\right]\left[1,\mathbf{0}_k'\right] \\
&\quad+\left[O_p\left(\|\mathbf{H}\|^2\|\mathbf{H}_0\|^2\right), O_p\left(\|\mathbf{H}\|^2\|\mathbf{H}_0\|^4\right)\right] + O_p\left(\frac{\|\mathbf{H}\|^2}{\sqrt{\mathbb{N}_c|\mathbf{H}_0|}}\right) \\
&= O_p\left(\|\mathbf{H}\|^2\right) + O_p\left(\frac{\|\mathbf{H}\|^2}{\sqrt{\mathbb{N}_c|\mathbf{H}_0|}}\right)
\end{aligned}$$

if $\mathbb{N}_c^{-1}\sum_{i_1j_1t_1}\sum_{(ijt)_2\neq(ijt)_1}\left|v_{(ijt)_1(ijt)_2}\right| \leq M$, and

$$\begin{aligned}
\Delta_2 &= \frac{1}{\mathbb{N}_c^2|\mathbf{H}_0||\mathbf{H}|}\sum_{i_1j_1t_1}\sum_{(ijt)_2\neq(ijt)_1} v_{(ijt)_1}v_{(ijt)_1(ijt)_2}\bar{f}_c^{-1}\left(\mathbf{x}_{(ijt)_2}\right)K_{0i_1j_1t_1} \\
&\quad\times\left[1\ \left(\mathbf{x}_{(ijt)_1}-\mathbf{x}_0\right)'\mathbf{H}_0^{-1}\right]e_1'D_n^{-1}\mathscr{X}'\mathbf{K}_{\mathbf{H}}(\mathbf{x}_{(ijt)_2})\mathbf{u} \\
&= O_p\left(\frac{1}{\mathbb{N}_c|\mathbf{H}|}+\frac{1}{\sqrt{\mathbb{N}_c}}\right),
\end{aligned}$$

as we have $\mathrm{E}(\Delta_2) = \mathbf{0}_k'$ and

$$
\begin{aligned}
&\operatorname{Var}(\Delta_2)\\
&= \frac{1}{\mathbb{N}_c^4 |\mathbf{H}_0|^2 |\mathbf{H}|^2} \sum_{i_1 j_1 t_1} \sum_{(ijt)_2 \neq (ijt)_1} \sum_{(ijt)_3} \sum_{(ijt)_4 \neq (ijt)_3} v_{(ijt)_1} v_{(ijt)_1 (ijt)_2} v_{(ijt)_3} v_{(ijt)_3 (ijt)_4}\\
&\quad \times \mathrm{E}\left\{ \bar{f}_c^{-1}\left(\mathbf{x}_{(ijt)_2}\right) \bar{f}_c^{-1}\left(\mathbf{x}_{(ijt)_4}\right) K_{0 i_1 j_1 t_1} K_{0 i_3 j_3 t_3} \right.\\
&\quad \times e_1' D_n^{-1} \mathscr{X}' \mathbf{K}_{\mathbf{H}}\left(\mathbf{x}_{(ijt)_1}\right) \Omega_u \mathbf{K}_{\mathbf{H}}\left(\mathbf{x}_{(ijt)_3}\right) \mathscr{X} D_n^{-1} e_1\\
&\quad \left. \times \begin{bmatrix} 1 & \left(\mathbf{x}_{(ijt)_1} - \mathbf{x}_0\right)' \mathbf{H}_0^{-1} \\ \mathbf{H}_0^{-1}\left(\mathbf{x}_{(ijt)_3} - \mathbf{x}_0\right) & \mathbf{H}_0^{-1}\left(\mathbf{x}_{(ijt)_3} - \mathbf{x}_0\right)\left(\mathbf{x}_{(ijt)_1} - \mathbf{x}_0\right)' \mathbf{H}_0^{-1} \end{bmatrix} \right\}\\
&\approx \frac{1}{\mathbb{N}_c^4 |\mathbf{H}_0|^2 |\mathbf{H}|^2} \sum_{i_1 j_1 t_1} \sum_{(ijt)_2 \neq (ijt)_1} \sum_{(ijt)_3 \neq (ijt)_1} \sum_{(ijt)_4 \neq (ijt)_3} \sum_{i_1 j_1 t_1} \sum_{(ijt)_3'} v_{(ijt)_1} v_{(ijt)_1 (ijt)_2} v_{(ijt)_3}\\
&\quad \times v_{(ijt)_3 (ijt)_4} \sigma_{(ijt)_1' (ijt)_3'} \mathrm{E}\left\{ \bar{f}^{-1}\left(\mathbf{x}_{(ijt)_2}\right) \bar{f}^{-1}\left(\mathbf{x}_{(ijt)_4}\right) K_{0 i_1 j_1 t_1} K_{0 i_3 j_3 t_3} \right.\\
&\quad \times K\left(\mathbf{H}^{-1}\left(\mathbf{x}_{(ijt)_1'} - \mathbf{x}_{(ijt)_1}\right)\right) K\left(\mathbf{H}^{-1}\left(\mathbf{x}_{(ijt)_3'} - \mathbf{x}_{(ijt)_3}\right)\right)\\
&\quad \left. \begin{bmatrix} 1 & \left(\mathbf{x}_{(ijt)_1} - \mathbf{x}_0\right)' \mathbf{H}_0^{-1} \\ \mathbf{H}_0^{-1}\left(\mathbf{x}_{(ijt)_3} - \mathbf{x}_0\right) & \mathbf{H}_0^{-1}\left(\mathbf{x}_{(ijt)_3} - \mathbf{x}_0\right)\left(\mathbf{x}_{(ijt)_1} - \mathbf{x}_0\right)' \mathbf{H}_0^{-1} \end{bmatrix} \right\}\\
&= O\left(\frac{1}{\mathbb{N}_c^2 |\mathbf{H}|^2} + \frac{1}{\mathbb{N}_c}\right).
\end{aligned}
$$

In addition, we obtain

$$
\begin{aligned}
&\operatorname{Var}\left(\left(\mathbb{N}_c |\mathbf{H}_0|\right)^{-1/2} D_{n0}^{-1} \mathbf{C}\right)\\
&= \frac{1}{\mathbb{N}_c |\mathbf{H}_0|} D_{n0}^{-1} \mathrm{E}\left[\mathscr{X}' \mathscr{H}^{-1} \mathbf{K}_{\mathbf{H}_0}^2(\mathbf{x}_0) \mathscr{H}^{-1} \mathscr{X}\right] D_{n0}^{-1}\\
&= \frac{1}{\mathbb{N}_c |\mathbf{H}_0|} \sum_{ijt} \mathrm{E}\left\{ v_{ijt}^2 K_{0ijt}^2 \begin{bmatrix} 1 & (\mathbf{x}_{ijt} - \mathbf{x}_0)' \mathbf{H}_0^{-1} \\ \mathbf{H}_0^{-1}(\mathbf{x}_{ijt} - \mathbf{x}_0) & \mathbf{H}_0^{-1}(\mathbf{x}_{ijt} - \mathbf{x}_0)(\mathbf{x}_{ijt} - \mathbf{x}_0)' \mathbf{H}_0^{-1} \end{bmatrix} \right\}\\
&\approx \frac{1}{\mathbb{N}_c} \sum_{ijt} v_{ijt}^2 f_{ijt}(\mathbf{x}_0) \begin{bmatrix} \nu_0 & \mathbf{0}_k' \\ \mathbf{0}_k & \kappa_{22} \mathbf{I}_k \end{bmatrix}.
\end{aligned}
$$

As $v = P^{-1}\mathbf{u}$ has a zero mean and variance of identity matrix, applying the martingale central limit theorem gives

$$
\left(\mathbb{N}_c |\mathbf{H}_0|\right)^{-1/2} D_{n0}^{-1} \mathbf{C} \xrightarrow{d} \mathscr{N}\left(\mathbf{0}_{k+1}, \bar{f}_c(\mathbf{x}_0) \begin{bmatrix} \nu_0 & \mathbf{0}_k' \\ \mathbf{0}_k & \kappa_{22} \mathbf{I}_k \end{bmatrix}\right),
$$

if $\max_{ijt} \mathrm{E}\left(\left\| \upsilon_{ijt} \right\|^{2+\delta}\right) < M$ where υ_{ijt} is the (i,j,t)th element of the $(\mathbb{N} \times 1)$ vector υ. Taking all the results above gives

$$\sqrt{\mathbb{N}_c\left|\mathbf{H}_0\right|}D_{n0}\left(\tilde{\gamma}-\gamma-\frac{\kappa_{12}}{2}\begin{bmatrix}\mathrm{tr}\left\{\mathbf{H}_0 m^{(2)}(\mathbf{x}_0)\mathbf{H}_0\right\}\\ \mathbf{0}_k\end{bmatrix}\right)$$
$$\xrightarrow{d}\mathcal{N}\left(\mathbf{0}_{k+1},\bar{f}_c^{-1}(\mathbf{x}_0)\begin{bmatrix}\nu_0 & \mathbf{0}_k'\\ \mathbf{0}_k & \kappa_{22}\kappa_{12}^{-2}\mathbf{I}_k\end{bmatrix}\right) \tag{7.22}$$

so that

$$\sqrt{\mathbb{N}_c\left|\mathbf{H}_0\right|}\left(\tilde{m}(\mathbf{x}_0)-m(\mathbf{x}_0)-\frac{\kappa_{12}}{2}\mathrm{tr}\left\{\mathbf{H}_0 m^{(2)}(\mathbf{x}_0)\mathbf{H}_0\right\}\right)$$
$$\xrightarrow{d}\mathcal{N}\left(\mathbf{0}_{k+1},\nu_0\bar{f}_c^{-1}(\mathbf{x}_0)\right). \tag{7.23}$$

From (7.18) and (7.23), we observe that $\hat{m}(\mathbf{x}_0)$ and $\tilde{m}(\mathbf{x}_0)$ have the same asymptotic bias term, but $\tilde{m}(\mathbf{x}_0)$ has a smaller asymptotic variance term if both steps use the same bandwidth, i.e., $\mathbf{H}=\mathbf{H}_0$, because applying Cauchy-Schwarz inequality gives

$$\frac{1}{\mathbb{N}_c^{-1}\sum_{ijt}v_{ijt}^2 f_{ijt}(\mathbf{x}_0)}\leq\frac{\mathbb{N}_c^{-1}\sum_{ijt}v_{ijt}^{-2}f_{ijt}(\mathbf{x}_0)}{\left[\mathbb{N}_c^{-1}\sum_{ijt}f_{ijt}(\mathbf{x}_0)\right]^2}$$

and $v_{ijt}^{-2}\leq\sigma_{ijt}^2$ for all i, j, and t.

The current estimator is not feasible as the true covariance matrix Ω_u is unknown. Assume that there exists $\hat{\Omega}_u\equiv\Omega_u(\hat{\xi})$, an estimate of $\Omega_u=\Omega_u(\xi)$, with the finite dimensional parameters ξ consistently estimated by $\hat{\xi}$ satisfying $\left\|\hat{\xi}-\xi\right\|=o_p(1)$. Let $\mathbf{m}(\mathbf{x}_0)$ equal $\tilde{m}(\mathbf{x}_0)$ with Ω_u replaced with $\hat{\Omega}_u$. We can show that the limit distribution result given in (7.23) also holds for $\mathbf{m}(\mathbf{x}_0)$. For the error structure given in (7.24), the two-step estimator applies to cases (iv)-(vii) defined in Sect. 7.4 below.

7.4 Pairwise Random Effects

Sections 7.2 and 7.3 prove the consistency of the pooled local linear estimator and the two-step estimator without specifying the error structure of u_{ijt} other than condition (7.15). In this section, we will further explore the asymptotic properties of the pooled local linear estimator under a specific error structure considered in Chap. 2 in this book, where the authors consider a pairwise random-effects error structure as follows

$$u_{ijt}=\mu_{ij}+v_{it}+\zeta_{jt}+\varepsilon_{ijt}. \tag{7.24}$$

In (7.24), μ_{ij}, v_{it}, and ζ_{jt} represent three unobserved pairwise random effects, ε_{ijt} is an i.i.d. error, and $\{\mu_{ij}\}$, $\{v_{it}\}$, $\{\zeta_{jt}\}$ and $\{u_{ijt}\}$ are all uncorrelated with $\{\mathbf{x}_{ijt}\}$. In addition, (i) μ_{ij}, v_{it}, ζ_{jt}, and ε_{ijt} all have zero mean and are mutually uncorrelated with each other; (ii) $\varepsilon_{ijt}\sim i.i.d.\left(0,\sigma_\varepsilon^2\right)$; (iii) $\mathrm{E}\left(\mu_{ij}\mu_{i'j'}\right)=\sigma_\mu^2$ if $(i,j)=(i',j')$ and 0 otherwise; (iv) $\mathrm{E}\left(v_{it}v_{i't'}\right)=\sigma_v^2$ if $(i,t)=(i',t')$ and 0 otherwise; (v) $\mathrm{E}\left(\zeta_{jt}\zeta_{j't'}\right)=\sigma_\zeta^2$ if $(j,t)=(j',t')$ and 0 otherwise. Although this particular type of error structure

nests several popularly used error structures in empirical applications, conditions (i)-(v) imposed here can be restrictive in empirical works.

The pairwise random-effects error structure defined above implies

$$\Omega_u = \sigma_\mu^2(\mathbf{I}_{N_1N_2} \otimes \mathbf{J}_T) + \sigma_\nu^2(\mathbf{I}_{N_1} \otimes \mathbf{J}_{N_2} \otimes \mathbf{I}_T) + \sigma_\zeta^2(\mathbf{J}_{N_1} \otimes \mathbf{I}_{N_2T}) + \sigma_\varepsilon^2\mathbf{I}_{\mathbb{N}} \ , \quad (7.25)$$

where $\mathbf{J}_n = \iota_n\iota_n'$ denotes an $(n \times n)$ matrix of ones and "$\otimes$" is the Kronecker product operator. Furthermore, we have $\mathrm{E}(u_{ijt}|\mathbf{X}) = \sigma_\mu^2 + \sigma_\nu^2 + \sigma_\zeta^2 + \sigma_\varepsilon^2$ for all (i,j,t) and

$$\mathbb{N}^{-1} \sum_{i_1 j_1 t_1} \sum_{(ijt)_2 \neq (ijt)_1} \left|\mathrm{E}\left(u_{i_1 j_1 t_1} u_{i_2 j_2 t_2}\right)\right| = \sigma_\mu^2 (T-1) + \sigma_\nu^2 (N_2-1) + \sigma_\zeta^2 (N_1-1) . \quad (7.26)$$

Such a pairwise random-effects error structure may lead to the violation of the key assumption

$$\mathbb{N}^{-1} |\mathbf{H}| \sum_{i_1 j_1 t_1} \sum_{(ijt)_2 \neq (ijt)_1} \left|\sigma_{(ijt)_1 (ijt)_2}(\mathbf{x}_0, \mathbf{x}_0)\right| = o(1) ,$$

which is used to derive the asymptotic properties for the pooled local linear estimator in Sect. 7.2 and the two-step estimator in Sect. 7.3. We show that the pooled local linear estimator can be inconsistent under the error structure (7.26).

Given the specific error structure, we are able to explicitly discuss and derive the consistency results of our proposed local linear estimator for seven cases listed below in the following three subsections.

$$\begin{array}{l} \text{(i) } N_1 \to \infty, \text{ and } N_2 \text{ and } T \text{ are fixed, } \mathbb{N}_1 = N_1; \\ \text{(ii) } N_2 \to \infty, \text{ and } N_1 \text{ and } T \text{ are fixed, } \mathbb{N}_2 = N_2; \\ \text{(iii) } T \to \infty, \text{ and } N_1 \text{ and } N_2 \text{ are fixed, } \mathbb{N}_3 = T; \\ \text{(iv) } N_1 \to \infty, N_2 \to \infty, \text{ and } T \text{ is fixed, } \mathbb{N}_4 = N_1N_2; \\ \text{(v) } N_1 \to \infty \text{ and } T \to \infty, \text{ and } N_2 \text{ is fixed, } \mathbb{N}_5 = N_1T; \\ \text{(vi) } N_2 \to \infty \text{ and } T \to \infty, \text{ and } N_1 \text{ is fixed, } \mathbb{N}_6 = N_2T; \\ \text{(vii) } N_1 \to \infty, N_2 \to \infty, \text{ and } T \to \infty, \mathbb{N}_7 = \mathbb{N} = N_1N_2T. \end{array} \quad (7.27)$$

In Sect. 7.4.1, we show that the proposed local linear estimator is inconsistent for cases (i)-(iii) since the variance term of $\hat{\gamma}$ is asymptotically non-negligible. It is easy to show that the assumption in (7.15) fails to hold for these three cases. For cases (iv)-(vii), Sects. 7.4.2 and 7.4.3 show that the pooled local linear estimator is consistent with an asymptotic bias of order $O_p(\|\mathbf{H}\|^2)$; however, the asymptotic variance of the pooled estimator may be different from $O_p((\mathbb{N}_c|\mathbf{H}|)^{-1})$ as derived in Sect. 7.2 where $\mathbb{N}_c$, for $c = 1,2,\ldots,7$, is defined in (7.27), because condition (7.15) may fail under the pairwise random-effects structure. For example, we find that, under case (iv) with $N_1 \geq N_2$, the optimal MSE rate of the local linear estimator and its corresponding optimal bandwidth rate depend on the number of regressors k and the limit of a sample index ratio $N_1/N_2^{k/4}$; under case (vii) with $N_1 = \max(N_1, N_2, T)$, the optimal rate depends on k and $N_1/(N_2T)^{k/4}$. The results for other cases can be derived similarly. Intuitively, the optimal bandwidth will balance the trade-off ef-

fects of the bandwidth on the magnitude of the asymptotic squared bias and the asymptotic variance of the estimator. For three-dimensional nonparametric panel data models, the optimal bandwidth will balance the asymptotic squared bias of order $O_p(\|\mathbf{H}\|^4)$ and the asymptotic variance of different orders under different conditions for $\mathbb{N}^{-1}|\mathbf{H}|\sum_{i_1 j_1 t_1}\sum_{(ijt)_2\neq(ijt)_1}|\sigma_{(ijt)_1(ijt)_2}(\mathbf{x}_0,\mathbf{x}_0)|$. For nonparametric cross-sectional, time series and traditional two-dimensional panel data models, the covariances of the error terms are asymptotically negligible if they are weakly dependent across either cross sections or time series (see, e.g., Cai et al., 2000; Robinson, 2012), and the negligibility of the covariance terms has nothing to do with the bandwidth. However, for error structure (7.24), the non-explosiveness of the covariance terms relies on the error structure, the bandwidth, and the sum of absolute values of covariances for the error term, as indicated in (7.15). For convenience, we include the bandwidth requirement in Table 7.1 for cases (iv)-(vii) here, and we will then discuss and derive the optimal MSE rate of the local linear estimator and its corresponding optimal bandwidth rate for each case. Note that the following three subsections focus on the limit results of the variance term of the pooled local linear estimator.

Table 7.1 Bandwidth requirement

Case	Bandwidth conditions
(iv)	$\|\mathbf{H}\|\to 0$ and $N_1N_2|\mathbf{H}|\to\infty$ as $\min(N_1,N_2)\to\infty$
(v)	$\|\mathbf{H}\|\to 0$ and $N_1T|\mathbf{H}|\to\infty$ as $\min(N_1,T)\to\infty$
(vi)	$\|\mathbf{H}\|\to 0$ and $N_2T|\mathbf{H}|\to\infty$ as $\min(N_2,T)\to\infty$
(vii)	$\|\mathbf{H}\|\to 0$ and $N_1N_2T|\mathbf{H}|\to\infty$ as $\min(N_1,N_2,T)\to\infty$

7.4.1 Cases (i)–(iii): The Sample Size Increases in One Index Only

In cases (i)–(iii), only one of the three sample indices approaches to infinity. Without loss of generality, we consider case (i) under which $\mathbb{N}_1 = N_1 \to \infty$ and N_2, T fixed. The results for cases (ii) and (iii) can be derived in the same way. Note that

$$\begin{aligned} &N_1^{-1}|\mathbf{H}|\sum_{i_1 j_1 t_1}\sum_{(ijt)_2\neq(ijt)_1}\left|\mathrm{E}\left(u_{i_1 j_1 t_1}u_{i_2 j_2 t_2}\right)\right| \\ &=|\mathbf{H}|\left[\sigma_\mu^2 N_2 T(T-1)+\sigma_v^2 N_2(N_2-1)T+\sigma_\zeta^2(N_1-1)N_2T\right]=O\left(N_1|\mathbf{H}|\right), \end{aligned}$$

which does not approach to 0 as $N_1\to\infty$ and therefore violates the assumption in (7.15).

Applying (7.25) to (7.10), (7.11), and (7.12) gives

$$
\begin{aligned}
&(N_1|\mathbf{H}|)^{-2}\mathrm{E}\left[\iota_{\mathbb{N}}'\mathbf{K}_{\mathbf{H}}(\mathbf{x}_0)\Omega_u\mathbf{K}_{\mathbf{H}}(\mathbf{x}_0)\iota_{\mathbb{N}}\right]\\
&\approx \left(\sigma_\mu^2+\sigma_v^2+\sigma_\zeta^2+\sigma_\varepsilon^2\right)\frac{\nu_0}{N_1^2|\mathbf{H}|}\sum_{ijt}f_{ijt}(\mathbf{x}_0)+\frac{\sigma_\mu^2}{N_1^2}\sum_{ijt}\sum_{t'\neq t}f_{(ijt)(ijt')}(\mathbf{x}_0,\mathbf{x}_0)\\
&\quad+\frac{\sigma_v^2}{N_1^2}\sum_{ijt}\sum_{j'\neq j}f_{(ijt)(ij't)}(\mathbf{x}_0,\mathbf{x}_0)+\frac{\sigma_\zeta^2}{N_1^2}\sum_{ijt}\sum_{i'\neq i}f_{(ijt)(i'jt)}(\mathbf{x}_0,\mathbf{x}_0)\\
&=\frac{\sigma_\zeta^2}{N_1^2}\sum_{ijt}\sum_{i'\neq i}f_{((ijt)(i'jt)}(\mathbf{x}_0,\mathbf{x}_0)+O((N_1|\mathbf{H}|)^{-1}+N_1^{-1})\\
&=\sigma_\zeta^2C_1(\mathbf{x}_0)+o(1),
\end{aligned}
\tag{7.28}
$$

$$
\begin{aligned}
&(N_1|\mathbf{H}|)^{-2}\mathrm{E}\left[\mathbf{H}^{-1}\left(\mathbf{X}-\mathbf{x}_0'\otimes\iota_{\mathbb{N}}\right)'\mathbf{K}_{\mathbf{H}}(\mathbf{x}_0)\Omega_u\mathbf{K}_{\mathbf{H}}(\mathbf{x}_0)\iota_{\mathbb{N}}\right]\\
&=\frac{\sigma_\zeta^2\kappa_{12}}{N_1^2}\sum_{ijt}\sum_{i'\neq i}\mathbf{H}\frac{\partial f_{(ijt)(i'jt)}(\mathbf{x}_0,\mathbf{x}_0)}{\partial\mathbf{x}_{i'jt}}+O\left(\|\mathbf{H}\|\left((N_1|\mathbf{H}|)^{-1}+N_1^{-1}\right)\right)=O\left(\|\mathbf{H}\|\right),
\end{aligned}
$$

and

$$
\begin{aligned}
&(N_1|\mathbf{H}|)^{-2}\mathbf{H}^{-1}\mathrm{E}\left[\left(\mathbf{X}-\mathbf{x}_0'\otimes\iota_{\mathbb{N}}\right)'\mathbf{K}_{\mathbf{H}}(\mathbf{x}_0)\Omega_u\mathbf{K}_{\mathbf{H}}(\mathbf{x}_0)\left(\mathbf{X}-\mathbf{x}_0{}'\otimes\iota_{\mathbb{N}}\right)\right]\mathbf{H}^{-1}\\
&\approx\frac{\sigma_\zeta^2\kappa_{12}^2}{N_1^2}\sum_{ijt}\sum_{i'\neq i}\mathbf{H}\frac{\partial f_{(ijt)(i'jt)}(\mathbf{x}_0,\mathbf{x}_0)}{\partial\mathbf{x}_{ijt}\partial\mathbf{x}_{i'jt}'}\mathbf{H}+O\left(\|\mathbf{H}\|^2\left((N_1|\mathbf{H}|)^{-1}+N_1^{-1}\right)\right)\\
&=O\left(\|\mathbf{H}\|^2\right),
\end{aligned}
$$

if $\|\mathbf{H}\|\to 0$ and $N_1|\mathbf{H}|\to\infty$ as $N_1\to\infty$, where

$$
C_1(\mathbf{x}_0)=\lim_{N_1\to\infty}N_1^{-2}\sum_{ijt}\sum_{i'\neq i}f_{(i,j,t),(i',j,t)}(\mathbf{x}_0,\mathbf{x}_0).
$$

It follows that

$$
\mathrm{Var}\left(\frac{D_n^{-1}\mathscr{C}}{N_1|\mathbf{H}|}\right)\approx\begin{bmatrix}C_1(\mathbf{x}_0) & O(\|\mathbf{H}\|)\\ O(\|\mathbf{H}\|) & O(\|\mathbf{H}\|^2)\end{bmatrix}.
$$

In addition, from the proofs given in Sect. 7.2, we have

$$
\frac{D_n^{-1}\mathscr{A}}{N_1|\mathbf{H}|}\xrightarrow{p}\bar{f}_1(\mathbf{x}_0)\begin{bmatrix}1 & \mathbf{0}_k'\\ \mathbf{0}_k & \kappa_{12}\mathbf{I}_k\end{bmatrix}
$$

and

$$
\frac{D_n^{-1}\mathscr{B}}{2N_1|\mathbf{H}|}=\frac{\kappa_{12}}{2}\begin{bmatrix}\mathrm{tr}\left\{\mathbf{H}\mathbf{m}^{(2)}(\mathbf{x}_0)\mathbf{H}\right\}\bar{f}_1(\mathbf{x}_0)\\ \mathbf{0}_k\end{bmatrix}+o_p\left(\|\mathbf{H}\|^2\right)
$$

if $\|\mathbf{H}\| \to 0$ as $N_1 \to \infty$, where we denote $\bar{f}_1(\mathbf{x}_0) = \lim_{N_1 \to \infty} N_1^{-1} \sum_{ijt} f_{ijt}(\mathbf{x}_0)$. Combining all the results above gives

$$\begin{aligned} \hat{m}(\mathbf{x}_0) - m(\mathbf{x}_0) &\approx \kappa_{12} \mathrm{tr}\left\{ \mathbf{H}\mathbf{m}^{(2)}(\mathbf{x}_0)\mathbf{H} \right\}/2 + \bar{f}_1^{-2}(\mathbf{x}_0) \frac{\iota_{\mathbb{N}}' \mathbf{K}_{\mathbf{H}}(\mathbf{x}_0)\mathbf{u}}{N_1 |\mathbf{H}|} \\ &= O_p(\|\mathbf{H}\|^2) + O_e(1) \end{aligned} \tag{7.29}$$

if $\|\mathbf{H}\| \to 0$ and $N_1 |\mathbf{H}| \to \infty$ as $N_1 \to \infty$, where the second term on the RHS of equation (7.29) is of order $O_e(1)$ since it has a finite positive variance, i.e.,

$$\mathrm{Var}[(N_1 |\mathbf{H}|)^{-1} \iota_{\mathbb{N}}' \mathbf{K}_{\mathbf{H}}(\mathbf{x}_0)\mathbf{u}] = (N_1 |\mathbf{H}|)^{-2} \mathrm{E}\left[\iota_{\mathbb{N}}' \mathbf{K}_{\mathbf{H}}(\mathbf{x}_0) \Omega_u \mathbf{K}_{\mathbf{H}}(\mathbf{x}_0) \iota_{\mathbb{N}} \right] = O_e(1)$$

by the result in (7.28). Therefore, the pooled local linear estimator $\hat{m}(\mathbf{x}_0)$ is not a consistent estimator of $m(\mathbf{x}_0)$ due to the large variance term. Because $N_1 |\mathbf{H}| \to \infty$ as $N_1 \to \infty$, condition (7.15) is violated and $\hat{m}(\mathbf{x}_0)$ fails to converge to $m(\mathbf{x}_0)$. Evidently, if $\sigma_\zeta^2 = 0$, i.e., the error term u_{ijt} does not contain an *index i-invariant* random effects, ζ_{jt}, we would have

$$\hat{m}(\mathbf{x}_0) - m(\mathbf{x}_0) = O_p\left(\|\mathbf{H}\|^2 + (N_1 |\mathbf{H}|)^{-1/2} \right)$$

so that $\hat{m}(\mathbf{x}_0)$ is a consistent estimator of $m(\mathbf{x}_0)$. As both N_2 and T are finite, one can estimate model (7.1)-(7.24) by taking $\zeta_{j,t}$'s as parameters to be estimated. That is, we take ζ_{jt} as *index i-invariant* fixed effects in model

$$y_{ijt} = m(\mathbf{x}_{ijt}) + \zeta_{jt} + \mu_{ij} + v_{it} + \varepsilon_{ijt}$$

and $\mu_{ij} + v_{it} + \varepsilon_{ijt}$ is the composite error. We will not provide further details along this line of research as it is beyond the scope of this chapter.

For cases (ii) and (iii), similar results hold: The pooled local linear estimator is inconsistent as its asymptotic variance term is of order $O_e(1)$. However,

$$\hat{m}(\mathbf{x}_0) - m(\mathbf{x}_0) = O_p(\|\mathbf{H}\|^2 + (\mathbb{N}_c |\mathbf{H}|)^{-1/2})$$

would hold if $\sigma_v^2 = 0$, i.e., the error term u_{ijt} does not contain an *index j-invariant* random effects, v_{it}, under case (ii); and if $\sigma_\mu^2 = 0$, i.e., the error term u_{ijt} does not contain an *index t-invariant* random effects, μ_{ij}, under case (iii). In Table 7.2, we summarize the asymptotic results of the pooled local linear estimators for cases (i)-(iii).

Table 7.2 Asymptotics of the estimators for cases (i) – (iii)

Case	Condition	Leading bias term	Leading variance term
(i)	$N_1 \to \infty$ $\|H\| \to 0$ $N_1 \|H\| \to \infty$	$\kappa_{1,2}\mathrm{tr}\left\{Hm^{(2)}(x_0)H\right\}/2$	$\frac{1'_{N_1N_2T}K_H(x_0)\zeta}{\bar{f}_1(x_0)N_1\|H\|} = O_e(1)$
(ii)	$N_2 \to \infty$ $\|H\| \to 0$ $N_2 \|H\| \to \infty$	$\kappa_{1,2}\mathrm{tr}\left\{Hm^{(2)}(x_0)H\right\}/2$	$\frac{1'_{N_1N_2T}K_H(x_0)\upsilon}{\bar{f}_2(x_0)N_2\|H\|} = O_e(1)$ [a]
(iii)	$T \to \infty$ $\|H\| \to 0$ $T \|H\| \to \infty$	$\kappa_{1,2}\mathrm{tr}\left\{Hm^{(2)}(x_0)H\right\}/2$	$\frac{1'_{N_1N_2T}K_H(x_0)\mu}{\bar{f}_3(x_0)T\|H\|} = O_e(1)$ [b]

[a] $\bar{f}_2(x_0) = \lim_{N_2\to\infty} N_2^{-1}\sum_{i,j,t} f_{i,j,t}(x_0)$.
[b] $\bar{f}_3(x_0) = \lim_{T\to\infty} T^{-1}\sum_{i,j,t} f_{i,j,t}(x_0)$.

7.4.2 *Cases (iv)-(vi): The Sample Size Increases in Two out of the Three Indices*

For cases (iv)–(vi), in which two of the three sample indices go to infinity, we first consider case (iv) in which $N_1 \to \infty$ and $N_2 \to \infty$ with T fixed and $\mathbb{N}_4 = N_1N_2$, and the results for cases (v) and (vi) can be derived in the same way. In addition to the bandwidth condition given in Table 7.1 for case (iv) that $\|\mathbf{H}\| \to 0$ and $N_1N_2|\mathbf{H}| \to \infty$ as N_1 and N_2 go to infinity, we further assume that $N_1 \geq N_2$ without loss of generality.

From (7.26), we have

$$\mathscr{R}_{\mathbb{N}_4} = \mathbb{N}_4^{-1}|\mathbf{H}| \sum_{i_1j_1t_1} \sum_{(ijt)_2 \neq (ijt)_1} \left|\mathrm{E}\left(u_{i_1j_1t_1}u_{i_2j_2t_2}\right)\right| = O\left(\max(N_1,N_2)|\mathbf{H}|\right),$$

which may approach to 0, a positive finite number, or infinity when $\max(N_1,N_2)|\mathbf{H}|$ approaches to 0, a positive finite number, or infinity, respectively. Under the three different limit arrangements, it turns out that the asymptotic results of the local linear estimator are different.

Under the first limit arrangement, $\max(N_1,N_2)|\mathbf{H}| = N_1|\mathbf{H}| \to 0$, so that $\mathscr{R}_{\mathbb{N}_4} \to 0$. Hence, the covariance terms in (7.10), (7.11) and (7.12) are ignorable under the error structure defined in (7.24), and we have

$$\mathrm{Var}\left(\frac{D_n^{-1}\mathscr{C}}{\sqrt{\mathbb{N}_4|\mathbf{H}|}}\right) \approx \Sigma_4(\mathbf{x}_0)\begin{pmatrix} \nu_0 & \mathbf{0}_k' \\ \mathbf{0}_k & \kappa_{22}\mathbf{I}_k(\sigma_\mu^2+\sigma_\nu^2+\sigma_\zeta^2+\sigma_\varepsilon^2)\bar{f}_4(\mathbf{x}_0)\end{pmatrix},$$

where

$$\Sigma_4(\mathbf{x}_0) = (\sigma_\mu^2 + \sigma_v^2 + \sigma_\zeta^2 + \sigma_\varepsilon^2)\bar{f}_4(\mathbf{x}_0) \quad \text{and} \quad \bar{f}_4(\mathbf{x}_0) = \lim_{\mathbb{N}_4 \to \infty} \mathbb{N}_4^{-1} \sum_{ijt} f_{ijt}(\mathbf{x}_0) \,.$$

Under some regularity conditions, the limiting distribution of the pooled local linear estimator for case (iv) becomes

$$\begin{aligned} &\sqrt{\mathbb{N}_4 |\mathbf{H}|} \left(\hat{m}(\mathbf{x_0}) - m(\mathbf{x}_0) - \kappa_{12} \mathrm{tr} \left\{ \mathbf{H} m^{(2)}(\mathbf{x}_0) \mathbf{H} \right\} /2 \right) \\ &\xrightarrow{d} \mathcal{N} \left(0, \nu_0 (\sigma_\mu^2 + \sigma_v^2 + \sigma_\zeta^2 + \sigma_\varepsilon^2)/\bar{f}_4(\mathbf{x}_0) \right), \end{aligned} \tag{7.30}$$

if $\|\mathbf{H}\| \to 0$, $\mathbb{N}_4 |\mathbf{H}| \to \infty$, $N_1 |\mathbf{H}| \to 0$ as $N_2 \to \infty$. The convergence rate of the estimator is

$$\hat{m}(\mathbf{x}_0) - m(\mathbf{x}_0) = O_p \left(\|\mathbf{H}\|^2 + (\mathbb{N}_4 |\mathbf{H}|)^{-1/2} \right) \,.$$

Next, under the second limit arrangement, when $N_1 |\mathbf{H}| \to c_1 \in (0, \infty)$ and $N_2 |\mathbf{H}| \to c_2 \in [0, c_1]$ with $N_1 \geq N_2$, we have $\mathscr{R}_{\mathbb{N}_4} = O(1)$. As in Sect. 7.2, we need to calculate (7.10), (7.11) and (7.12) under the error structure defined in (7.24). Straightforward calculation gives

$$\begin{aligned} &(\mathbb{N}_4 |\mathbf{H}|)^{-1} \mathbb{E} \left[1_{\mathbb{N}}' \, \mathbf{K}_{\mathbf{H}}(\mathbf{x}_0) \Omega_u \mathbf{K}_{\mathbf{H}}(\mathbf{x}_0) 1_{\mathbb{N}} \right] \\ &\approx \nu_0 (\sigma_\mu^2 + \sigma_v^2 + \sigma_\zeta^2 + \sigma_\varepsilon^2) \frac{1}{\mathbb{N}_4} \sum_{ijt} f_{ijt}(\mathbf{x}_0) + N_1 |\mathbf{H}| \frac{\sigma_\zeta^2}{N_1^2 N_2} \sum_{ijt} \sum_{i' \neq i} f_{(ijt)(i'jt)}(\mathbf{x}_0, \mathbf{x}_0) \\ &+ N_2 |\mathbf{H}| \frac{\sigma_v^2}{N_1 N_2^2} \sum_{ijt} \sum_{j' \neq j} f_{(ijt)(ij't)}(\mathbf{x}_0, \mathbf{x}_0) + |\mathbf{H}| \frac{\sigma_\mu^2}{N_1 N_2} \sum_{ijt} \sum_{t' \neq t} f_{(ijt)(ijt')}(\mathbf{x}_0, \mathbf{x}_0) \,, \end{aligned}$$

where the first three terms are positive and bounded from above and the last term is of order $O(|\mathbf{H}|)$. Denoting

$$C_4(\mathbf{x}_0) = \nu_0 (\sigma_\mu^2 + \sigma_v^2 + \sigma_\zeta^2 + \sigma_\varepsilon^2) \bar{f}_4(\mathbf{x}_0) + c_1 \sigma_\zeta^2 \bar{f}_{4,\zeta}(\mathbf{x}_0, \mathbf{x}_0) + c_2 \sigma_v^2 \bar{f}_{4,v}(\mathbf{x}_0, \mathbf{x}_0)$$

with

$$\begin{aligned} \bar{f}_{4,\zeta}(\mathbf{x}_0, \mathbf{x}_0) &= \lim_{\mathbb{N}_4 \to \infty} \frac{1}{N_1^2 N_2} \sum_{ijt} \sum_{i' \neq i} f_{(ijt)(i'jt)}(\mathbf{x}_0, \mathbf{x}_0) \\ \bar{f}_{4,v}(\mathbf{x}_0, \mathbf{x}_0) &= \lim_{\mathbb{N}_4 \to \infty} \frac{1}{N_1 N_2^2} \sum_{ijt} \sum_{j' \neq j} f_{(ijt)(ij't)}(\mathbf{x}_0, \mathbf{x}_0), \end{aligned}$$

we can show that

$$\mathrm{Var} \left(\frac{D_n^{-1} \mathscr{C}}{\sqrt{\mathbb{N}_4 |\mathbf{H}|}} \right) \approx \begin{pmatrix} C_4(\mathbf{x}_0) & O(\|\mathbf{H}\|) \\ O(\|\mathbf{H}\|) & \kappa_{22} \mathbf{I}_k (\sigma_\mu^2 + \sigma_v^2 + \sigma_\zeta^2 + \sigma_\varepsilon^2) \bar{f}_4(\mathbf{x}_0) \end{pmatrix}. \tag{7.31}$$

Therefore, $\mathscr{A}^{-1} \mathscr{C} = O_p((\mathbb{N}_4 |\mathbf{H}|)^{-1/2})$ and then

$$\hat{m}(\mathbf{x}_0) - m(\mathbf{x}_0) = O_p(\|\mathbf{H}\|^2 + (\mathbb{N}_4|\mathbf{H}|)^{-1/2}) .$$

Under some regularity conditions, we have

$$\begin{aligned} &\sqrt{\mathbb{N}_4|\mathbf{H}|}\left(\hat{m}(\mathbf{x}_0) - m(\mathbf{x}_0) - \kappa_{12}\mathrm{tr}\left\{\mathbf{H}\mathbf{m}^{(2)}(\mathbf{x}_0)\mathbf{H}\right\}/2\right) \\ &\xrightarrow{d} \mathscr{N}\left(0, \nu_0\frac{\sigma_\mu^2+\sigma_v^2+\sigma_\zeta^2+\sigma_\varepsilon^2}{\bar{f}_4(\mathbf{x}_0)} + c_1\frac{\sigma_\zeta^2\bar{f}_{4,\zeta}(\mathbf{x}_0,\mathbf{x}_0)}{\bar{f}_4^2(\mathbf{x}_0)} + c_2\frac{\sigma_v^2\bar{f}_{4,v}(\mathbf{x}_0,\mathbf{x}_0)}{\bar{f}_4^2(\mathbf{x}_0)}\right) \end{aligned} \tag{7.32}$$

if $\|\mathbf{H}\| \to 0$, $\mathbb{N}_4|\mathbf{H}| \to \infty$, $N_1|\mathbf{H}| \to c_1 \in (0,\infty)$, and $N_2|\mathbf{H}| \to c_2 \in [0,c_1]$ as $N_2 \to \infty$ and $N_1 \geq N_2$.

Finally, under the third limit arrangement, when $N_1|\mathbf{H}| \to \infty$ and $N_2/N_1 \to r_{2,1} \in [0,1]$, we have $\mathscr{R}_{\mathbb{N}_4} \to \infty$. It is clear that, when $N_1|\mathbf{H}| \to \infty$, (7.10) is explosive due to large covariance terms. For the $(1,1)$th element of $\mathrm{Var}\left(\mathscr{C}/\sqrt{\mathbb{N}_4|\mathbf{H}|N_1|\mathbf{H}|}\right)$, we obtain

$$\begin{aligned} &(\mathbb{N}_4|\mathbf{H}|)^{-1}(N_1|\mathbf{H}|)^{-1}\mathrm{E}\left[\iota_{\mathbb{N}}'\mathbf{K}_{\mathbf{H}}(\mathbf{x}_0)\Omega_u\mathbf{K}_{\mathbf{H}}(\mathbf{x}_0)\iota_{\mathbb{N}}\right] \\ \approx\ & \frac{\nu_0}{N_1|\mathbf{H}|}(\sigma_\mu^2+\sigma_v^2+\sigma_\zeta^2+\sigma_\varepsilon^2)\frac{1}{\mathbb{N}_4}\sum_{ijt}f_{ijt}(\mathbf{x}_0) + \frac{\sigma_\zeta^2}{N_1^2N_2}\sum_{ijt}\sum_{i'\neq i}f_{(ijt)(i'jt)}(\mathbf{x}_0,\mathbf{x}_0) \\ &+\sigma_v^2\frac{N_2}{N_1}\frac{1}{N_1N_2^2}\sum_{ijt}\sum_{j'\neq j}f_{(ijt)(ij't)}(\mathbf{x}_0,\mathbf{x}_0) + \frac{\sigma_\mu^2}{N_1}\frac{1}{N_1N_2}\sum_{ijt}\sum_{t_2'\neq t}f_{(ijt)(ijt')}(\mathbf{x}_0,\mathbf{x}_0) \\ \to\ & C_{4,3}(\mathbf{x}_0) , \end{aligned}$$

where $C_{4,3}(\mathbf{x}_0) = \sigma_\zeta^2\bar{f}_{4,\zeta}(\mathbf{x}_0,\mathbf{x}_0) + r_{2,1}\sigma_v^2\bar{f}_{4,v}(\mathbf{x}_0,\mathbf{x}_0)$, and

$$\mathrm{Var}\left(\frac{D_n^{-1}\mathscr{C}}{\sqrt{\mathbb{N}_4|\mathbf{H}|}\sqrt{N_1|\mathbf{H}|}}\right) \approx \begin{pmatrix} C_{4,3}(\mathbf{x}_0) & O(\|\mathbf{H}\|) \\ O(\|\mathbf{H}\|) & O\left(\|\mathbf{H}\|^2 + (N_1|\mathbf{H}|)^{-1}\right) \end{pmatrix}.$$

Therefore,we have

$$\mathscr{A}^{-1}\mathscr{C} = \frac{\sqrt{N_1|\mathbf{H}|}}{\sqrt{\mathbb{N}_4|\mathbf{H}|}}\left(\frac{D_n^{-1}\mathscr{A}}{\mathbb{N}_4|\mathbf{H}|}\right)^{-1}\left(\frac{D_n^{-1}\mathscr{C}}{\sqrt{\mathbb{N}_4|\mathbf{H}|}\sqrt{N_1|\mathbf{H}|}}\right) = O_p\left(\frac{1}{\sqrt{N_2}}\right).$$

This implies $\hat{m}(\mathbf{x}_0) - m(\mathbf{x}_0) = O_p(\|\mathbf{H}\|^2 + 1/\sqrt{N_2})$. Similarly, under some suitable regularity conditions, we obtain the limiting distribution of the estimator

$$\begin{aligned} &\sqrt{N_2}\left(\hat{m}(\mathbf{x}_0) - m(\mathbf{x}_0) - \kappa_{12}\mathrm{tr}\left\{\mathbf{H}\mathbf{m}^{(2)}(\mathbf{x}_0)\mathbf{H}\right\}/2\right) \\ &\xrightarrow{d} \mathscr{N}\left(0, \frac{\sigma_\zeta^2\bar{f}_{4,\zeta}(\mathbf{x}_0,\mathbf{x}_0)}{\bar{f}_4^2(\mathbf{x}_0)} + r_{2,1}\frac{\sigma_v^2\bar{f}_{4,v}(\mathbf{x}_0,\mathbf{x}_0)}{\bar{f}_4^2(\mathbf{x}_0)}\right) \end{aligned} \tag{7.33}$$

if $\|\mathbf{H}\| \to 0$, $\mathbb{N}_4|\mathbf{H}| \to \infty$, $N_1 \geq N_2$, $N_2/N_1 \to r_{2,1} \in [0,1]$, and $N_1|\mathbf{H}| \to \infty$ as $N_2 \to \infty$.

Now we discuss the optimal convergence rate and the optimal bandwidth choice of the local linear estimator $\widehat{m}(\mathbf{x}_0)$ under the three different limit arrangements. The stochastic leading bias and variance term of the estimator are these:

$$\text{Bias term} \equiv \mathscr{A}^{-1}\mathscr{B} = O_p(\|\mathbf{H}\|^2)$$

$$\text{Variance term} \equiv \mathscr{A}^{-1}\mathscr{C} = \begin{cases} O_p((N_1N_2|\mathbf{H}|)^{-1/2}), & \text{if } N_1|\mathbf{H}| \to c \in [0,+\infty), \\ O_p(N_2^{-1/2}), & \text{if } N_1|\mathbf{H}| \to +\infty. \end{cases}$$

Evidently, the asymptotic variance of $\hat{m}(\mathbf{x}_0)$ is never smaller than order $O(N_2^{-1})$ no matter how large k is if $N_1|\mathbf{H}| \to +\infty$. As in the conventional nonparametric literature, the optimal bandwidth is set to balance the asymptotic squared bias term and the asymptotic variance term of $\hat{m}(\mathbf{x}_0)$ and the optimal convergence rate of $\hat{m}(\mathbf{x}_0)$ in MSE is the convergence rate reached when the optimal bandwidth is used for calculation. Let $\mathbf{H}_{opt}$ be the optimal bandwidth. Below, we discuss how the choice of $\mathbf{H}_{opt}$ is affected by N_1, N_2 and k, and delay the detailed proof to Sect. 7.6. Without loss of generality, we assume that $\mathbf{H} = h\mathbf{I}_k$ (i.e., the bandwidths for all the k regressors are the same) and $N_1 \geq N_2$.

- When $1 \leq k \leq 3$, we have $N_1/N_2^{k/4} \to \infty$ as $N_1 \geq N_2 \to \infty$. It can be shown that the optimal MSE rate of the pooled local linear estimator is always $O_e\left(N_2^{-1}\right)$. However, there is a range for h to achieve the optimal MSE rate. A simple choice for h is that $h_{opt} = c_{opt}N_1^{-1/k}$ for some finite constant $c_{opt} > 0$, we have $N_1|\mathbf{H}_{opt}| \to c > 0$ with asymptotic variance of the estimator given in (7.32). (For other choices of optimal bandwidth, please refer to the technical Appendix at the end of this chapter.)
- When $k = 4$, the ratio, $N_1/N_2^{k/4} = N_1/N_2$, can approach to either ∞ or a finite positive constant $c > 0$. It can be shown that the optimal MSE rate of the pooled local linear estimator is always $O_e\left(N_2^{-1}\right)$, which can be obtained by setting the rate of the optimal bandwidth as follows:
  - If $N_1/N_2 \to \infty$ and the optimal bandwidth satisfies $h_{opt} = O_e(N_2^{-1/4})$, we have $N_1|\mathbf{H}_{opt}| \to \infty$ with the asymptotic variance of the estimator given in (7.33).
  - If $N_1/N_2 \to \infty$ and the optimal bandwidth satisfies $h_{opt} = O_e(N_1^{-1/4})$, we have $N_1|\mathbf{H}_{opt}| \to c > 0$ with the asymptotic variance of the estimator given in (7.32).
  - If $N_1/N_2 \to c \geq 1$ and the optimal bandwidth satisfies $h_{opt} = O_e(N_1^{-1/4})$, we have $N_1|\mathbf{H}_{opt}| \to c > 0$ with the asymptotic variance of the estimator given in (7.32).
- When $k \geq 5$, the ratio $N_1/N_2^{k/4}$ can approach to 0, a positive finite number c, or ∞. It can be shown that the optimal MSE rate of the pooled local linear estimator is either $O_e((N_1N_2)^{-4/(k+4)})$ or $O_e(N_2^{-1})$ if $N_1/N_2^{k/4} \to 0$ or $N_1/N_2^{k/4} \to c \in (0,\infty]$, respectively. More specifically,

- If $N_1/N_2^{k/4} \to 0$, the optimal MSE rate is $O_e((N_1N_2)^{-4/(k+4)})$ with the optimal bandwidth $h_{opt} = O_e((N_1N_2)^{-1/(k+4)})$, which corresponds to $N_1|\mathbf{H}| \to 0$ with the asymptotic variance of the estimator given in (7.30).
- If $N_1/N_2^{k/4} \to c \in (0,\infty)$, the optimal MSE rate is $O_e(N_2^{-1})$ with the optimal bandwidth $h_{opt} = O_e(N_2^{-1/4})$, which corresponds to $N_1|\mathbf{H}| \to c \in (0,\infty)$ with the asymptotic variance of the estimator given in (7.32).
- If $N_1/N_2^{k/4} \to \infty$, the optimal MSE rate is $O_e(N_2^{-1})$ with the optimal bandwidth $h_{opt} = O_e(N_2^{-1/4})$, which corresponds to $N_1|\mathbf{H}| \to \infty$ with the asymptotic variance of the estimator given in (7.33).

We clearly see that the optimal MSE of $\hat{m}(\mathbf{x}_0)$ depends on k and the ratio of $N_1/N_2^{k/4}$. When the number of regressors, k, is less than or equal to 4, the optimal MSE of $\hat{m}(\mathbf{x}_0)$ is always $O_e\left(N_2^{-1}\right)$, which does not depend on k. When $k \geq 5$ and $N_1/N_2^{k/4} \to 0$ hold, the optimal MSE of $\hat{m}(\mathbf{x}_0)$ is $O_e\left((N_1N_2)^{-4/(k+4)}\right)$. Moreover, if $k \geq 5$ and $N_2^{k/4}/N_1 \to c \geq 0$, the optimal MSE of $\hat{m}(\mathbf{x}_0)$ continues to be of order $O_e\left(N_2^{-1}\right)$.

We give two examples on determining the optimal convergence rate. For the first example, when $k = 5$ and $N_2 = N_1^{1/2}$, it is easy to check that

$$\frac{N_1}{N_2^{k/4}} = \frac{N_2^2}{N_2^{5/4}} = N_2^{3/4} \to \infty,$$

which means that the optimal MSE rate of the estimator is $O_e(N_2^{-1})$. For the second example, when $k = 5$ and $N_2 = N_1$, it is easy to check that

$$\frac{N_1}{N_2^{k/4}} = \frac{N_2}{N_2^{5/4}} = N_2^{-1/4} \to 0,$$

which means that the optimal bandwidth rate is

$$h_{opt} = O((N_1N_2)^{-1/9}) = O(N_2^{-2/9})$$

and the optimal MSE rate of the estimator is

$$O((N_1N_2)^{-4/9}) = O(N_2^{-8/9}).$$

Table 7.3 summarizes the asymptotic results of the pooled local linear estimators for cases (iv) under four sets of different conditions on the sample indices and orders. Due to space limitation, we only report their asymptotic results for case (v) and (vi) in Table 7.4 and 7.5 respectively, and omit their derivations and the discussion about the optimal convergence rate of the estimator.

Table 7.3 Asymptotics of the estimators for case (iv): $N_1, N_2 \to \infty$ with T fixed ($N_1 \geq N_2$)

Condition	Bias term rate	Variance term rate	Asymptotic variance
$\|\mathbf{H}\| \to 0$ $\mathbb{N}_4 \|\mathbf{H}\| \to \infty$ [a] $\max(N_1, N_2) \|\mathbf{H}\| \to 0$	$O_p(\|\mathbf{H}\|^2)$	$O_p((N_4 \|\mathbf{H}\|)^{-1/2})$	$v_0 \frac{\sigma_\mu^2+\sigma_v^2+\sigma_\zeta^2+\sigma_\varepsilon^2}{\bar{f}_4(x_0)}$ [b]
$\|\mathbf{H}\| \to 0$ $\mathbb{N}_4 \|\mathbf{H}\| \to \infty$ $N_1 \|\mathbf{H}\| \to c_1 \in (0, \infty)$ $N_2 \|\mathbf{H}\| \to c_2 \in [0, c_1]$	$O_p(\|\mathbf{H}\|^2)$	$O_p((N_4 \|\mathbf{H}\|)^{-1/2})$	$v_0 \frac{\sigma_\mu^2+\sigma_v^2+\sigma_\zeta^2+\sigma_\varepsilon^2}{\bar{f}_4(x_0)}$ $+c_1 \frac{\sigma_\zeta^2 \bar{f}_{4,\zeta}(x_0,x_0)}{\bar{f}_4^2(x_0)}$ [c] $+c_2 \frac{\sigma_v^2 \bar{f}_{4,v}(x_0,x_0)}{\bar{f}_4^2(x_0)}$ [d]
$\|\mathbf{H}\| \to 0$ $\mathbb{N}_4 \|\mathbf{H}\| \to \infty$ $N_2/N_1 \to r_{2,1} \in [0, 1]$ $N_1 \|\mathbf{H}\| \to \infty$	$O_p(\|\mathbf{H}\|^2)$	$O_p((N_2)^{-1/2})$	$\frac{\sigma_\zeta^2 \bar{f}_{4,\zeta}(x_0,x_0)}{\bar{f}_4^2(x_0)}$ $+r_{2,1} \frac{\sigma_v^2 \bar{f}_{4,v}(x_0,x_0)}{\bar{f}_4^2(x_0)}$

[a] $\mathbb{N}_4 = N_1 N_2$.
[b] $\bar{f}_4(x_0) = \lim_{\mathbb{N}_4 \to \infty} \mathbb{N}_4^{-1} \sum_{i,j,t} f_{ijt}(x_0)$.
[c] $\bar{f}_{4,\zeta}(x_0,x_0) = \lim_{\mathbb{N}_4 \to \infty} \frac{1}{N_1 \mathbb{N}_4} \sum_{i,j,t} \sum_{i' \neq i} f_{(i,j,t),(i',j,t)}(x_0,x_0)$.
[d] $\bar{f}_{4,v}(x_0,x_0) = \lim_{\mathbb{N}_4 \to \infty} \frac{1}{N_2 \mathbb{N}_4} \sum_{i,j,t} \sum_{j' \neq j} f_{(i,j,t),(i,j',t)}(x_0,x_0)$.

7.4.3 Case (vii): The Sample Size Increases in All Three Indices

For case (vii), in which all three sample indices approach to infinity, we derive the asymptotics of the pairwise random effects local linear estimator under the general sample and bandwidth condition given in Table 7.1 that $\|\mathbf{H}\| \to \infty$ and $N_1 N_2 T \|\mathbf{H}\| \to \infty$ as $\min(N_1, N_2, T) \to \infty$. Without loss of generality, we further assume that $N_1 \geq N_2$ and $N_1 \geq T$ with $N_2/N_1 \to r_{2,1} \in [0, 1]$ and $T/N_1 \to r_{T,1} \in [0, 1]$, and let $\mathbb{N} \equiv N_1 N_2 T$.

Similarly to cases (iv)-(vi),

$$\mathscr{R}_{\mathbb{N}} = \mathbb{N}^{-1} |\mathbf{H}| \sum_{i_1 j_1 t_1} \sum_{(ijt)_2 \neq (ijt)_1} \left| \mathrm{E}\left(u_{i_1 j_1 t_1} u_{i_2 j_2 t_2}\right) \right|$$

may approach to 0, a positive finite number, or infinity when $\max(N_1, N_2, T)|\mathbf{H}| = N_1 |\mathbf{H}|$ approaches to 0, a positive finite number, or infinity, respectively. Under the three limit arrangements, it turns out that the asymptotic results of the local linear estimator are different.

Under the first limit arrangement, when $\max(N_1, N_2, T) |\mathbf{H}| = N_1 |\mathbf{H}| \to 0$, we have $\mathscr{R}_{\mathbb{N}} \to 0$. Like the benchmark model that we consider in Sect. 7.2, we obtain

$$\hat{m}(\mathbf{x}_0) - m(\mathbf{x}_0) = O_p\left(\|\mathbf{H}\|^2 + (\mathbb{N} |\mathbf{H}|)^{-1/2} \right).$$

Table 7.4 Asymptotics of the estimators for case (v): $N_1, T \to \infty$ with N_2 fixed

Condition	Bias term rate	Variance term rate	Asymptotic variance
$\|\mathbf{H}\| \to 0$ $\mathbb{N}_5 \|\mathbf{H}\| \to \infty$ [a] $\max(N_1, N_2) \|\mathbf{H}\| \to 0$	$O_p(\|\mathbf{H}\|^2)$	$O_p((N_5 \|\mathbf{H}\|)^{-1/2})$	$v_0 \frac{\sigma_\mu^2+\sigma_v^2+\sigma_\zeta^2+\sigma_\varepsilon^2}{\bar{f}_5(x_0)}$ [b]
$\|\mathbf{H}\| \to 0$ $\mathbb{N}_5 \|\mathbf{H}\| \to \infty$ $N_1 \|\mathbf{H}\| \to c_1 \in [0,\infty)$ $T \|\mathbf{H}\| \to c_T \in [0,\infty)$	$O_p(\|\mathbf{H}\|^2)$	$O_p((N_5 \|\mathbf{H}\|)^{-1/2})$	$v_0 \frac{\sigma_\mu^2+\sigma_v^2+\sigma_\zeta^2+\sigma_\varepsilon^2}{\bar{f}_5(x_0)}$ $+c_1 \frac{\sigma_\zeta^2 \bar{f}_{5,\zeta}(x_0,x_0)}{\bar{f}_5^2(x_0)}$ [c] $+c_T \frac{\sigma_\mu^2 \bar{f}_{5,\mu}(x_0,x_0)}{\bar{f}_5^2(x_0)}$ [d]
$\|\mathbf{H}\| \to 0$ $\mathbb{N}_5 \|\mathbf{H}\| \to \infty$ $N_1 \geq T$ $T/N_1 \to r_{T,1} \in [0,1]$ $N_1 \|\mathbf{H}\| \to \infty$	$O_p(\|\mathbf{H}\|^2)$	$O_p((T)^{-1/2})$	$\frac{\sigma_\zeta^2 \bar{f}_{5,\zeta}(x_0,x_0)}{\bar{f}_5^2(x_0)}$ $+r_{T,1} \frac{\sigma_\mu^2 \bar{f}_{5,\mu}(x_0,x_0)}{\bar{f}_5^2(x_0)}$
$\|\mathbf{H}\| \to 0$ $\mathbb{N}_5 \|\mathbf{H}\| \to \infty$ $N_1 < T$ $N_1/T \to r_{1,T} \in [0,1]$ $T \|\mathbf{H}\| \to \infty$	$O_p(\|\mathbf{H}\|^2)$	$O_p((N_1)^{-1/2})$	$r_{1,T} \frac{\sigma_\zeta^2 \bar{f}_{5,\zeta}(x_0,x_0)}{\bar{f}_5^2(x_0)}$ $+\frac{\sigma_\mu^2 \bar{f}_{5,\mu}(x_0,x_0)}{\bar{f}_5^2(x_0)}$

Notes: the results for both sample index order, $N_1 \geq T$ and $N_1 < T$, are included.
[a] $\mathbb{N}_5 = N_1 T$.
[b] $\bar{f}_5(x_0) = \lim_{\mathbb{N}_5 \to \infty} \mathbb{N}_5^{-1} \sum_{i,j,t} f_{ijt}(x_0)$.
[c] $\bar{f}_{5,\zeta}(x_0,x_0) = \lim_{\mathbb{N}_5 \to \infty} \frac{1}{N_1 \mathbb{N}_5} \sum_{i,j,t} \sum_{i' \neq i} f_{(i,j,t),(i',j,t)}(x_0,x_0)$.
[d] $\bar{f}_{5,\mu}(x_0,x_0) = \lim_{\mathbb{N}_5 \to \infty} \frac{1}{T \mathbb{N}_5} \sum_{i,j,t} \sum_{t' \neq t} f_{(i,j,t),(i,j,t')}(x_0,x_0)$.

Under some regularity conditions, we conjecture that the limiting distribution of the pooled local linear estimator is

$$\begin{aligned} &\sqrt{\mathbb{N}|\mathbf{H}|}\left(\hat{m}(\mathbf{x_0}) - m(\mathbf{x}_0) - \kappa_{12} \mathrm{tr}\left\{\mathbf{H} m^{(2)}(\mathbf{x}_0)\mathbf{H}\right\}/2\right) \\ &\xrightarrow{d} \mathscr{N}\left(0, v_0(\sigma_\mu^2 + \sigma_v^2 + \sigma_\zeta^2 + \sigma_\varepsilon^2)/\bar{f}_7(\mathbf{x}_0)\right) \end{aligned} \tag{7.34}$$

if $\|\mathbf{H}\| \to 0$, $\mathbb{N}|\mathbf{H}| \to \infty$, $N_1|\mathbf{H}| \to 0$ holds, where

$$\bar{f}_7(\mathbf{x}_0) = \lim_{\mathbb{N} \to \infty} \mathbb{N}^{-1} \sum_{ijt} f_{ijt}(\mathbf{x}_0) \quad \text{and} \quad \Sigma_7(\mathbf{x}_0) = (\sigma_\mu^2 + \sigma_v^2 + \sigma_\zeta^2 + \sigma_\varepsilon^2)\bar{f}_7(\mathbf{x}_0)\,.$$

Next, under the second limit arrangement, when $\max(N_1, N_2, T)|\mathbf{H}| = N_1|\mathbf{H}| \to c \in (0,\infty)$, we have $\mathscr{R}_\mathbb{N} = O_e(1)$. We consider a general scenario in which $N_1|\mathbf{H}| \to$

Table 7.5 Asymptotics of the estimators for case (vi): $N_2, T \to \infty$ with N_1 fixed

Condition	Bias term rate	Variance term rate	Asymptotic variance
$\|\mathbf{H}\| \to 0$ $\mathbb{N}_6 \|\mathbf{H}\| \to \infty$ [a] $\max(N_2, T) \|\mathbf{H}\| \to 0$	$O_p(\|\mathbf{H}\|^2)$	$O_p((N_6 \|\mathbf{H}\|)^{-1/2})$	$\nu_0 \frac{\sigma_\mu^2+\sigma_\nu^2+\sigma_\zeta^2+\sigma_\varepsilon^2}{\bar{f}_6(x_0)}$ [b]
$\|\mathbf{H}\| \to 0$ $\mathbb{N}_6 \|\mathbf{H}\| \to \infty$ $N_2 \|\mathbf{H}\| \to c_2 \in [0,\infty)$ $T \|\mathbf{H}\| \to c_T \in [0,\infty)$	$O_p(\|\mathbf{H}\|^2)$	$O_p((N_6 \|\mathbf{H}\|)^{-1/2})$	$\nu_0 \frac{\sigma_\mu^2+\sigma_\nu^2+\sigma_\zeta^2+\sigma_\varepsilon^2}{\bar{f}_6(x_0)}$ $+c_2 \frac{\sigma_\nu^2 \bar{f}_{6,\nu}(x_0,x_0)}{\bar{f}_6^2(x_0)}$ [c] $+c_T \frac{\sigma_\mu^2 \bar{f}_{6,\mu}(x_0,x_0)}{\bar{f}_6^2(x_0)}$ [d]
$\|\mathbf{H}\| \to 0$ $\mathbb{N}_6 \|\mathbf{H}\| \to \infty$ $N_2 \geq T$ $T/N_2 \to r_{T,2} \in [0,1]$ $N_2 \|\mathbf{H}\| \to \infty$	$O_p(\|\mathbf{H}\|^2)$	$O_p((T)^{-1/2})$	$\frac{\sigma_\nu^2 \bar{f}_{6,\nu}(x_0,x_0)}{\bar{f}_6^2(x_0)}$ $+r_{T,2} \frac{\sigma_\mu^2 \bar{f}_{6,\mu}(x_0,x_0)}{\bar{f}_6^2(x_0)}$
$\|\mathbf{H}\| \to 0$ $\mathbb{N}_6 \|\mathbf{H}\| \to \infty$ $N_2 < T$ $N_2/T \to r_{2,T} \in [0,1]$ $T \|\mathbf{H}\| \to \infty$	$O_p(\|\mathbf{H}\|^2)$	$O_p((N_2)^{-1/2})$	$r_{2,T} \frac{\sigma_\nu^2 \bar{f}_{6,\nu}(x_0,x_0)}{\bar{f}_6^2(x_0)}$ $+\frac{\sigma_\mu^2 \bar{f}_{6,\mu}(x_0,x_0)}{\bar{f}_6^2(x_0)}$

Notes: the results for both sample index order, $N_2 \geq T$ and $N_2 < T$, are included.
[a] $\mathbb{N}_6 = N_2 T$.
[b] $\bar{f}_6(x_0) = \lim_{\mathbb{N}_6 \to \infty} \mathbb{N}_6^{-1} \sum_{i,j,t} f_{ijt}(x_0)$.
[c] $\bar{f}_{6,\nu}(x_0,x_0) = \lim_{\mathbb{N}_6 \to \infty} \frac{1}{N_2 \mathbb{N}_6} \sum_{i,j,t} \sum_{j' \neq j} f_{(i,j,t),(i,j',t)}(x_0,x_0)$.
[d] $\bar{f}_{6,\mu}(x_0,x_0) = \lim_{\mathbb{N}_6 \to \infty} \frac{1}{T \mathbb{N}_6} \sum_{i,j,t} \sum_{t' \neq t} f_{(i,j,t),(i,j,t')}(x_0,x_0)$.

$c_1 \in (0,\infty)$, $N_2 \|\mathbf{H}\| \to c_2 \in [0,c_1]$, and $T \|\mathbf{H}\| \to c_T \in [0,c_1]$ with $N_1 \geq N_2$ and $N_1 \geq T$, and find that

$$\operatorname{Var}\left(\frac{D_n^{-1}\mathscr{C}}{\sqrt{\mathbb{N}\|\mathbf{H}\|}}\right) \approx \begin{pmatrix} C_{7,2}(\mathbf{x}_0) & O(\|\mathbf{H}\|) \\ O(\|\mathbf{H}\|) & \kappa_{22}\mathbf{I}_k(\sigma_\mu^2+\sigma_\nu^2+\sigma_\zeta^2+\sigma_\varepsilon^2)\bar{f}_7(\mathbf{x}_0) \end{pmatrix}, \tag{7.35}$$

where

$$\begin{aligned} C_{7,2}(\mathbf{x}_0) = {} & \nu_0(\sigma_\mu^2+\sigma_\nu^2+\sigma_\zeta^2+\sigma_\varepsilon^2)\bar{f}_7(\mathbf{x}_0) \\ & + c_1\sigma_\zeta^2 \bar{f}_{7,\zeta}(\mathbf{x}_0,\mathbf{x}_0) + c_2\sigma_\nu^2 \bar{f}_{7,\nu}(\mathbf{x}_0,\mathbf{x}_0) + c_T\sigma_\mu^2 \bar{f}_{7,\mu}(\mathbf{x}_0,\mathbf{x}_0), \end{aligned}$$

in which we assume the existence of the following limits

$$\bar{f}_{7,\zeta}(\mathbf{x}_0,\mathbf{x}_0) \equiv \lim_{\mathbb{N}\to\infty}\frac{1}{N_1\mathbb{N}}\sum_{ijt}\sum_{i'\neq i} f_{(ijt)(i'jt)}(\mathbf{x}_0,\mathbf{x}_0),$$

$$\bar{f}_{7,v}(\mathbf{x}_0,\mathbf{x}_0) \equiv \lim_{\mathbb{N}\to\infty}\frac{1}{N_2\mathbb{N}}\sum_{ijt}\sum_{j'\neq j} f_{(ijt)(ij't)}(\mathbf{x}_0,\mathbf{x}_0),$$

$$\bar{f}_{7,\mu}(\mathbf{x}_0,\mathbf{x}_0) \equiv \lim_{\mathbb{N}\to\infty}\frac{1}{T\mathbb{N}}\sum_{ijt}\sum_{t'\neq t} f_{(ijt)(ijt')}(\mathbf{x}_0,\mathbf{x}_0).$$

Therefore, the variance term $\mathscr{A}^{-1}\mathscr{C}$ is still of the order $O_p((\mathbb{N}|\mathbf{H}|)^{-1/2})$.

Now we show that (7.35) holds by calculating (7.10), (7.11) and (7.12) for current case. First, we have

$$\begin{aligned}
&(\mathbb{N}|\mathbf{H}|)^{-1}\mathrm{E}\left[\iota_{\mathbb{N}}'\,\mathbf{K}_{\mathbf{H}}(\mathbf{x}_0)\Omega_u\mathbf{K}_{\mathbf{H}}(\mathbf{x}_0)\iota_{\mathbb{N}}\right]\\
\approx\ & \nu_0(\sigma_\mu^2+\sigma_v^2+\sigma_\zeta^2+\sigma_\varepsilon^2)\lim_{\mathbb{N}\to\infty}\frac{1}{\mathbb{N}}\sum_{ijt} f_{ijt}(\mathbf{x}_0)\\
&+N_1|\mathbf{H}|\,\sigma_\zeta^2\lim_{\mathbb{N}\to\infty}\frac{1}{N_1^2N_2T}\sum_{ijt}\sum_{i'\neq i} f_{(ijt)(i'jt)}(\mathbf{x}_0,\mathbf{x}_0)\\
&+N_2|\mathbf{H}|\,\sigma_v^2\lim_{\mathbb{N}\to\infty}\frac{1}{N_1N_2^2T}\sum_{ijt}\sum_{j'\neq j} f_{(ijt)(ij't)}(\mathbf{x}_0,\mathbf{x}_0)\\
&+T|\mathbf{H}|\sigma_\mu^2\lim_{\mathbb{N}\to\infty}\frac{1}{N_1N_2T^2}\sum_{ijt}\sum_{t'\neq t} f_{(ijt)(ijt')}(\mathbf{x}_0,\mathbf{x}_0)\\
=\ & \nu_0(\sigma_\mu^2+\sigma_v^2+\sigma_\zeta^2+\sigma_\varepsilon^2)\bar{f}_7(\mathbf{x}_0)\\
&+c_1\sigma_\zeta^2\bar{f}_{7,\zeta}(\mathbf{x}_0,\mathbf{x}_0)+c_2\sigma_v^2\bar{f}_{7,v}(\mathbf{x}_0,\mathbf{x}_0)+c_T\sigma_\mu^2\bar{f}_{7,\mu}(\mathbf{x}_0,\mathbf{x}_0),
\end{aligned}$$

where the four terms above are of the order $O(1)$, $O(N_1|\mathbf{H}|)$, $O(N_2|\mathbf{H}|)$, and $O(T|\mathbf{H}|)$ respectively. Second, it is easy to see

$$(\mathbb{N}|\mathbf{H}|)^{-1}\mathrm{E}\left[\mathbf{H}^{-1}\left(\mathbf{X}-\mathbf{x}_0'\otimes\iota_{\mathbb{N}}\right)'\mathbf{K}_{\mathbf{H}}(\mathbf{x}_0)\Omega_u\mathbf{K}_{\mathbf{H}}(\mathbf{x}_0)\iota_{\mathbb{N}}\right]=O(\|\mathbf{H}\|)$$

and

$$\begin{aligned}
&(\mathbb{N}|\mathbf{H}|)^{-1}\mathbf{H}^{-1}\mathrm{E}\left[\left(\mathbf{X}-\mathbf{x}_0'\otimes\iota_{\mathbb{N}}\right)'\mathbf{K}_{\mathbf{H}}(\mathbf{x}_0)\Omega_u\mathbf{K}_{\mathbf{H}}(\mathbf{x}_0)\left(\mathbf{X}-\mathbf{x}_0{}'\otimes\iota_{\mathbb{N}}\right)\right]\mathbf{H}^{-1}\\
\approx\ & \kappa_{22}\mathbf{I}_k(\sigma_\mu^2+\sigma_v^2+\sigma_\zeta^2+\sigma_\varepsilon^2)\bar{f}_7(\mathbf{x}_0).
\end{aligned}$$

Under some suitable regularity conditions, we conjecture that the limiting distribution of the estimator is given by

$$\begin{aligned}
&\sqrt{\mathbb{N}|\mathbf{H}|}\left(\widehat{m}(\mathbf{x}_0)-m(\mathbf{x}_0)-\kappa_{12}\mathrm{tr}\left\{\mathbf{H}\mathbf{m}^{(2)}(\mathbf{x}_0)\mathbf{H}\right\}/2\right)\\
\xrightarrow{d}\ &\mathscr{N}\left(0,\nu_0\frac{\sigma_\mu^2+\sigma_v^2+\sigma_\zeta^2+\sigma_\varepsilon^2}{\bar{f}_7(\mathbf{x}_0)}+c_1\frac{\sigma_\zeta^2\bar{f}_{7,\zeta}(\mathbf{x}_0,\mathbf{x}_0)}{\bar{f}_7^2(\mathbf{x}_0)}\right.\\
&\qquad\left.+c_2\frac{\sigma_v^2\bar{f}_{7,v}(\mathbf{x}_0,\mathbf{x}_0)}{\bar{f}_7^2(\mathbf{x}_0)}+c_T\frac{\sigma_\mu^2\bar{f}_{7,\mu}(\mathbf{x}_0,\mathbf{x}_0)}{\bar{f}_7^2(\mathbf{x}_0)}\right)
\end{aligned}\tag{7.36}$$

if $\|\mathbf{H}\|\to 0$, $\mathbb{N}|\mathbf{H}|\to\infty$, and $N_1|\mathbf{H}|\to c_1\in(0,\infty)$, $N_2|\mathbf{H}|\to c_2\in[0,c_1]$, and $T|\mathbf{H}|\to c_T\in[0,c_1]$ with $N_1\geq N_2$ and $N_1\geq T$ as $\mathbb{N}\to\infty$. The convergence rate of the estimator is

$$\widehat{m}(\mathbf{x}_0)-m(\mathbf{x}_0)=O_p(\|\mathbf{H}\|^2+(\mathbb{N}|\mathbf{H}|)^{-1/2})\,.$$

Finally, under the third limit arrangement, when $\max(N_1,N_2,T)|\mathbf{H}|=N_1|\mathbf{H}|\to\infty$, we have $\mathscr{R}_\mathbb{N}\to\infty$. Similarly to case (iv), to have a bounded variance for $D_n^{-1}\mathscr{C}$, the normalizing term now should be $1/(\sqrt{\mathbb{N}|\mathbf{H}|}\sqrt{N_1|\mathbf{H}|})$ so that

$$\mathrm{Var}\left(\frac{D_n^{-1}\mathscr{C}}{\sqrt{\mathbb{N}|\mathbf{H}|}\sqrt{N_1|\mathbf{H}|}}\right)\approx\begin{pmatrix}C_{7,3}(\mathbf{x}_0) & O(\|\mathbf{H}\|)\\ O(\|\mathbf{H}\|) & O((N_1|\mathbf{H}|)^{-1}+\|\mathbf{H}\|^2)\end{pmatrix},\tag{7.37}$$

where

$$C_{7,3}(\mathbf{x}_0)=\sigma_\zeta^2\bar{f}_{7,\zeta}(\mathbf{x}_0,\mathbf{x}_0)+r_{2,1}\sigma_v^2\bar{f}_{7,v}(\mathbf{x}_0,\mathbf{x}_0)+r_{T,1}\sigma_\mu^2\bar{f}_{7,\mu}(\mathbf{x}_0,\mathbf{x}_0)\,.$$

Therefore, the variance term $\mathscr{A}^{-1}\mathscr{C}$ is of the order $O_p((N_2T)^{-1/2})$ by

$$\mathscr{A}^{-1}\mathscr{C}=\frac{\sqrt{N_1|\mathbf{H}|}}{\sqrt{\mathbb{N}|\mathbf{H}|}}\left(\frac{D_n^{-1}\mathscr{A}}{\mathbb{N}|\mathbf{H}|}\right)^{-1}\left(\frac{D_n^{-1}\mathscr{C}}{\sqrt{\mathbb{N}|\mathbf{H}|}\sqrt{N_1|\mathbf{H}|}}\right)=O_p\left(\frac{1}{\sqrt{N_2T}}\right).$$

Now we show that equation (7.37) holds by checking on (7.10), (7.11) and (7.12) under the current condition. First, we have

$$\begin{aligned}
&(\mathbb{N}|\mathbf{H}|)^{-1}(N_1|\mathbf{H}|)^{-1}\mathrm{E}\left[\iota_\mathbb{N}'\,\mathbf{K}_\mathbf{H}(\mathbf{x}_0)\Omega_u\mathbf{K}_\mathbf{H}(\mathbf{x}_0)\iota_\mathbb{N}\right]\\
\approx\ &O((N_1|\mathbf{H}|)^{-1})+\sigma_\zeta^2\lim_{\mathbb{N}\to\infty}\frac{1}{N_1^2N_2T}\sum_{ijt}\sum_{i'\neq i}f_{(ijt)(i'jt)}(\mathbf{x}_0,\mathbf{x}_0)\\
&+\frac{N_2}{N_1}\sigma_v^2\lim_{\mathbb{N}\to\infty}\frac{1}{N_1N_2^2T}\sum_{ijt}\sum_{j'\neq j}f_{(ijt)(ij't)}(\mathbf{x}_0,\mathbf{x}_0)\\
&+\frac{T}{N_1}\sigma_\mu^2\lim_{\mathbb{N}\to\infty}\frac{1}{N_1N_2T^2}\sum_{ijt}\sum_{t'\neq t}f_{(ijt)(ijt')}(\mathbf{x}_0,\mathbf{x}_0)\\
\to\ &\sigma_\zeta^2\bar{f}_{7,\zeta}(\mathbf{x}_0,\mathbf{x}_0)+r_{2,1}\sigma_v^2\bar{f}_{7,v}(\mathbf{x}_0,\mathbf{x}_0)+r_{T,1}\sigma_\mu^2\bar{f}_{7,\mu}(\mathbf{x}_0,\mathbf{x}_0).
\end{aligned}$$

Second, it is readily seen that

$$
\begin{aligned}
&(\mathbb{N}|\mathbf{H}|)^{-1}(N_1|\mathbf{H}|)^{-1}\mathrm{E}\left[\mathbf{H}^{-1}\left(\mathbf{X}-\mathbf{x}_0'\otimes\iota_{\mathbb{N}}\right)'\mathbf{K}_{\mathbf{H}}(\mathbf{x}_0)\Omega_u\mathbf{K}_{\mathbf{H}}(\mathbf{x}_0)\iota_{\mathbb{N}}\right]\\
&=(N_1|\mathbf{H}|)^{-1}\kappa_{22}\mathbb{N}^{-1}\mathbf{H}\sum_{ijt}\sigma_{ijt}^2(\mathbf{x}_0)f_{ijt}^{(1)}(\mathbf{x}_0)(1+o(1))\\
&\quad+\kappa_{12}(N_1|\mathbf{H}|)^{-1}\mathbb{N}^{-1}|\mathbf{H}|\sum_{i_1j_1t_1}\sum_{(ijt)_2\neq(ijt)_1}\sigma_{(ijt)_1(ijt)_2}(\mathbf{x}_0,\mathbf{x}_0)\times\\
&\quad\times\mathbf{H}\frac{\partial f_{(ijt)_1(ijt)_2}(\mathbf{x}_0,\mathbf{x}_0)}{\partial\mathbf{x}_{(ijt)_2}}(1+o(1))\\
&=O((N_1|\mathbf{H}|)^{-1}\|\mathbf{H}\|)+O(\|\mathbf{H}\|)=O(\|\mathbf{H}\|)
\end{aligned}
$$

and

$$
\begin{aligned}
&(\mathbb{N}|\mathbf{H}|)^{-1}(N_1|\mathbf{H}|)^{-1}\mathbf{H}^{-1}\mathrm{E}\left[\left(\mathbf{X}-\mathbf{x}_0'\otimes\iota_{\mathbb{N}}\right)'\mathbf{K}_{\mathbf{H}}(\mathbf{x}_0)\Omega_u\times\right.\\
&\left.\quad\times\mathbf{K}_{\mathbf{H}}(\mathbf{x}_0)\left(\mathbf{X}-\mathbf{x}_0{}'\otimes\iota_{\mathbb{N}}\right)\right]\mathbf{H}^{-1}\\
&=\kappa_{22}\mathbf{I}_k(N_1|\mathbf{H}|)^{-1}\mathbb{N}^{-1}\sum_{ijt}\sigma_{ijt}^2(\mathbf{x}_0)f_{ijt}(\mathbf{x}_0)(1+o(1))\\
&\quad+\kappa_{12}^2N_1^{-1}\mathbb{N}^{-1}\sum_{i_1j_1t_1}\sum_{(ijt)_2\neq(ijt)_1}\sigma_{(ijt)_1(ijt)_2}(\mathbf{x}_0,\mathbf{x}_0)\times\\
&\quad\times\mathbf{H}\frac{\partial^2 f_{(ijt)_1(ijt)_2}(\mathbf{x}_0,\mathbf{x}_0)}{\partial\mathbf{x}_{(ijt)_1}\partial\mathbf{x}'_{(ijt)_2}}\mathbf{H}(1+o(1))\\
&=O((N_1|\mathbf{H}|)^{-1})+O(\|\mathbf{H}\|^2).
\end{aligned}
$$

Under some suitable regularity conditions, we have the limiting distribution of the local linear estimator

$$
\begin{aligned}
&\sqrt{N_2T}\left(\widehat{m}(\mathbf{x}_0)-m(\mathbf{x}_0)-\kappa_{12}\mathrm{tr}\left\{\mathbf{H}\mathbf{m}^{(2)}(\mathbf{x}_0)\mathbf{H}\right\}/2\right) \qquad (7.38)\\
&\xrightarrow{d}\mathscr{N}\left(0,\frac{\sigma_\xi^2\bar{f}_{7,\zeta}(\mathbf{x}_0,\mathbf{x}_0)}{\bar{f}_7^2(\mathbf{x}_0)}+r_{2,1}\frac{\sigma_v^2\bar{f}_{7,v}(\mathbf{x}_0,\mathbf{x}_0)}{\bar{f}_7^2(\mathbf{x}_0)}+r_{T,1}\frac{\sigma_\mu^2\bar{f}_{7,\mu}(\mathbf{x}_0,\mathbf{x}_0)}{\bar{f}_7^2(\mathbf{x}_0)}\right)
\end{aligned}
$$

if $\|\mathbf{H}\|\to 0$, $\mathbb{N}|\mathbf{H}|\to\infty$, $N_1\geq N_2$, $N_1\geq T$, $N_2/N_1\to r_{2,1}\in[0,1]$, $T/N_1\to r_{T,1}\in[0,1]$, and $N_1|\mathbf{H}|\to\infty$ as $\min(N_1,N_2,T)\to\infty$. The convergence rate of the estimator is

$$\hat{m}(\mathbf{x}_0)-m(\mathbf{x}_0)=O_p(\|\mathbf{H}\|^2+(N_2T)^{-1/2}).$$

Now we discuss the optimal convergence rate of the pooled local linear estimator $\widehat{m}(\mathbf{x}_0)$. The bias and variance term of the estimator are of the order:

$$
\begin{aligned}
\text{Bias term}&\equiv\mathscr{A}^{-1}\mathscr{B}=O_p(\|\mathbf{H}\|^2)\\
\text{Variance term}&\equiv\mathscr{A}^{-1}\mathscr{C}=\begin{cases}O_p((\mathbb{N}|\mathbf{H}|)^{-1/2}),\ \text{if } N_1|\mathbf{H}|\to c\in[0,+\infty),\\ O_p((N_2T)^{-1/2}),\ \text{if } N_1|\mathbf{H}|\to+\infty.\end{cases}
\end{aligned}
$$

Similarly to case (iv), the variance term of estimator $\widehat{m}(\mathbf{x}_0)$ is never smaller than $O((N_2T)^{-1/2})$. The optimal bandwidth is set to balance the squared bias and variance term of the estimator. Similarly to the result for case (iv), the limit of the ratio,

$$\frac{N_1}{(N_2T)^{k/4}} \to \begin{cases} 0, \\ c \in (0,\infty), \\ \infty, \end{cases}$$

and the number of regressors k determine the optimal MSE rate of the pooled local linear estimator, its optimal bandwidth rate, and its asymptotic variance. Below, we discuss how the optimal MSE rate of the estimator and its corresponding optimal bandwidth rate are determined by N_1, N_2, T, and k. Without loss of generality, we assume that $\mathbf{H} = h\mathbf{I}_k$, $N_1 \geq N_2$, and $N_1 \geq T$.

- When $k = 1$, we have $N_1/(N_2T)^{1/4} \to \infty$. It can be shown that the optimal MSE rate of the pooled local linear estimator is always $O_e\left((N_2T)^{-1}\right)$. However, there is a range for h to achieve the optimal MSE rate. A simple choice for h is that $h_{opt} = O_e(N_1^{-1})$, we have $N_1|\mathbf{H}_{opt}| \to c > 0$ with the asymptotic variance of the estimator given in (7.36).
- When $k = 2$, the ratio $N_1/(N_2T)^{1/2}$ can approach to either ∞ or $c > 0$. It can be shown that the optimal MSE rate of local linear estimator is always $O_e\left((N_2T)^{-1}\right)$, which can be obtained by setting the rate of optimal bandwidth as follows:
  - If $N_1/(N_2T)^{1/2} \to \infty$ and the optimal bandwidth satisfies $h_{opt} = O_e((N_2T)^{-1/4})$, we have $N_1|\mathbf{H}_{opt}| \to \infty$ with the asymptotic variance of the estimator given in (7.38).
  - If $N_1/(N_2T)^{1/2} \to \infty$ and the optimal bandwidth satisfies $h_{opt} = O_e(N_1^{-1/2})$, we have $N_1|\mathbf{H}_{opt}| \to c > 0$ with the asymptotic variance of the estimator given in (7.36).
  - If $N_1/(N_2T)^{1/2} \to c \geq 1$ and the optimal bandwidth satisfies $h_{opt} = O_e(N_1^{-1/2})$, we have $N_1|\mathbf{H}_{opt}| \to c > 0$ with the asymptotic variance of the estimator given in (7.36).
- When $k \geq 3$, the ratio $N_1/N_2^{k/4}$ can approach to 0, a positive finite number c, or ∞. It can be shown that the optimal MSE rate of local the linear estimator is either $O((N_1N_2T)^{-4/(k+4)})$ or $O((N_2T)^{-1})$ if $N_1/(N_2T)^{k/4} \to 0$ or $N_1/(N_2T)^{k/4} \to c \in (0,\infty]$, respectively. More specifically,
  - If $N_1/(N_2T)^{k/4} \to 0$, the optimal MSE rate is $O((N_1N_2T)^{-4/(k+4)})$ with optimal bandwidth $h_{opt} = O_e((N_1N_2T)^{-1/(k+4)})$, which corresponds to $N_1|\mathbf{H}| \to 0$ with the asymptotic variance of the estimator given in (7.34).
  - If $N_1/(N_2T)^{k/4} \to c \in (0,\infty)$, the optimal MSE rate is $O((N_2T)^{-1})$ with optimal bandwidth $h_{opt} = O_e((N_2T)^{-1/4})$, which corresponds to $N_1|\mathbf{H}| \to c \in (0,\infty)$ with the asymptotic variance of the estimator given in (7.36).

– If $N_1/(N_2T)^{k/4} \to \infty$, the optimal MSE rate $O_e((N_2T)^{-1})$ with optimal bandwidth $h_{opt} = O_e((N_2T)^{-1/4})$, which corresponds to $N_1|\mathbf{H}| \to \infty$ with the asymptotic variance of the estimator given in (7.38).

Table 7.6 summarizes the asymptotic results of the pooled local linear estimators for case (vii) under five sets of different conditions on the sample indices and orders, which includes the scenario that N_1, N_2, or T is the largest sample index.

7.5 Some Extensions

For three-dimensional nonparametric panel data models with random effects, our proofs given in Sects. 7.2 and 7.4 indicate that the consistency and limiting distribution of the local linear estimator depend on the error structure, the bandwidth, and the relative sample sizes across different dimensions, and that the local linear estimator for the random-effects model can fail to be consistent due to the non-diminishing variance term. In this section, we briefly extend our review in three directions. In Sect. 7.5.1, we explain how to obtain consistent estimation when the local linear estimator fails to be consistent. In Sect. 7.5.2, we give a brief discussion of estimating four and higher-order nonparametric panel data models with random effects. Sect. 7.5.3 discusses nonparametric panel data models with fixed effects.

7.5.1 Mixed Fixed and Random Effects Models

For the pairwise error structure considered in Sect. 7.4, the local linear estimator is inconsistent when the sample size increases in only one index. For this case, we suggest readers should estimate model (7.1) with (7.24) as a mixed fixed and random effects model.

For the purpose of illustration, we consider the case that $N_1 \to \infty$ and both N_2 and T are finite. Due to the existence of index i-invariant random effects, ζ_{jt}, the local linear estimator is inconsistent, as shown in Sect. 7.4.1. Rewriting model (7.1) gives

$$y_{ijt} = \zeta_{jt} + m(\mathbf{x}_{ijt}) + \upsilon_{it},\ \upsilon_{it} = \mu_{ij} + v_{it} + \varepsilon_{ijt}\ , \tag{7.39}$$

where we will estimate $m(\cdot)$ as well as $\{\zeta_{jt} : j = 1,\ldots,N_2, t = 1,\ldots,T\}$, and treat υ_{it} as the composite error. Then, model (7.39) becomes a mixed fixed and random effects model.

For two-dimensional panel data with a large number of cross sectional units and finite time periods, two different methods are used to estimate the unknown function $m(\cdot)$ from a nonparametric fixed effects panel data model. One method is to cancel out the unobserved time-invariant fixed effects before the estimation of $m(\cdot)$; see, e.g., Henderson et al. (2008), Qian and Wang (2012), and Rodriguez-Poo and Soberon (2015). The other method is to treat the unobserved fixed effects as parame-

Table 7.6 Asymptotics of the estimators for case (vii): $N_1, N_2, T \to \infty$

Condition	Bias term rate	Variance term rate	Asymptotic variance
$\|\mathbf{H}\| \to 0$ $\mathbb{N}\|\mathbf{H}\| \to \infty$ [a] $\max(N_1, N_2, T)\|\mathbf{H}\| \to 0$	$O_p(\|\mathbf{H}\|^2)$	$O_p((\mathbb{N}\|\mathbf{H}\|)^{-1/2})$	$\nu_0 \frac{\sigma_\mu^2+\sigma_v^2+\sigma_\zeta^2+\sigma_\varepsilon^2}{\bar{f}_7(x_0)}$ [b]
$\|\mathbf{H}\| \to 0$ $\mathbb{N}\|\mathbf{H}\| \to \infty$ $N_1\|\mathbf{H}\| \to c_1 \in [0,\infty)$ $N_2\|\mathbf{H}\| \to c_2 \in [0,\infty)$ $T\|\mathbf{H}\| \to c_T \in [0,\infty)$	$O_p(\|\mathbf{H}\|^2)$	$O_p((\mathbb{N}\|\mathbf{H}\|)^{-1/2})$	$\nu_0 \frac{\sigma_\mu^2+\sigma_v^2+\sigma_\zeta^2+\sigma_\varepsilon^2}{\bar{f}_7(x_0)}$ $+c_1 \frac{\sigma_\zeta^2 \bar{f}_{7,\zeta}(x_0,x_0)}{\bar{f}_7^2(x_0)}$ [c] $+c_2 \frac{\sigma_v^2 \bar{f}_{7,v}(x_0,x_0)}{\bar{f}_7^2(x_0)}$ [d] $+c_T \frac{\sigma_\mu^2 \bar{f}_{7,\mu}(x_0,x_0)}{\bar{f}_7^2(x_0)}$ [e]
$\|\mathbf{H}\| \to 0$ $\mathbb{N}\|\mathbf{H}\| \to \infty$ $N_1 \geq N_2, N_1 \geq T$ $N_2/N_1 \to r_{2,1} \in [0,1]$ $T/N_1 \to r_{T,1} \in [0,1]$ $N_1\|\mathbf{H}\| \to \infty$	$O_p(\|\mathbf{H}\|^2)$	$O_p((N_2T)^{-1/2})$	$\frac{\sigma_\zeta^2 \bar{f}_{7,\zeta}(x_0,x_0)}{\bar{f}_7^2(x_0)}$ $+r_{2,1} \frac{\sigma_v^2 \bar{f}_{7,v}(x_0,x_0)}{\bar{f}_7^2(x_0)}$ $+r_{T,1} \frac{\sigma_\mu^2 \bar{f}_{7,\mu}(x_0,x_0)}{\bar{f}_7^2(x_0)}$
$\|\mathbf{H}\| \to 0$ $\mathbb{N}\|\mathbf{H}\| \to \infty$ $N_2 \geq N_1, N_2 \geq T$ $N_1/N_2 \to r_{1,2} \in [0,1]$ $T/N_2 \to r_{T,2} \in [0,1]$ $N_2\|\mathbf{H}\| \to \infty$	$O_p(\|\mathbf{H}\|^2)$	$O_p((N_1T)^{-1/2})$	$r_{1,2} \frac{\sigma_\zeta^2 \bar{f}_{7,\zeta}(x_0,x_0)}{\bar{f}_7^2(x_0)}$ $+\frac{\sigma_v^2 \bar{f}_{7,v}(x_0,x_0)}{\bar{f}_7^2(x_0)}$ $+r_{T,2} \frac{\sigma_\mu^2 \bar{f}_{7,\mu}(x_0,x_0)}{\bar{f}_7^2(x_0)}$
$\|\mathbf{H}\| \to 0$ $\mathbb{N}\|\mathbf{H}\| \to \infty$ $T \geq N_1, T \geq N_2$ $N_1/T \to r_{1,T} \in [0,1]$ $N_2/T \to r_{2,T} \in [0,1]$ $T\|\mathbf{H}\| \to \infty$	$O_p(\|\mathbf{H}\|^2)$	$O_p((N_1N_2)^{-1/2})$	$r_{1,T} \frac{\sigma_\zeta^2 \bar{f}_{7,\zeta}(x_0,x_0)}{\bar{f}_7^2(x_0)}$ $+r_{2,T} \frac{\sigma_v^2 \bar{f}_{7,v}(x_0,x_0)}{\bar{f}_7^2(x_0)}$ $+\frac{\sigma_\mu^2 \bar{f}_{7,\mu}(x_0,x_0)}{\bar{f}_7^2(x_0)}$

[a] $\mathbb{N} = N_1N_2T$.
[b] $\bar{f}_7(x_0) = \lim_{\mathbb{N}\to\infty} \mathbb{N}^{-1} \sum_{i,j,t} f_{ijt}(x_0)$.
[c] $\bar{f}_{7,\zeta}(x_0,x_0) = \lim_{\mathbb{N}\to\infty} \frac{1}{N_1\mathbb{N}} \sum_{i,j,t} \sum_{i'\neq i} f_{(i,j,t),(i',j,t)}(x_0,x_0)$.
[d] $\bar{f}_{7,v}(x_0,x_0) = \lim_{\mathbb{N}\to\infty} \frac{1}{N_2\mathbb{N}} \sum_{i,j,t} \sum_{j'\neq j} f_{(i,j,t),(i,j',t)}(x_0,x_0)$.
[e] $\bar{f}_{7,\mu}(x_0,x_0) = \lim_{\mathbb{N}\to\infty} \frac{1}{T\mathbb{N}} \sum_{i,j,t} \sum_{t'\neq t} f_{(i,j,t),(i,j,t')}(x_0,x_0)$.

ters to be estimated; see, e.g., Su and Ullah (2006), Sun et al. (2009), and Chen et al. (2013). Su and Ullah (2006) and Chen et al. (2013) assumed that the unobserved

fixed effects add up to zero for identification purposes, while Sun et al. (2009) assumed that the unobserved fixed effects are i.i.d. with zero mean and finite variance. For model (7.39) above, ζ_{jt} is i.i.d. with zero mean and finite variance, so Sun et al. (2009) setup is proper. Intuitively, applying the nonparametric least squares dummy variable approach proposed in Sun et al. (2009), we can show the consistency of the local linear estimator for $m(\cdot)$ as condition (7.15) holds true:

$$\begin{aligned} & N_1^{-1}|\mathbf{H}| \sum_{i_1 j_1 t_1} \sum_{(ijt)_2 \neq (ijt)_1} \left| \mathrm{E}\left(\upsilon_{i_1 j_1 t_1} \upsilon_{i_2 j_2 t_2}\right) \right| \\ = & |\mathbf{H}| \left[\sigma_\mu^2 N_2 T(T-1) + \sigma_v^2 N_2 (N_2-1) T\right] = O(|\mathbf{H}|). \end{aligned}$$

7.5.2 Four and Higher-dimensional Cases

We define an s-dimensional panel data model with random effects as follows

$$y_{\mathbf{i}} = m(\mathbf{x}_{\mathbf{i}}) + u_{\mathbf{i}}\,, \tag{7.40}$$

where $\mathbf{i} = (i_1, i_2, \ldots, i_s)$ with $1 \leq i_j \leq N_j$ for all j and the total sample size is $\mathbb{N} = N_1 \cdot N_2 \cdots N_s$, $\mathbf{x}_{\mathbf{i}}$ is a $(k \times 1)$ strictly exogenous vector of continuous variables, both $y_{\mathbf{i}}$ and $u_{\mathbf{i}}$ are scalar, $u_{\mathbf{i}}$ is a random error with zero mean, and $m(x) = E(y_{\mathbf{i}} | \mathbf{x}_{\mathbf{i}} = x)$ is a smooth unknown function to be estimated. As in Sect. 7.2, without knowing the specific structure of the error term and the data generating mechanism of the regressor, one can derive the consistency of the local linear estimator of $m(\cdot)$ by arguing that both $\{\mathbf{x}_i\}$ and $\{u_{\mathbf{i}}\}$ are well-behaved weakly dependent random fields across all the indices with large samples. For example, the near-epoch dependence of a random field is a valid measure of weak dependence for high-dimensional data as defined in Jenish (2012). We expect that the argument on the more efficient two-step estimator given in Sect. 7.3 is also applicable for model (7.40) with $s \geq 4$. However, the concept of weakly dependent random fields can be too abstract to compete with the pairwise error structure and/or its nested special cases in economic fields. If this is the case, the consistency and the limiting distribution of the local linear estimator will depend on the dimensionality of the panel data.

Chap. 2 in this handbook gives a brief discussion for a four-dimensional parametric panel data model with random effects, where the authors assume a pairwise error structure $u_{ijst} = \mu_{ijs} + v_{ist} + \zeta_{jst} + \lambda_{ijt} + \varepsilon_{ijst}$, where μ_{ijs},v_{ist}, ζ_{jst}, λ_{ijt}, and ε_{ijst} are mutually uncorrelated and self-uncorrelated across any indices with zero mean and variances equal to σ_μ^2, σ_v^2, σ_ζ^2, σ_λ^2, and σ_ε^2, respectively. For this error structure, the covariance matrix becomes

$$\begin{aligned} \Omega_u = & \sigma_\mu^2 \left(\mathbf{I}_{N_1 N_2 N_3} \otimes \mathbf{J}_T\right) + \sigma_v^2 \left(\mathbf{I}_{N_1} \otimes \mathbf{J}_{N_2} \otimes \mathbf{I}_{N_3 T}\right) \\ & + \sigma_\zeta^2 \left(\mathbf{J}_{N_1} \otimes \mathbf{I}_{N_2 N_3 T}\right) + \sigma_\lambda^2 \left(\mathbf{I}_{N_1 N_2} \otimes \mathbf{J}_{N_3} \otimes \mathbf{I}_T\right) + \sigma_\varepsilon^2 \mathbf{I}_{\mathbb{N}}\,, \end{aligned}$$

where $\mathbb{N} = N_1 N_2 N_3 T$ is the sample size. In parallel to (7.26), we have

$$\mathbb{N}^{-1}|\mathbf{H}|\sum_{i_1 j_1 s_1 t_1}\sum_{(ijst)_2\neq(ijst)_1}\left|\mathrm{E}\left(u_{i_1 j_1 s_1 t_1}u_{i_2 j_2 s_2 t_2}\right)\right|$$
$$=|\mathbf{H}|\left[\sigma_\mu^2(T-1)+\sigma_v^2(N_2-1)+\sigma_\zeta^2(N_1-1)+\sigma_\lambda^2(N_3-1)\right]. \quad (7.41)$$

Again, we can see that (7.41) can be explosive if only one index increases with the sample size, so that the local linear estimator becomes inconsistent and the mixed fixed and random effects modelling approach explained in Sect. 7.5.1 should be applied. When two or more indices increase with the sample size, the local linear estimator is expected to be consistent and approximately normally distributed at the conventional convergence rate $O_p\left(\|\mathbf{H}\|^2+\sqrt{\mathbb{N}|\mathbf{H}|}\right)$ if the term in (7.41) converges to zero as the sample size increases. If the term in (7.41) converges to a positive constant or is explosive as the sample size increases, one has to look into the rate of the asymptotic variance term case by case; i.e., only two indices grow, or only three indices grow, or all four indices grow. We will forgo the detailed results to save space here.

To sum up, for the s-dimension panel data model (7.40) with a pairwise error structure, $u_{\mathbf{i}}=\mu_{\mathbf{1}}+\ldots+\mu_{\mathbf{s}}+\varepsilon_{\mathbf{i}}$, where the idiosyncratic error, ε_i, and the s error terms $\mu_{\mathbf{j}}$ are mutually uncorrelated and self-uncorrelated with zero mean and variance equal to σ_ε^2 and σ_j^2, respectively, and removing j from $\mathbf{s}=(1,2,\ldots,s)$ gives $\mathbf{j}$. Then, (7.41) becomes

$$\mathbb{N}^{-1}|\mathbf{H}|\sum_{\mathbf{i}_1}\sum_{\mathbf{i}_2\neq\mathbf{i}_1}\left|\mathrm{E}\left(u_{\mathbf{i}_1}u_{\mathbf{i}_2}\right)\right|=|\mathbf{H}|\sum_{l=1}^{s-1}\sigma_l^2(N_l-1)\ .$$

Again, the mixed fixed and random effects modelling approach is recommended if $s-1$ indices are fixed and only one index increases with the sample size. If two or more indices increase with sample size, the local linear estimator is always consistent but its limiting distribution varies with respect to the dimension of the regressors, the relative sample size across different indices and the bandwidth.

7.5.3 Fixed Effects Models

Consider a nonparametric fixed effects panel data model in matrix form,

$$\mathbf{y}=m(\mathbf{X})+\mathbf{D}\mu+\mathbf{u}, \quad (7.42)$$

where μ contains all unobserved fixed effects and $\mathbf{D}$ is the corresponding dummy variable matrix, $\mathbf{u}$ is an $(\mathbb{N}\times 1)$ vector of i.i.d. errors. If $m(\mathbf{X})=\mathbf{X}\theta$ is known up to a finite number of unknown parameters θ, (7.42) becomes the parametric fixed effects model considered in Chap. 1 in this volume, where the authors explain that the traditional least squares dummy variable (or LSDV) estimator continues to be an effective estimator as long as the modified dummy variables, $\mathbf{D}$, and re-

gressors, $\mathbf{X}$, are linearly independent. Let $M_{\mathbf{D}} = \mathbf{I}_{\mathbb{N}} - \mathbf{D}\left(\mathbf{D}^T\mathbf{D}\right)^{-1}\mathbf{D}^T$ and the LSDV estimator is given by $\hat{\theta} = \left(\mathbf{X}^T M_D \mathbf{X}\right)^{-1}\mathbf{X}^T M_D \mathbf{y}$, where the LSDV estimator can essentially be applied to two and higher-dimension fixed effects panel data models. However, without concrete proofs, we are not sure whether the nonparametric estimation methods listed in Sect. 7.5.1 for two-dimensional panel data models will work equally well for higher-dimensional panel data models.

Take the first estimation method, for example, where one cancels out the unobserved fixed effects first before applying a nonparametric kernel estimation. Say that we multiply with $M_{\mathbf{D}}$ both sides of model (7.42), so that $M_{\mathbf{D}}\mathbf{y} = M_{\mathbf{D}} m(\mathbf{X}) + M_{\mathbf{D}}\mathbf{u}$, where the typical element of $M_{\mathbf{D}} m(\mathbf{X})$ equals a linear combination of $m(\mathbf{x_i})$ for all $\mathbf{i}$, where $\mathbf{i}$ is defined in the same way as in Sect. 7.5.2. One direct consequence of this method is that $m(\cdot)$ is not fully identified if $m(\mathbf{X}) = m_1(\mathbf{X}) + m_2(\mathbf{X})$ and $M_{\mathbf{D}} m_2(\mathbf{X}) = 0$ for some non zero function $m_1(\cdot)$. In addition, the additive structure resulting from this method requires the usage of a backfitting algorithm or a marginal integration method. It is well known that the performance of the backfitting algorithm and the marginal integration method will deteriorate for correlated regressors. Therefore, it is possible that this estimation method can perform worse for higher-dimensional panel data models with fixed effects. On the other hand, the limiting distribution of the nonparametric LSDV estimation method can be affected by the relative sample size in each index or dimension, and we leave the theoretical work to our future research.

On the last point, for two-dimensional parametric panel data models, the LSDV estimator is asymptotically equivalent to the random effects estimator if the sample size is large in both dimensions. It will be interesting to check whether this result continues to hold for panel data models with three or more dimensions.

7.6 Conclusion

This chapter proposes a pooled local linear estimator for a three-dimensional nonparametric panel data model with random effects. Our results indicate that the pooled local linear estimator can be inconsistent when the unobserved random effects exhibit serial and/or cross dependence to the magnitude that the sum of all the error term covariances in absolute values is explosive to infinity too quickly. In addition, when the pooled local linear estimator is consistent, the optimal convergence rate of the estimator, its corresponding optimal bandwidth and asymptotic variance depend on the number of regressors and the limit of certain sample indices ratio. Therefore, one needs to examine the consistency of the pooled local linear estimator and its asymptotic properties case by case. Furthermore, when the pooled local linear estimator is consistent, we propose a two-step estimator along the lines of Su et al. (2013) and show that this estimator is asymptotically more efficient than the pooled local linear estimator.

Acknowledgements

Wei Lin's research was partially supported by National Science Foundation of China grants No. 71501134 (Young Scientists Fund Project) and No. 71671183 (General Program Project).

References

Cai, Z., Fan, J., and Yao, Q. (2000). Functional-coefficient regression models for nonlinear time series. *Journal of the American Statistical Association*, 95:941–956.

Chen, J., Gao, J., and D., L. (2013). Estimation in partially linear single-index panel data models with fixed effects. *Journal of Business & Economics Statistics*, 31:315–330.

Dedecker, J. (1998). A central limit theorem for stationary random fields. *Probability Theory Related Fields*, 110:397–426.

El Machkouri, M., Volný, D., and Wu, W. B. (2013). A central limit theorem for stationary random fields. *Stochastic Processes and their Applications*, 123:1–14.

Henderson, D. J., Carroll, R. J., and Li, Q. (2008). Nonparametric estimation and testing of fixed effects panel data models. *Journal of Econometrics*, 144:257–275.

Henderson, D. J. and Ullah, A. (2005). A nonparametric random effects estimator. *Economics Letters*, 88:403–407.

Jenish, N. (2012). Nonparametric spatial regression under near-epoch dependence. *Journal of Econometrics*, 167:224–239.

Jenish, N. and Prucha, I. R. (2009). Central limit theorems and uniform laws of large numbers for arrays of random fields. *Journal of Econometrics*, 150:86–98.

Lin, X. and Carroll, R. J. (2000). Nonparametric function estimation for clustered data when the predictor is measured without/with error. *Journal of the American Statistical Association*, 95:520–534.

Martins-Filho, C. and Yao, F. (2009). Nonparametric regression estimation with general parametric error covariance. *Journal of Multivariate Analysis*, 100:309–333.

Masry, E. (1996). Multivariate local polynomial regression for time series: Uniform strong consistency and rates. *Journal of Time Series Analysis*, 17:571–599.

Qian, J. and Wang, L. (2012). Estimating semiparametric panel data models by marginal integration. *Journal of Econometrics*, 167:483–493.

Robinson, P. (2012). Nonparametric trending regression with cross-sectional dependence. *Journal of Econometrics*, 169:4–14.

Rodriguez-Poo, J. M. and Soberon, A. (2015). Nonparametric estimation of fixed effects panel data varying coefficient models. *Journal of Multivariate Analysis*, 133:95–122.

Ruckstuhl, A. F., Welsh, A. H., and Carroll, R. J. (2000). Nonparametric function estimation of the relationship between two repeatedly measured variables. *Statistica Sinica*, 10:51–71.

Su, L. and Ullah, A. (2006). Profile likelihood estimation of partially linear panel data models with fixed effects. *Economics Letters*, 92:75–81.

Su, L. and Ullah, A. (2007). More efficient estimation of nonparametric panel data models with random effects. *Economics Letters*, 96:375–380.

Su, L., Ullah, A., and Wang, Y. (2013). Nonparametric regression estimation with general parametric error covariance: A more efficient two-step estimator. *Empirical Eocnomics*, 45:1009–1024.

Sun, Y., Carroll, R., and Li, D. (2009). Semiparametric estimation of fixed effects panel data varying coefficient models. *Advances in Econometrics*, 25:101–129.

Sun, Y., Zhang, Y., and Li, Q. (2015). Nonparametric panel data regression models. In Baltagi, B. H., editor, *Oxford Handbook on Panel Data*, chapter 10, pages 285–324. Oxford University Press.

Tran, L. T. (1990). Kernel density estimation on random fields. *Journal of Multivariate Analysis*, 34:37–53.

Yao, W. and Li, R. (2013). New local estimation procedure for a non-parametric regression for longitudinal data. *Journal of the Royal Statistical Society (Series B)*, 75:123–138.

Appendix

In this section, we provide a sketchy proof for the results of the optimal MSE rate of the pooled local linear estimator with $N_1 \to \infty$, $N_2 \to \infty$ and T fixed, which are stated in Sect. 7.4.2. Without loss of generality, we set $\mathbf{H} = h\mathbf{I}_k$ and $N_1 \geq N_2$.

First, we consider the case that $1 \leq k \leq 3$:

- If we set $\mathbf{H}$ satisfying $N_1|\mathbf{H}| \to \infty$ as $N_2 \to \infty$, $\text{MSE}(\hat{m}(\mathbf{x}_0)) = O_p\left(\|\mathbf{H}\|^4 + N_2^{-1}\right)$. Letting $h \asymp N_2^{-(1+\varepsilon)/4}$ for some small $\varepsilon \in (0, 4/k-1)$, we have $N_2\|\mathbf{H}\|^4 \asymp N_2^{-\varepsilon} \to 0$ while $N_1|\mathbf{H}| \asymp N_1 N_2^{-k(1+\varepsilon)/4} \geq N_2^{1-k(1+\varepsilon)/4} \to \infty$. Hence,

$$\text{MSE}(\hat{m}(\mathbf{x}_0)) = O_e\left(N_2^{-1}\right)$$

  for any $\mathbf{H}$ satisfying $N_1|\mathbf{H}| \to \infty$ as $N_2 \to \infty$.
- If we set $\mathbf{H}$ satisfying $N_1|\mathbf{H}| \to 0$ as $N_2 \to \infty$, we see $N_1 N_2|\mathbf{H}| = o(N_2)$ and $\|\mathbf{H}\|^4 = o\left(N_2^{-1}\right)$ as $N_2 h^4 \leq N_2 h^k \leq N_1 h^k \to 0$. Hence,

$$\text{MSE}(\hat{m}(\mathbf{x}_0)) = O_p\left((N_1 N_2|\mathbf{H}|)^{-1}\right)$$

  is always larger when $N_1|\mathbf{H}| \to 0$ than when $N_1|\mathbf{H}| \to \infty$ and the optimal bandwidth $\mathbf{H}_{opt}$ must not satisfy $N_1|\mathbf{H}_{opt}| \to 0$. It cannot be set that $h_{opt} = c(N_1 N_2)^{\frac{-1}{k+4}}$

since $N_1|\mathbf{H}_{opt}| = c^k \left(N_1/N_2^{k/4}\right)^{4/(k+4)} \geq c^k N_2^{(4-k)/(k+4)} \to \infty$ which contradicts our assumption $N_1|\mathbf{H}_{opt}| \to 0$.

- If we set $\mathbf{H}$ satisfying $N_1|\mathbf{H}| \to c > 0$ as $N_2 \to \infty$, we have $N_1 N_2|\mathbf{H}| = O(N_2)$ and $h = c_* N_1^{-1/k}$ for some finite $c_* > 0$. Hence, $\text{MSE}(\hat{m}(\mathbf{x}_0)) = O_p\left(N_1^{-4/k} + N_2^{-1}\right) = O_e\left(N_2^{-1}\right)$ as $N_1^{4/k} N_2^{-1} \geq N_2^{4/k-1} \to \infty$. Without knowing which is the bigger between $C_4(\mathbf{x}_0)$ and $C_{4,3}(\mathbf{x}_0)$, we cannot compare $\text{MSE}(\hat{m}(\mathbf{x}_0))$ under $N_1|\mathbf{H}| \to c > 0$ and $N_1|\mathbf{H}| \to \infty$.

To sum up, minimizing the MSE of $\hat{m}(\mathbf{x}_0)$ depends on the values of N_1, N_2 and k. Specifically, the minimum of $\text{MSE}(\hat{m}(\mathbf{x}_0)) = O_e\left(N_2^{-1}\right)$ with $\mathbf{H}_{opt}$ satisfying $N_1|\mathbf{H}_{opt}| \to \infty$ or $N_1|\mathbf{H}_{opt}| \to c > 0$. Moreover, it is interesting to observe that the optimal mean squared error of the pooled local linear estimator is always bounded by $O_e\left(N_2^{-1}\right)$ for the three-dimensional panel data model with random effects when two sample indices approach to infinity. The pooled local linear estimator does not suffer the curse-of-dimensionality problem, which means that the asymptotic variance term vanishes at the same speed for $k \in \{1, 2, 3\}$.

Next, we consider the case that $k = 4$:

- If $N_1|\mathbf{H}| \to c > 0$ as $N_2 \to \infty$, we have $h \asymp N_1^{-1/4}$, so that $\text{MSE}(\hat{m}(\mathbf{x}_0)) = O_p\left(N_1^{-1} + N_2^{-1}\right) = O_e\left(N_2^{-1}\right)$.
- If $N_1|\mathbf{H}| \to 0$ as $N_2 \to \infty$, $N_1 N_2|\mathbf{H}| = o(N_2)$ and $\|\mathbf{H}\|^4 = o\left(N_2^{-1}\right)$ as $N_2 h^4 \leq N_1 h^4 \to 0$. Hence, $\text{MSE}(\hat{m}(\mathbf{x}_0)) = O_p\left((N_1 N_2|\mathbf{H}|)^{-1}\right)$ is always larger when $N_1|\mathbf{H}| \to 0$ than when $N_1|\mathbf{H}| \to c > 0$ and the optimal bandwidth $\mathbf{H}_{opt}$ must not satisfy $N_1|\mathbf{H}_{opt}| \to 0$.
- If $N_1|\mathbf{H}| \to \infty$ as $N_2 \to \infty$, $\text{MSE}(\hat{m}(\mathbf{x}_0)) = O_p\left(\|\mathbf{H}\|^4 + N_2^{-1}\right)$.

  (i) If $N_1/N_2 \to \infty$, setting $h \asymp N_2^{-1/4}$, we have $\text{MSE}(\hat{m}(\mathbf{x}_0)) = O_e\left(N_2^{-1}\right)$ and $N_1|\mathbf{H}| \asymp N_1 N_2^{-1} \to \infty$.

  (ii) If $N_1/N_2 \to c \geq 1$ for some finite c, we have $N_2\|\mathbf{H}\|^4 \to \infty$ and $\text{MSE}(\hat{m}(\mathbf{x}_0)) = O_e\left(\|\mathbf{H}\|^4\right)$ is larger than $O_p\left(N_2^{-1}\right)$.

To sum up, the optimal $\text{MSE}(\hat{m}(\mathbf{x}_0)) = O_e\left(N_2^{-1}\right)$. If $N_1/N_2 \to \infty$, the optimal bandwidth satisfies either $N_1|\mathbf{H}_{opt}| \to c > 0$ or $N_1|\mathbf{H}_{opt}| \to \infty$ depending on the relative size of $C_4(\mathbf{x}_0)$ and $C_{4,3}(\mathbf{x}_0)$; if $N_1/N_2 \to c \geq 1$ holds, the optimal bandwidth satisfies $N_1|\mathbf{H}_{opt}| \to c > 0$ or $h_{opt} \asymp N_1^{-1/4}$.

Finally, we consider the case that $k \geq 5$:

- If $N_1|\mathbf{H}| \to \infty$, $\text{MSE}(\hat{m}(\mathbf{x}_0)) = O_p\left(N_2^{-1}\right)$ if $\|\mathbf{H}\|^4 N_2 \to c \geq 0$ and $\text{MSE}(\hat{m}(\mathbf{x}_0)) = O_p\left(\|\mathbf{H}\|^4\right)$ if $\|\mathbf{H}\|^4 N_2 \to \infty$. If $N_1|\mathbf{H}| \to 0$,

$$\text{MSE}(\hat{m}(\mathbf{x}_0)) = O_p\left(\|\mathbf{H}\|^4 + (N_1 N_2|\mathbf{H}|)^{-1}\right).$$

If $N_1|\mathbf{H}| \to c > 0$, we have $\text{MSE}(\hat{m}(\mathbf{x}_0)) = O_p\left(N_2^{-1}\right)$ if $N_2^{k/4}/N_1 \to c \geq 0$ and $\text{MSE}(\hat{m}(\mathbf{x}_0)) = O_e\left(N_1^{-4/k}\right)$ if $N_2^{k/4}/N_1 \to \infty$.

- Assume that $N_2^{k/4}/N_1 \to \infty$.

  (i) Setting $h \asymp (N_1N_2)^{-1/(k+4)}$ gives $N_1|\mathbf{H}| \to 0$ and

$$\text{MSE}(\hat{m}(\mathbf{x}_0)) = O_p\left((N_1N_2)^{-4/(k+4)}\right).$$

  (ii) If $\|\mathbf{H}\|^4 N_2 \to c \geq 0$, we have $N_1|\mathbf{H}| = O\left(N_1N_2^{-4/k}\right) = o(1)$, so if $N_1|\mathbf{H}| \to \infty$, we have

$$\|\mathbf{H}\|^4 N_2 \to \infty \quad \text{and} \quad (N_1N_2)^{4/(k+4)}\|\mathbf{H}\|^4 = \left(N_1N_2h^{k+4}\right)^{4/(k+4)} \to \infty.$$

  (iii) If $N_1|\mathbf{H}| \to c > 0$,

$$N_1^{-4/k}(N_1N_2)^{4/(k+4)} = \left(N_2^{k/4}/N_1\right)^{16/[k(k+4)]} \to \infty.$$

  Therefore, the optimal $\text{MSE}(\hat{m}(\mathbf{x}_0)) = O_p\left((N_1N_2)^{-4/(k+4)}\right)$ with the optimal bandwidth $h_{opt} \asymp (N_1N_2)^{-1/(k+4)}$.

- Assume that $N_2^{k/4}/N_1 \to c \geq 0$. We have $\text{MSE}(\hat{m}(\mathbf{x}_0)) = O_p\left(N_2^{-1}\right)$ if $\mathbf{H}$ satisfies $h \asymp N_2^{-1/4}$ and $N_1|\mathbf{H}| \to \infty$ or if $N_1|\mathbf{H}| \to c > 0$.

Hence, the optimal MSE of $\hat{m}(\mathbf{x}_0)$ is of order $O_p\left((N_1N_2)^{-4/(k+4)}\right)$ if $N_2^{k/4}/N_1 \to \infty$ and is of order $O_p\left(N_2^{-1}\right)$ if $N_2^{k/4}/N_1 \to c \in [0,\infty)$.

Chapter 8
Multi-dimensional Panels in Quantile Regression Models

Antonio F. Galvao and Gabriel V. Montes-Rojas

Abstract This chapter studies estimation and inference methods for multi-dimensional quantile regression panel data models. First, we discuss the fixed effects (FE) model. This model imposes a relatively restrictive asymptotic condition on the growth of the time series dimension relative to the cross section dimension. Nevertheless, extending the FE to three or more dimensions allows for larger data availability, and might help to relax the stringent condition on the time series. We also present a model for the smoothed FE quantile regression case. Second, we present a random effects (RE) model. This model has the advantage of allowing for small time-series dimension. Finally, we present a correlated RE model. In this case, the unobservable individual-specific effects are modeled as a function of observables and a disturbance.

8.1 Introduction

Standard panel data consisting of observations across time for different individuals allow the possibility of controlling for unobserved individual heterogeneity. Such heterogeneity can be an important phenomenon, and failure to control for it may result in misleading inference. This problem is particularly severe when the unobserved heterogeneity is correlated with explanatory variables. Recently, rich panel data sets have become widely available and very popular. They provide a large number of data points, allow analysis of the dynamics of adjustment, as well as control for individual specific heterogeneity.

Quantile regression (QR) models have provided a valuable tool in economics and statistics as a way of capturing heterogeneous effects that covariates may have on the

Antonio F. Galvao
University of Iowa, Iowa, USA

Gabriel V. Montes-Rojas
Universitat Autònoma de Barcelona, Barcelona, Spain

L. Matyas (ed.), *The Econometrics of Multi-dimensional Panels*, Advanced Studies in Theoretical and Applied Econometrics 50, DOI 10.1007/978-3-319-60783-2_8

outcome of interest, exposing a wide variety of forms of conditional heterogeneity under weak distributional assumptions. Koenker (2004) introduced a general approach to the estimation of QR for panel data models. The panel QR has attracted considerable interest in both the theoretical and applied literatures. It allows one to explore a range of conditional quantiles, thereby exposing a variety of forms of conditional heterogeneity, and to control for unobserved individual-specific effects. Controlling for individual heterogeneity, while exploring heterogeneous covariate effects within the QR framework, offers a more flexible approach to the analysis of panel data than that afforded by the classical fixed and random effects mean-based estimation. QR panel data models are able to capture these two types of heterogeneity in a single framework.

The extension of the two-dimension to the three or higher-dimensional panel data framework has implications for modeling, estimation and inference of conditional quantile models. In turn, these depends on the nature of the multi-dimensional setting, i.e., nested or non-nested, and the type of estimation and inference analyses to be implemented. Moreover, whether the individual-specific effects are correlated or not with the covariates, and at which level, are important elements for the analysis of panel data QR. This chapter studies panel data QR models for multi-dimensional panels and looks into three-dimensional (and higher-dimensional) settings. We concentrate on the standard linear models and discuss the multi-dimensional panels for both fixed and random effects models, and also for correlated random-effects models.

First, we discuss the fixed effects (FE) QR model. The FE-QR allows for individual-specific effects in which no parametric assumptions on the relationship between the specific effects and the covariates are made. Unfortunately, the standard FE-QR estimator is subject to the incidental parameters problem. In addition, there is no general transformation that can suitably eliminate the specific effects in the QR model. Thus, it has been customary to impose a relatively restrictive condition on the growth of the time series dimension relative to the cross section dimension. Nevertheless, extending the standard FE to three dimensions allows for larger data availability, and might help to relax the stringent condition on the time series. In this case, it is even possible that the time dimension is fixed, while the other two dimensions satisfy alternative requirements for asymptotic analysis.

Second, we present the random effects (RE) QR model. The RE-QR model imposes that the specific components are independent of the regressors. In spite of this restriction, in the RE-QR model the unobserved specific effects affect the unobservable variable, which induces heterogeneity across the conditional quantile function of the dependent variable. In addition, the RE-QR has the advantage of allowing for time-invariant regressors, and allows the time-series dimension to be small and fixed.

Finally, we briefly consider the correlated random effects (CRE) QR model. In this case, the unobservable individual-specific effects are modeled as a function of observables and a disturbance. In addition, we will suggest specific guidelines for practitioners in applied work.

The chapter is organized as follows. Section 8.2 describes the fixed effects models considering individual-specific heterogeneity in multi-dimensional panels and analyzes these panel data structures. Section 8.3 studies random-effects models, while Sect. 8.4 explores correlated random-effects frameworks. Finally, Sect. 8.5 summarizes some specific issues for practitioners.

8.2 Fixed Effects Models

In this section we consider a multi-dimensional FE-QR model. In particular, we present a three-dimensional panel data set where the variables of the model are observed along three indices given by the index set (i, j, t) where $i \in \{1, ..., N_1\}$, $j \in \{1, ..., N_2\}$ and $t \in \{1, ..., T\}$, respectively. A FE-QR model with individual-specific and time-specific effects can be written as

$$Q_\tau(y_{ijt}|x_{ijt}, \alpha_i, \gamma_j, \lambda_t) = x_{ijt}\beta(\tau) + \alpha_i(\tau) + \gamma_j(\tau) + \lambda_t(\tau), \tag{8.1}$$

where y_{ijt} is a dependent variable, x_{ijt} is a p-dimensional vector of explanatory variables, α_i and γ_j are the i-th and j-th individual-specific effects, respectively, λ_t the time-specific effect, and $Q_\tau(y_{ijt}|x_{ijt}, \alpha_i, \gamma_j, \lambda_t)$ is the conditional τ-quantile of y_{ijt} given $(x_{ijt}, \alpha_i, \gamma_j, \lambda_t)$. For future reference, we will define $\pi_{ijt} := (\alpha_i, \gamma_j, \lambda_t)$, $\pi_{ij} := (\alpha_i, \gamma_j)$, and $N = N_1 \cdot N_2$. This notation simplifies the discussion on the asymptotic properties, since it implicitly allows us to write $N \to \infty$ when both N_1 and N_2 diverge to infinity, or when one of these dimensions is fixed and the other grows to infinity. In practice, it is often the case that only one individual dimension is large (e.g., firm-employee linked data when the number of employees is much larger than the number of firms). Thus, one dimension, say N_1, might be small or considered fixed, while the other(s), say N_2, is considered large. Therefore, given the specific model of interest to establish the asymptotic properties of the desired estimator, one may consider different scenarios:

$$\begin{aligned}
&(i)\ N_1 \to \infty \text{ and } N_2 \text{ fixed},\ T \to \infty, \\
&(ii)\ N_1 \text{ fixed and } N_2 \to \infty,\ T \to \infty, \\
&(iii)\ N_1 \to \infty \text{ and } N_2 \to \infty,\ T \text{ fixed}, \\
&(iv)\ N_1 \to \infty \text{ and } N_2 \to \infty,\ T \to \infty.
\end{aligned} \tag{8.2}$$

Model (8.1) can be written as

$$y_{ijt} = x_{ijt}\beta(\tau) + \alpha_i(\tau) + \gamma_j(\tau) + \lambda_t(\tau) + \varepsilon_{ijt}(\tau), \tag{8.3}$$

where $\varepsilon_{ijt}(\tau)$ has a zero conditional τ-quantile given $(x_{ijt}, \alpha_j, \gamma_j, \lambda_t)$. In general, each α_i, γ_j, λ_t and β can depend on τ, but we assume τ to be fixed throughout the section and suppress such dependence for notational simplicity whenever there is

no confusion. Model (8.3) assumes that each individual-specific i and j effect enters additively in a linear model.

Koenker (2004) defines the conditional τ-quantile of interest for the dependent variable y conditional on x, for $\tau \in (0,1)$. Two different models can be proposed using this set-up depending on the interpretation of the individual-specific effects. The first is a model in which individual effects do not vary across τ as in Koenker (2004). In this case, the multi-dimensional effects have a pure *location* shift effect on the conditional quantiles. In this case, the pair (i, j) contains intrinsic characteristics which are assumed to be constant when studying conditional heterogeneity. Thus covariate heterogeneity is analyzed at a different level to individual heterogeneity. The second is a model in which individual effects are τ-specific as in Kato et al. (2012). This is a more flexible approach in which (i, j) effects are allowed to vary across the conditional heterogeneity. This is called the *distributional* shift model. However, the QR restrictions on estimation and asymptotic properties reveal that the large T requirement applies to both models, and the choice of each model is a trade-off between flexibility and degrees of freedom (i.e., number of observations with respect to the number of parameters, including the individual effects, to be estimated).

The location conditional τ-quantile of interest for the dependent variable y conditional on x is

$$Q_\tau(y_{ijt}|x_{ijt}) = \beta(\tau)'x_{ijt} + \pi_{ijt}, \tag{8.4}$$

in which π has a pure location shift effect on the conditional quantiles. This quantile model assumes the restrictions $Q_{\tau 1}(y_{ijt}|x_{ijt}) - Q_{\tau_2}(y_{ijt}|x_{ijt}) = (\beta(\tau_1) - \beta(\tau_2))'x_{ijt}$ for all $\tau_1, \tau_2 \in (0,1)$, that is, covariate heterogeneity is present only through changes in the slope parameters β.

The distributional conditional τ-quantile of interest for the dependent variable y conditional on x is

$$Q_\tau(y_{ijt}|x_{ijt}) = \beta(\tau)'x_{ijt} + \pi_{ijt}(\tau), \tag{8.5}$$

in which π has a location-scale shift effect on the conditional quantiles. This quantile model assumes that $Q_{\tau_1}(y_{ijt}|x_{ijt}) - Q_{\tau_2}(y_{ijt}|x_{ijt}) = (\beta(\tau_1) - \beta(\tau_2))'x_{ijt} + \pi_{ijt}(\tau_1) - \pi_{ijt}(\tau_2)$, for all $\tau_1, \tau_2 \in (0,1)$, that is, covariate heterogeneity is present through changes in the slope parameters β and ij-specific intercepts π.

For multi-dimensional panels, we can also consider a "mixed" model in which some intercept parameters vary with τ, while others do not. The choice of the τ-specific and τ-invariant components would depend on the nature of the covariate heterogeneity to be studied.

It is generally assumed that the innovation term ε is independent across individuals, which applied to our case means independence across i and j, but not identically distributed. If the disturbances are assumed to be identically distributed, then $\beta(\tau) = \beta, \forall \tau \in (0,1)$, i.e., the slope parameters are equal across quantiles, all the conditional quantiles are parallel and they only change depending on the location. In general, however, a more flexible model allows for ε to be dependent on the conditioning set $(x, \alpha, \gamma, \lambda)$, in which case $\beta(\tau) \neq \beta$ for some $\tau \in (0,1)$. A canonical ex-

ample of this situation is the location-scale model in which $\beta(\tau) = \beta + g(x)F_{\varepsilon}^{-1}(\tau)$, for some function $g(\cdot)$ of the covariates.

Now consider the FE-QR model, for which the conditional quantile model of interest is τ-specific,

$$Q_{\tau}(y_{ijt}|x_{it}, \pi_{ij}(\tau)) = \beta(\tau)'x_{ijt} + \pi_{ij}(\tau). \tag{8.6}$$

It is standard in the panel QR literature to treat π as fixed by conditioning on it, as in Hahn and Newey (2004), Fernandez-Val (2005), and Kato et al. (2012). Below we consider the fixed effects estimation of β, which is implemented by treating each individual-specific effect also as a parameter to be estimated. However, given the required estimation of π_{ij}, the FE-QR estimator is, unfortunately, subject to the incidental parameters problem (see Neyman and Scott, 1948; Lancaster, 2000, for a review) and will be inconsistent if the number of individual-specific effects diverges to infinity while the number of time periods T is fixed. It is important to note that, in contrast to mean regression there is no general transformation that can suitably eliminate the specific effects in the QR model. This intrinsic difficulty was recognized by Abrevaya and Dahl (2008), among others, and was clarified by Koenker and Hallock (2000). They remarked that "Quantiles of convolutions of random variables are rather intractable objects, and preliminary differencing strategies familiar from Gaussian models have sometimes unanticipated effects" (p.19).

Therefore, given these difficulties, in the QR panel data literature, it is usual to allow T to increase to infinity at a higher rate than N to achieve consistent estimators. As a result, the standard FE-QR model given by equation (8.6) does not consider a time-specific effect λ_t as a parameter to be estimated. In a multi-dimensional panel, careful consideration of asymptotics of the relative dimensions should be considered. In addition, as we will discuss below, because of the incidental parameters problem, one will be able to control for at most two out of the three specific effects $(\alpha, \gamma, \lambda)$. The allowed specific FE will also depend on the asymptotics considered, as given in cases $(i) - (iv)$ in (8.2) above.

Koenker (2004) and Kato et al. (2012) follow large (N, T) asymptotics (for other recent developments, see, e.g., Galvao, 2011; Galvao et al., 2013; Galvao and Wang, 2015). In the nonlinear and QR literatures, the large panel data asymptotics is used in an attempt to cope with the incidental parameters problem. Canay (2011) proposed a two-step estimator of the common parameters. The difference is that in his model, no individual effect is allowed to change across quantiles, and requires an additional restriction on the conditional average. Graham et al. (2009) show that when $T = 2$ and the explanatory variables are independent of the error term, the FE-QR estimator does not suffer from the incidental parameters problem. However, their argument does not seem to extend to general cases. Rosen (2012) addressed a set identification problem of the common parameters when T is fixed. Chernozhukov et al. (2013) considered identification and estimation of the quantile structural function defined in Imbens and Newey (2009) of a non-separable panel model with discrete explanatory variables. They studied the bounds of the quantile structural function when T is fixed, and the asymptotic behavior of the bounds when T goes to infinity.

In a multi-dimensional setting, the data structure determines the choice of the individual effects that one is able to control for and determines the nature of the model. In several cases, the researcher may be interested in exploring a particular time-invariant covariate set (say across i but not across j). As such, one may choose to explore heterogeneity across a certain dimension and not the other(s). In addition, the FE quantile panel models will produce different models to index heterogeneity depending on the conditional set. For instance, consider the model in equation (8.3). If $(\alpha, \gamma, \lambda)$ are controlled for, τ corresponds to an index of heterogeneity in the conditional quantile function of $y|(x, \alpha, \gamma, \lambda)$, which in fact depends on the quantiles of ε. In addition, the choice of the conditional model may depend on the stringent requirements on the time series encountered in the literature for asymptotic analysis. In multi-dimensional panels, additional dimensions are available to the researcher.

8.2.1 Estimation and Implementation

Koenker (2004) and Kato et al. (2012) consider the estimation of the FE using standard QR for a given quantile-τ as follows

$$\left(\hat{\pi}, \hat{\beta}\right) = \arg\min_{\pi,\beta} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \sum_{t=1}^{T} \rho_\tau(y_{ijt} - x'_{ijt}\beta - d'_{ijt}\pi), \tag{8.7}$$

where d_{ijt} is a set of dummy variables that identifies the individual FE for i and j given by π, and $\rho_\tau(u) := u(\tau - I(u < 0))$ as in Koenker and Bassett (1978). Note that the coefficients β correspond to the τ-quantile slopes $\beta(\tau)$. The estimation of the regression parameters can be implemented through a QR estimation augmented by the inclusion of the d_{ijt} dummy variables. Note that standard procedures for the estimation of the variance-covariance matrix of this augmented dummy variables estimator are feasible, and then the inference procedures described in the next section could follow from these estimation procedures.

The optimization for solving (8.7) can be very large depending on N_1, N_2 and T. However, as Koenker (2004) observe, in typical applications, the design matrix is very sparse. Standard sparse matrix storage schemes only require the space for the non-zero elements and their indexing locations. This considerably reduces the computational effort and memory requirements. Galvao and Wang (2015) address the computational difficulties and implementation problems without sacrificing the desirable asymptotic properties of the FE-QR strategy. They propose an efficient minimum distance QR estimator, which is very simple to implement in practice. This estimator is defined as the weighted average of the specific QR slope estimators, with weights given by the inverses of the corresponding individual variance-covariance matrices. Moreover, the implementation is not affected by the presence of unbalanced data as the dummy variables strategy works for both balanced and unbalanced panels.

The work in Koenker (2004) also introduced an alternative approach to estimate QR models for panel data with FE that may be subject to shrinkage by ℓ_1 regularization methods. It is well known that the optimal estimator for the random effects Gaussian model involves shrinking the individual effects toward a common value. When there is an intercept in the model, this common value can be taken to be the conditional central tendency of the response at a point determined by the centering of the other covariates. In the QR model, this would be some corresponding conditional quantile of the response. Particularly, when N is large relative to T, shrinkage may be advantageous in controlling the variability introduced by the large number of estimated individual-specific parameters. In this case, the model with shrinkage is

$$\left(\hat{\pi},\hat{\beta}\right) = \arg\min_{\pi,\beta} \sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\sum_{t=1}^{T} \rho_\tau(y_{ijt} - x'_{ijt}\beta - d'_{ijt}\pi) + \eta\left(\sum_{i=1}^{N_1}|\alpha_i| + \sum_{j=1}^{N_2}|\gamma_j|\right), \quad (8.8)$$

where $\eta \geq 0$ is a (scalar) penalty or regularization parameter. Note that for $\eta \to 0$ we obtain the FE estimator described above, while as $\eta \to \infty$ we obtain an estimate of the model purged of the FE. In the multi-dimensional case, the penalty is allowed to be different depending on the specific dimensions. For instance we could consider $(\eta_\alpha, \eta_\gamma)$, where the penalty parameter becomes $\eta_\alpha \sum_{i=1}^{N_1}|\alpha_j| + \eta_\gamma \sum_{j=1}^{N_2}|\gamma_j|$. In this case, we could have different degrees of tolerance for shrinking in the different dimensions. Usually for the dimension in which we believe that only a few FE should be non-zero is where $\eta_{(\cdot)}$ should be the largest.

8.2.2 Inference Procedures

As stated above, in standard FE-QR models, the asymptotic analysis for both models requires the time series dimension, T, to increase to infinity to achieve asymptotically unbiased estimators. In a standard two-dimension FE panel, Kato et al. (2012) show that we are required to impose more restrictive conditions on T than that found in the linear panel data FE literature. They show that a sufficient condition to prove asymptotic normality is $N^2(\log N)^3/T \to 0$, which reflects the fact that the rate of the remainder term of the Bahadur representation of the FE-QR estimator is of order $(T/\log N)^{-3/4}$. The slower convergence rate of the remainder term is due to the non-smoothness of the scores. It is important to note that the growth condition on T for establishing $\sqrt{NT}$-consistency of the FE-QR estimator (or other FE estimators in general) is determined so that it "kills" the remainder term. Thus, the rate of the remainder term is essential in the asymptotic analysis of the FE estimation when N and T jointly go to infinity. This restriction requires the cross-sectional dimensions to grow slower than the time dimension.

In a multi-dimensional setting, asymptotic valid inference will depend on the growth rate of the sample sizes described in equation (8.2). First, note that scenarios

(i) and (ii) in (8.2) require the same conditions on the sample size as stated in Kato et al. (2012). Nevertheless, although these models still require stringent conditions on N relative to T, they allow the researcher to control for π_{ij} because one of these dimensions is finite. However, since scenarios (i) and (ii) require large T, one is not allowed to control for λ_t.

Second, consider the case (iii) in (8.2). This case is also similar to that in Kato et al. (2012) with two dimensions diverging, but the time series dimension is fixed. In this case, one is able to control for only one individual specific effect, i or j, and the requirements on the sample size growth are imposed on the two individual dimensions relative to each other, such that one of the dimensions takes the role of the time series. Note that in this case, since the time series is given, one is also able to control for λ_t.

Finally, consider the case (iv) in (8.2). In this case, one is able to control for the two individual effects, π_{ij}, but note the time effect λ_t, as T diverges to infinity. However, it is important to note that the number of parameters in $\{(\alpha_i, \gamma_j)\}_{i=1,j=1}^{N_1,N_2}$ is $N_1 + N_2$, and this is in general considerably smaller than $N_1 \cdot N_2$. As the conditions discussed are imposed on N $(= N_1 \cdot N_2)$ relative to T, these requirements are more stringent than those that would be required to estimate $N_1 + N_2$ parameters. The main intuition is that although the number of parameters to be estimated grows with the sample size, in the three-dimension panel, the number of parameters to be estimated $(N_1 + N_2)$ is smaller than the sample size $(N = N_1 \cdot N_2)$. Another remark on case (iv) is that, since there are three dimensions in the panel data, one is able to exchange the roles of the indices (i, j, t) and estimate one of the individual effects and the time effect, and hence impose the restriction on the remaining dimension to grow fast relative to the other two.

In summary, in cases (i) and (ii) choosing λ_t to be excluded from model (8.6) is based on the idea that in FE models one mainly wishes to control for individual heterogeneity, which in this case is captured by π_{ij}. In case (iii), we could, however, exclude one component in π_{ij}, say α_i, and let its dimension, say N_1, to increase at a higher rate than the other dimensions, say N_2. In this case, one is able to control for λ_t because T is fixed. Finally, in case (iv), one is able to control for two effects only. In this case, one might be able to control for one individual effect and the time effect by exchanging the roles of the indices when considering the relative sample size growth. Therefore, different asymptotic conditions arise depending on different models. In all cases, we need one particular dimension to grow at a faster rate than the number of parameters to be estimated.

In practice, the asymptotic variance of FE-QR estimators depends on the density of the innovation term. For this estimation, different techniques have been suggested and implemented in the literature to produce a consistent estimation of the variance-covariance matrix.

Now we describe the asymptotic normality of $\hat{\beta}$ for case (iv) in (8.2), which can be obtained as in Kato et al. (2012). The other cases are parallel to the standard FE panel data in Kato et al. (2012). Under some regularity conditions, and $N^2(\log N)^3/T \to 0$, we have that

$$\sqrt{NT}(\hat{\beta}-\beta) \xrightarrow{d} N\{0, \tau(1-\tau)\Gamma^{-1}V\Gamma^{-1}\}\,.$$

Let

$$\tilde{\pi}_{ij} := \mathrm{E}[f_{ij}(0|x_{ij1})x_{ij1}]/f_{ij}(0) \quad \text{and} \quad \Gamma_N := N^{-1}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\mathrm{E}[f_{ij}(0|x_{ij1})x_{ij1}(x'_{ij1}-\tilde{\pi}'_{ij})]\,,$$

and where $f_{ij}(u|x)$ is the density of $u_{ijt} = y_{ijt} - \pi_{ij} - x_{jit}\beta$ conditional on x_{ij1} and $f_{ij}(u)$ is the marginal density of u_{ijt}. Let Γ_N be nonsingular for each N, and the limit $\Gamma := \lim_{N\to\infty}\Gamma_N$ exists and is nonsingular; and let the limit

$$V := \lim_{N\to\infty} N^{-1}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\mathrm{E}[(x_{ij1}-\pi_{ij})(x_{ij1}-\pi_{ij})']$$

exist and be nonsingular.

Let $\hat{u}_{ijt} = y_{ijt} - \hat{\pi}_{ij} - x'_{jit}\hat{\beta}$. Kato et al. (2012) propose a kernel estimation procedure of the variance of the slope parameters, $V_{\beta(\tau)}$. Let $\mathscr{K} : \mathbb{R} \to \mathbb{R}$ denote a kernel function (probability density function). Let $\{h_N\}$ denote a sequence of positive numbers (bandwidths) such that $h_N \to 0$ as $N \to \infty$ and use the notation $\mathscr{K}_{h_N}(u) = h_N^{-1}\mathscr{K}(u/h_N)$. Assume that the kernel $\mathscr{K}$ is continuous, bounded and of bounded variation on $\mathbb{R}$, and that $h_N \to 0$ and $\log N/(Th_N) \to 0$ as $N \to \infty$.

Define

$$\hat{V}_{1\beta} := \frac{1}{NT}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\sum_{t=1}^{T}\mathscr{K}_{h_n}(\hat{u}_{ijt})x_{ijt}(x_{ijt}-\hat{\pi}_{ij})',$$

$$\hat{V}_{0\beta} := \frac{1}{NT}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\sum_{t=1}^{T}(x_{ijt}-\hat{\pi}_{ij})(x_{ijt}-\hat{\pi}_{ij})',$$

where

$$\hat{\pi}_{ij} := \frac{1}{\hat{f}_{ij}T}\sum_{t=1}^{T}\mathscr{K}_{h_N}(\hat{u}_{ijt})x_{ijt}, \quad \hat{f}_{ij} := \frac{1}{T}\sum_{t=1}^{T}\mathscr{K}_{h_N}(\hat{u}_{ijt}).$$

Then, one can consistently estimate the variance-covariance matrix as

$$\hat{V}_{\beta(\tau)} = \tau(1-\tau)\hat{V}_{1\beta}^{-1}\hat{V}_{0\beta}\hat{V}_{1\beta}^{-1}.$$

In practice, as noted above, the variance-covariance matrix can be implemented from standard QR models with the inclusion of individual-specific dummy variables.

8.2.3 Smoothed Quantile Regression Panel Data

A distinctive feature of FE-QR is that its objective function is not differentiable. Nevertheless, the asymptotic analysis depends on the smoothness of objective functions. Kato et al. (2012) formally established the asymptotic properties of the standard FE-QR estimator. However, they required a restrictive condition, such that T grows faster than N^2, to show the asymptotic normality of the estimator, and did not succeed in deriving the bias. The difficulty in handling the standard QR estimator in panel models is partly explained by the fact that the higher order stochastic expansion of the scores is an essential technical tool in the analysis (Hahn and Newey, 2004; Hahn and Kuersteiner, 2004) but such expansion is difficult to implement in the QR case because the Taylor series method is not directly applicable. It is also important to note that the higher order asymptotic behavior of QR estimators is non-standard and rather complicated (Arcones, 1998; Knight, 1998).

An alternative method proposed by Galvao and Kato (2016) is to slightly modify the QR objective function to make it smooth. While this seems an *ad-hoc* change of the objective function, its asymptotic gains are remarkable. The idea of smoothing non-differentiable objective functions goes back to Amemiya (1982) and Horowitz (1992, 1998). Under suitable regularity conditions, the smoothed FE-QR estimator has an order $O(T^{-1})$ bias and hence its limiting normal distribution has a bias in the mean (even) when N and T grow at the same rate. They propose a one-step bias correction estimator based on the analytic form of the asymptotic bias. This is of particular interest in multi-dimensional settings, where the dimension of the individuals is large.

In an attempt to cope with the incidental parameters problem, Galvao and Kato (2016) adopt a different approach and propose a model where N and T grow at the same rate. Instead of the standard QR estimator, the asymptotic properties of the estimator are defined by a minimizer of a smoothed version of the QR objective function.

Smoothing the QR objective function was employed in Horowitz (1998) to study the bootstrap refinement for inference in conditional quantile models. The basic insight of Horowitz (1998) is to smooth over the indicator function $I(y_{ijt} \leq \pi_{ij} + x_{it}'\beta)$ by using a kernel function. To do so, let $K(\cdot)$ be a kernel function and $G(\cdot)$ be the survival function of $K(\cdot)$, i.e.,

$$\int_{-\infty}^{\infty} K(u)du = 1,\ G(u) := \int_{u}^{\infty} K(v)dv.$$

$K(\cdot)$ is not required to be non-negative. Let $\{h_N\}$ be a sequence of positive numbers (bandwidths) such that $h_N \to 0$ as $N \to \infty$ and write $G_{h_N}(\cdot) = G(\cdot/h_N)$. Note that $G_{h_N}(y_{ijt} - \pi_{ij} - \gamma_j - x_{ijt}'\beta)$ is a smoothed counterpart of $I(y_{ijt} \leq \pi_{ij} + x_{ijt}'\beta)$. Then, we consider the estimator

$$(\hat{\pi}, \hat{\beta}) := \arg\min_{(\pi,\beta)} \sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\sum_{t=1}^{T}(y_{ijt} - \pi_{ij} - x_{ijt}'\beta)\{\tau - G_{h_N}(y_{ijt} - \pi_{ij} - x_{ijt}'\beta)\}, \quad (8.9)$$

The estimates $\hat{\beta}$ are the FE smoothed quantile regression (FE-SQR) estimator of β.

Galvao and Kato (2016) investigate the asymptotic properties of the FE-SQR estimator defined by (8.9) and provide conditions under which the FE-SQR estimator is consistent and has a limiting normal distribution with a bias in the mean when N and T grow at the same rate. In particular, assuming that $N/T \to \rho$ for some $\rho > 0$, and under some regularity conditions,

$$\sqrt{NT}(\hat{\beta} - \beta) \xrightarrow{d} N(\sqrt{\rho}b, \Gamma^{-1} V \Gamma^{-1}) , \tag{8.10}$$

where

$$b := \Gamma^{-1} \left[\lim_{N\to\infty} \left\{ \frac{1}{N} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} s_{ij} \left(\omega_{ij}^{(1)} \pi_{ij} - \omega_{ij}^{(2)} + \frac{s_{ij} \omega_{ij}^{(3)} \nu_{ij}}{2} \right) \right\} \right] , \tag{8.11}$$

with

$$\begin{aligned} s_{ij} &:= 1/f_{ij}(0) , \quad \pi_{ij} := s_{ij} \mathrm{E}[f_{ij}(0|x_{ij1}) x_{ij1}] , \\ \nu_{ij} &:= f_{ij}^{(1)}(0) \pi_{ij} - \mathrm{E}[f_{ij}^{(1)}(0|x_{ij1}) x_{ij1}] , \\ \Gamma_N &:= N^{-1} \textstyle\sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \mathrm{E}[f_{ij}(0|x_{ij1}) x_{ij1} (x'_{ij1} - \pi'_{ij})] , \end{aligned}$$

and the limit $\Gamma := \lim_{N\to\infty} \Gamma_N$, and

$$V := \lim_{N\to\infty} N^{-1} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} V_{ij}$$

with V_{ij} denoting the covariance matrix of the term

$$T^{-1/2} \sum_{t=1}^{T} \{\tau - I(u_{ijt} \le 0)\}(x_{ijt} - \pi_{ij}) .$$

Moreover,

$$\begin{aligned} \omega_{ij}^{(1)} &:= \sum_{1\le|k|\le T-1} \left(1 - \frac{|k|}{T}\right) \left\{ \tau f_{ij}(0) - \int_{-\infty}^{0} f_{ij,k}(0,u) du \right\} , \\ \omega_{ij}^{(2)} &:= \sum_{1\le|k|\le T-1} \left(1 - \frac{|k|}{T}\right) \left\{ \tau \mathrm{E}[f_{ij}(0|x_{ij1}) x_{ij1}] \right. \\ &\qquad \left. - \mathrm{E}\left[x_{ij1} \int_{-\infty}^{0} f_{ij,k}(0,u|x_{ij1}, x_{ij,1+k}) du \right] \right\} , \\ \omega_{ij}^{(3)} &:= \sum_{|k|\le T-1} \left(1 - \frac{|k|}{T}\right) \mathrm{Cov}\{I(u_{ij1} \le 0), I(u_{ij,1+k} \le 0)\} . \end{aligned}$$

The exact form of the term V_{ij} is given by

$$V_{ij} = \sum_{|k|\leq T-1} \left(1 - \frac{|k|}{T}\right) \mathrm{E}\left[\{\tau - I(u_{ij1} \leq 0)\}\{\tau - I(u_{ij,1+k} \leq 0)\} \times \right.$$
$$\left. \times (x_{ij1} - \pi_{ij})(x_{ij,1+k} - \pi_{ij})'\right].$$

If there is no time series dependence, i.e., for each i, j, the process $\{(y_{ijt}, x_{ijt}), t = 0, \pm 1, \pm 2, \dots\}$ is i.i.d., then

$$V_{ij} = \tau(1-\tau)\mathrm{E}[(x_{ij1} - \pi_{ij})(x_{ij1} - \pi_{ij})'], \quad \omega_{ij}^{(1)} = 0, \quad \omega_{ij}^{(2)} = 0, \quad \text{and}$$
$$\omega_{ij}^{(3)} = \tau(1-\tau)$$

8.2.3.1 Bias Correction – Analytical Method

As stated in the literature, the problem of the limiting distribution of $\sqrt{NT}(\hat{\beta} - \beta)$ not being centered at zero is that usual confidence intervals based on the asymptotic approximation will be incorrect. In particular, even if b is small, the asymptotic bias can be of moderate size when the ratio N/T is large. In this subsection, we shall consider the bias correction to the FE-SQR estimator.

Consider a one-step bias correction based on the analytic form of the asymptotic bias. Put $\hat{u}_{ijt} := y_{it} - \hat{\pi}_{ij} - x_{ijt}'\hat{\beta}$. The terms $f_{ij} := f_{ij}(0), s_{ij}, \pi_{ij}, \nu_{ij}$ and Γ can be estimated by

$$\hat{f}_{ij} := \frac{1}{T}\sum_{t=1}^{T} K_{h_N}(\hat{u}_{ijt}), \quad \hat{s}_{ij} := \frac{1}{\hat{f}_{ij}}, \quad \hat{\pi}_{ij} := \frac{\hat{s}_{ij}}{T}\sum_{t=1}^{T} K_{h_N}(\hat{u}_{ijt}) x_{ijt},$$
$$\hat{\nu}_{ij} := \frac{1}{Th_N^2}\sum_{t=1}^{T} K^{(1)}(\hat{u}_{ijt}/h_N)(x_{ijt} - \hat{\pi}_{ij}),$$
$$\hat{\Gamma}_N := \frac{1}{NT}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\sum_{t=1}^{T} K_{h_N}(\hat{u}_{ijt}) x_{ijt}(x_{ijt}' - \hat{\pi}_{ij}'),$$

where $K^{(1)}(u) = dK(u)/du$. The estimation of the terms $\omega_{ij}^{(1)}, \omega_{ij}^{(2)}$ and $\omega_{ij}^{(3)}$ is a more delicate issue, since it reduces to the estimation of long run covariances. As in Hahn and Kuersteiner (2004), we make use of a truncation strategy. Define

$$\phi_{ij}(k) := \int_{-\infty}^{0} f_{ij,k}(0,u)du,$$
$$\varphi_{ij}(k) := \mathrm{E}\left[x_{ij1}\int_{-\infty}^{0} f_{ij,k}(0,u|x_{ij1}, x_{ij,1+k})du\right],$$
$$\rho_{ij}(k) := \mathrm{E}[I(u_{ij1} \leq 0)I(u_{ij,1+k} \leq 0)].$$

Since $\phi_{ij}(k) \approx \mathrm{E}[K_{h_N}(u_{ij1})I(u_{ij,1+k} \leq 0)]$, it can be estimated by

$$\hat{\phi}_{ij}(k) := \frac{1}{T} \sum_{t=\max\{1,-k+1\}}^{\min\{T,T-k\}} K_{h_N}(\hat{u}_{ijt}) I(\hat{u}_{ij,t+k} \leq 0).$$

Similarly, $\varphi_{ij}(k)$ can be estimated by

$$\hat{\varphi}_{ij}(k) := \frac{1}{T} \sum_{t=\max\{1,-k+1\}}^{\min\{T,T-k\}} K_{h_N}(\hat{u}_{ijt}) I(\hat{u}_{ij,t+k} \leq 0) x_{ijt}.$$

The term $\rho_{ij}(k)$ can be estimated by its sample analogue:

$$\hat{\rho}_{ij}(k) := \frac{1}{T} \sum_{t=\max\{1,-k+1\}}^{\min\{T,T-k\}} I(\hat{u}_{ijt} \leq 0) I(\hat{u}_{ij,t+k} \leq 0).$$

Take a sequence m_N such that $m_N \to \infty$ sufficiently slowly. Then, $\omega_{ij}^{(1)}$, $\omega_{ij}^{(2)}$ and $\omega_{ij}^{(3)}$ can be estimated by

$$\hat{\omega}_{ij}^{(1)} := \sum_{1 \leq |k| \leq m_N} \left(1 - \frac{|k|}{T}\right) \{\tau \hat{f}_{ij} - \hat{\phi}_{ij}(k)\},$$

$$\hat{\omega}_{ij}^{(2)} := \sum_{1 \leq |k| \leq m_N} \left(1 - \frac{|k|}{T}\right) \{\tau \hat{f}_{ij} \hat{\pi}_{ij} - \hat{\varphi}_{ij}(k)\},$$

$$\hat{\omega}_{ij}^{(3)} := \tau(1-\tau) + \sum_{1 \leq |k| \leq m_N} \left(1 - \frac{|k|}{T}\right) \{-\tau^2 + \hat{\rho}_{ij}(k)\}.$$

The bias term b is thus estimated by

$$\hat{b} := \hat{\Gamma}_N^{-1} \left\{ \frac{1}{N} \sum_{i=1}^{N_1} \sum_{i=1}^{N_2} \hat{s}_{ij} \left(\hat{\omega}_{ij}^{(1)} \hat{\pi}_{ij} - \hat{\omega}_{Nij}^{(2)} + \frac{\hat{s}_{ij} \hat{\omega}_{ij}^{(3)} \hat{v}_{ij}}{2} \right) \right\}.$$

We define the one-step bias corrected estimator by $\hat{\beta}^1 := \hat{\beta} - \hat{b}/T$. In practice, there is no need to compute the terms $\tau \hat{f}_{ij}$ and $\tau \hat{f}_{ij} \hat{\pi}_{ij}$ in $\hat{\omega}_{ij}^{(1)}$ and $\hat{\omega}_{ij}^{(2)}$, respectively, as they are canceled out by the difference $\hat{\omega}_{ij}^{(1)} \hat{\pi}_{ij} - \hat{\omega}_{ij}^{(2)}$. Additionally, there is no need to use the same kernel and the same bandwidth to estimate β and b.

Galvao and Kato (2016) show that the bias corrected estimator, $\hat{\beta}^1$, has the limiting normal distribution with mean zero and the same covariance matrix as $\hat{\beta}$ as

$$\sqrt{NT}(\hat{\beta}^1 - \beta) \xrightarrow{d} N(0, \Gamma^{-1} V \Gamma^{-1}),$$

when $m_N \to \infty$ such that $m_N^2 (\log N)/(T h_N^2) \to 0$.

8.2.3.2 Jackknife

Galvao and Kato (2016) also consider the half-panel jackknife method originally proposed by Dhaene and Jochmans (2015), as an estimator. This method is an automatic way of removing the bias of $\hat{\beta}$. Suppose for a moment that T is even. Partition $\{1,\ldots,T\}$ into two subsets, $S_1 := \{1,\ldots,T/2\}$ and $S_2 := \{T/2+1,\ldots,T\}$. Let $\hat{\beta}_{S_l}$ be the FE-SQR estimate based on the data $\{(y_{ijt},x_{ijt}), 1 \leq i \leq N_1, 1 \leq j \leq N_2, t \in S_l\}$ for $l = 1,2$. The half-panel jackknife estimator is defined as $\hat{\beta}_{1/2} := 2\hat{\beta} - \bar{\beta}_{1/2}$, where $\bar{\beta}_{1/2} := (\hat{\beta}_{S_1} + \hat{\beta}_{S_2})/2$. For simplicity, suppose for a moment that we use the same bandwidth to construct $\hat{\beta}$ and $\hat{\beta}_{S_l}$ ($l = 1,2$). Then, from the asymptotic representation of the FE-SQR estimator, it can be shown that under some regularity conditions

$$\sqrt{NT}(\hat{\beta}_{1/2} - \beta) \xrightarrow{d} N(0, \Gamma^{-1} V \Gamma^{-1}).$$

The half-panel does not require the non-parametric estimation of the bias term and at the same time is easy to implement empirically.

8.3 Random Effects Models

Random effects (RE) models have recently been considered in the QR panel data framework. As noted in Chap. 2, RE models have two main advantages. First, the RE approach does not suffer from the incidental parameters problem, that is the number of parameters to take into account does not increase with the sample size (i.e., N). This is an important restriction to be lifted for QR models as noted previously, because for FE-QR models the existing sufficient conditions under which the asymptotic bias of the FE-QR vanishes require $T >> N$. For the general multi-dimensional setting, this is an important restriction for models in which the number of intercepts is of the order $O(N_1 + N_2)$ or $O(N_1 \cdot N_2)$. Koenker (2004) argues that the latter "would certainly be useful for groups of individuals: a distributional shift for men versus women, or for blacks versus whites. However, in most applications the $[T]$, the number of observations in the time series, would be relatively modest and then it is quite unrealistic to attempt to estimate a τ-dependent, distributional, individual effect" (p.76). In most applications, the time series dimension T is indeed relatively small compared to the number of individuals.

Second, the RE model also makes possible the identification of parameters associated with individual (and time) invariant variables. In a multi-dimensional framework, this may be of interest for the applied researcher that wants to control for (i,j,t) heterogeneity while exploring covariate heterogeneity across conditional quantiles.

Galvao and Poirier (2015) develop a RE model for QR panel data with time invariant regressors. They establish identification, and develop practical estimation and inference procedures. In this section, we extend the RE model to the multi-

dimensional context and apply a simple pooled QR estimator to estimate the coefficients of interest and establish its statistical properties. We also suggest a cluster robust variance-covariance matrix estimator for inference, and establish its uniform consistency. The RE model is interesting because it allows the researcher to control for time-invariant regressors, as well as use small a panel where the time dimension, T, is small.

8.3.1 Model

Consider now a linear RE-QR model with scalar multi-dimensional specific effects (for simplicity, we follow the notation of Chap. 1 rather than Chap. 2). Let z_{ij} be a set of covariates that does not vary across t.

Following Galvao and Poirier (2015), we begin the discussion with a random coefficients representation of the form

$$y_{ijt} = c(U_{ijt}) + x'_{ijt}\beta(U_{ijt}) + z'_{ij}\delta(U_{ijt}), \tag{8.12}$$

where U_{ijt} represents the heterogeneity in responses and can depend on both ε_{ijt} and π_{ij}, as

$$U_{ijt} \equiv U(\pi_{ij}, \varepsilon_{ijt}), \tag{8.13}$$

with $U(\cdot,\cdot)$ being a scalar and unspecified non-parametric function. Note that equation (8.13) allows the unobserved heterogeneity to depend on both the independent unobserved component, ε_{ijt}, and the individual-specific components, π_{ij}, in an unrestricted form. The functions $c(\cdot), \beta(\cdot)$ and $\delta(\cdot)$ in (8.12) quantify the distributional effects for the intercept, and the time-varying and time-invariant regressors, x_{ijt} and z_{ij} respectively. Note that for the RE-QR models, since the time series dimension T is fixed, one can easily include a time-specific effect λ_t as regressors with corresponding parameters to be estimated inside the vector x_{ijt} in equation (8.12), hence we only consider π_{ij} in equation (8.13).

The RE assumption in standard linear mean panel data models restricts the unobserved component, π_{ij}, to being uncorrelated with all regressors. We generalize this assumption to the model in (8.12)–(8.13) by assuming the following independence condition

$$(\pi_{ij}, \varepsilon_{ijt}) \perp\!\!\!\perp (x_{ijt}, z_{ij}). \tag{8.14}$$

The stronger independence assumption in (8.14) is used due to the non-linearity in π_{ij} of equations (8.12)–(8.13). Thus the unobserved heterogeneity $U(\pi_{ij}, \varepsilon_{ijt})$ is independent of (x_{ijt}, z_{ij}), which gives rise to the following quantile representation

$$Q_\tau(y_{ijt}|x_{ijt}, z_{ij}) = c(\tau) + x'_{ijt}\beta(\tau) + z'_{ij}\delta(\tau)\,, \tag{8.15}$$

where the presence of τ on the right-hand side follows from our previous normalization of U_{ijt} and from $Q_\tau(U_{ijt}|x_{ijt}, z_{ij}) = Q_\tau(U_{ijt}) = \tau$. Equation (8.15) establishes

the linear RE-QR model, given equations (8.12)–(8.13), and condition (8.14). For notation convenience define $w_{ijt} = [1, x'_{ijt}, z'_{ij}]'$ and $\theta(.) = [c(.), \beta(.), \delta(.)]$.

Before we present the estimation, it is important to discuss the differences between the FE-QR and the RE-QR. Galvao and Poirier (2015) show that the relationship between RE and FE is more delicate for quantile models than for standard mean-regression models. In a linear panel model, traditional conditional mean FE and RE estimation are based on the same linear model, which often takes the form

$$y_{ijt} = x'_{ijt}\beta + \pi_{ij} + \varepsilon_{ijt}.$$

Under FE, i.e., $\text{Cov}(\alpha_{ij}, x_{ijt}) \neq 0$, the "within" estimator can recover β. On the other hand, if the RE assumption of $\text{Cov}(\alpha_{ij}, x_{ijt}) = 0$ is assumed, the pooled regression estimator will also be consistent for β. However, in QR models, the FE and RE models differ substantially.

To see this difference, consider the additive-in-π_{ij} linear FE-QR model, which can be represented as

$$y_{it} = \pi_{ij} + x'_{ijt}\beta(\varepsilon_{ijt}). \tag{8.16}$$

There are two important points regarding model (8.16). First, note that under FE or RE, the *conditional* model yields $Q_\tau(y_{ijt}|x_{ijt}, \pi_{ij}) = \pi_{ij} + x'_{ijt}\beta(\tau)$, a linear expression. Under some regularity conditions, and no restriction on the relationship between π_{ij} and x_{ijt}, Kato et al. (2012) show that $\beta(\tau)$ can be estimated consistently by a QR with individual-specific dummy variables when T is large. Note that the inclusion of individual-specific dummy variables precludes one from having time-invariant regressors. Nevertheless, even if $\pi_{ij} \perp\!\!\!\perp x_{ijt}$ holds in (8.16) a simple RE estimator would be unable to consistently estimate $\beta(\tau)$ since the conditional quantile function would be misspecified if we do not condition on the individual specific effects π_{ij}. This implies that the pooled QR and the FE-QR estimators estimate different quantities.

Second, note that there are differences between $Q_\tau(y_{ijt}|x_{ijt}, \pi_{ij})$, the *conditional* model, and $Q_\tau(y_{ijt}|x_{ijt})$, the *marginal* model. Even if $\pi_{ij} \perp\!\!\!\perp x_{ijt}$ holds in (8.16), the conditional quantile of y_{ijt} given x_{ijt} might not be linear in x_{ijt} since $Q_\tau(y_{ijt}|x_{ijt}) \neq Q_\tau(\pi_{ij}) + x'_{ijt}\beta(\tau)$, because the quantile of a sum is generally different from the sum of the quantiles by the non-linearity of the quantile operator.[1] Again, the pooled QR and the FE-QR estimators estimate different quantities.

In equations (8.12)–(8.13), consider an alternative model which is non-additive in the individual-specific effect, α_i, as

$$y_{ijt} = x'_{ijt}\beta(U(\pi_{ij}, \varepsilon_{ijt})).$$

Under assumption (8.14), the conditional quantile of y_{ijt} given x_{ijt} in the above equation, i.e., the marginal model, is linear in x_{ijt} and $\beta(\tau)$. Again, the conditional

[1] The conditional quantile of the sum will be equal to the sum of conditional quantiles if conditional co-monotonicity between $\pi_{ij}i$ and $x'_{it}\beta(\varepsilon_{ijt})$ holds conditional on x_{ij}. This is ruled out by the conditional independence of π_{ij} and ε_{ijt} given x_{ij} since co-monotonic variables cannot be independent (see Galvao and Poirier (2015) for a proof of this result).

model will yield a different effect of x_{ijt} on y_{ijt}, so a FE estimator with individual-specific dummy variables would not recover the same coefficients on x_{ijt} due to the non-additivity of π_{ij}. Thus, the FE-QR and RE-QR estimators will converge to different quantities, since they must rely on different modeling assumptions, which stands in contrast to the linear mean-regression panel case, where both the FE and RE estimators are consistent. This is due to the fact that under RE, the marginal and conditional models for linear, mean-regression panels yield the same effect of x_{ijt} on y_{ijt},[2] while these effects differ for all models considered here.

Another feature of the non-additive RE-QR model is that a failure of the RE assumption (8.14) will imply that the conditional quantile of y_{ijt} is no longer linear in w_{ijt}, because the composite unobserved heterogeneity term U_{ijt} will generally be correlated with w_{ijt}. We can then write the conditional quantile of y_{ijt} as follows

$$\begin{aligned} Q_\tau(y_{ijt}|w_{ijt}) &= w'_{ijt}\theta(Q_\tau(U_{ijt}|w_{ijt})) \\ &\equiv w'_{ijt}\tilde{\theta}(\tau; w_{ijt}), \end{aligned}$$

where $\tilde{\theta}(\tau; w_{ijt})$ is a non-parametric function of τ and w_{ijt}. Graham et al. (2015) discuss the non-parametric identification and estimation of this model, and more specifically of the unconditional quantile effect $\theta(\tau)$.[3]

8.3.2 Estimation and Implementation

Based on the identification condition (8.14) and the model given in (8.15), a simple pooled QR estimator for $\theta(\tau) = [c(\tau), \beta(\tau)', \delta(\tau)']'$ can be employed. The estimator is defined as follows:

$$\hat{\theta}(\tau) \equiv \arg\min_{\theta} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \sum_{t=1}^{T} \rho_\tau(y_{ijt} - w'_{ijt}\theta), \tag{8.17}$$

where $\rho_\tau(u) \equiv \{\tau - 1(u \leq 0)\}u$ is, again, the check function (Koenker and Bassett, 1978). Therefore, the practical estimation procedure for the coefficients of interest is very simple and can thus be implemented through standard QR estimation. First, one stacks the data, and second, applies a simple QR. Nevertheless, given that the individual effects induce clustering, the inference needs to be adjusted. We describe inference procedures in the next section.

[2] This can be seen from $\mathrm{E}[y_{ijt}|x_{ijt}, \pi_{ij}] = \pi_{ij} + x'_{ijt}\beta$ and $\mathrm{E}[y_{ijt}|x_{ijt}] = \mathrm{E}[\pi_{ij}] + x'_{ijt}\beta$.

[3] These differences between the RE and FE models in the QR case make testing for the presence of RE very important in the QR context. Galvao and Poirier (2015) provide such a test.

8.3.3 Inference Procedures

Under some standard regularity conditions, Galvao and Poirier (2015) derive the asymptotic normality of the RE-QR estimator as follows. For a given quantile τ of interest, as $N \to \infty$, $\hat{\theta}(\cdot)$

$$\sqrt{N}(\hat{\theta}(\tau) - \theta(\tau)) \xrightarrow{d} N(0, \Gamma(\tau)^{-1} V(\tau) \Gamma(\tau)^{-1}),$$

where $\Gamma(\tau) \equiv \mathrm{E}\left[\frac{1}{T}\sum_{t=1}^{T} w_{ijt} w'_{ijt} f_{y_{ijt}}(w'_{ijt}\theta(\tau)|w_{ij})\right]$, and

$$V(\tau,\tau') = \frac{1}{T^2}\sum_{s=1}^{T}\sum_{t=1}^{T} \mathrm{E}\left[(1(v_{it}(\tau) \leq 0) - \tau)(1(v_{is}(\tau') \leq 0) - \tau') X_{it} X'_{is}\right]. \quad (8.18)$$

The existence of the RE in the model generates cluster-dependence, and thus the standard errors require a cluster-robust variance-covariance matrix estimation.

For given quantiles of interest, the variance-covariance matrix of $\hat{\theta}(\tau)$ is

$$\Gamma(\tau)^{-1} V(\tau) \Gamma(\tau)^{-1}$$

with components $\Gamma(\tau) = \mathrm{E}\left[\frac{1}{T}\sum_{t=1}^{T} w_{ijt} w'_{ijt} f_{y_{ijt}}(w'_{ijt}\theta(\tau)|w_{ij})\right]$, and (8.18) can be rewritten as

$$\begin{aligned} V(\tau) &= \mathrm{E}\left[\frac{1}{T^2}\sum_{s=1}^{T}\sum_{t=1}^{T}(1(v_{ijt}(\tau) \leq 0) - \tau)(1(v_{ijs}(\tau) \leq 0) - \tau) w_{ijs} x'_{ijt}\right] \\ &= \frac{\tau(1-\tau)}{T^2}\sum_{t=1}^{T} \mathrm{E}[w_{ijt} w'_{ijt}] \\ &\quad + \frac{1}{T^2}\sum_{s \neq t} \mathrm{E}\left[\mathrm{Cov}(1(v_{ijs}(\tau) \leq 0), 1(v_{ijt}(\tau) \leq 0)|w_{ijs}, w_{ijt}) w_{ijs} w'_{ijt}\right], \end{aligned}$$

where $v_{it}(\tau) \equiv y_{ijt} - (c(\tau) + x'_{ijt}\beta(\tau) + z'_{ij}\gamma(\tau))$ and $f(.)$ is the conditional density of y_{ijt} given $w_{ij} = [w_{ij1}, w_{ij2}, ..., w_{ijT}]$. Note that in the second component, the second term disappears if there is no intra-unit dependence of the QR residuals. Thus, using simple standard errors for the pooled QR estimator without correcting for the cluster-dependence will produce incorrect inference unless the second term is zero.

To conduct practical inference, consider

$$\hat{\Gamma}(\tau) = \frac{1}{NT}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\sum_{t=1}^{T} w_{ijt} w'_{ijt} \frac{1}{h_N} K\left(\frac{\hat{v}_{ijt}(\tau)}{h_N}\right),$$

where $\hat{v}_{ijt}(\tau)$ are the estimated residuals, and $K(\cdot)$ is a kernel function of bounded variation, and h_N is a bandwidth. This is a variant of the Powell (1986) kernel estimator for QR in cross-sectional models. The component $V(\tau)$ can be estimated by

$$\hat{V}(\tau) = \frac{\tau(1-\tau)}{NT^2}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\sum_{t=1}^{T} w_{ijt}w'_{ijt} + \frac{1}{NT^2}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\sum_{s\neq t} w_{ijs}w'_{ijt}(1(\hat{v}_{ijs}(\tau)\leq 0)-\tau)(1(\hat{v}_{ijt}(\tau)\leq 0)-\tau).$$

8.4 Correlated Random Effects Models

Another alternative class of models for QR panel data is the correlated random effects. The correlated RE model of Chamberlain (1982, 1984) views the unobservable individual specific component π_{ij} as a linear projection onto the observables plus a disturbance. The intuition behind these models is that a rich set of covariates is able to explain unobserved heterogeneity and what is left is idiosyncratic noise. This idea has also been implemented in QR panel data models.

Abrevaya and Dahl (2008) introduced an alternative approach to the FE-QR, which estimates QR models for panel data employing the correlated random effects (CRE) model of Chamberlain (1982). The unobservable individual specific effect is modeled as a linear projection onto the observables and a disturbance. Geraci and Bottai (2007) consider a RE approach for a single quantile assuming that the outcome variable is distributed as an asymmetric Laplace distribution conditional on covariates and individual effects. Arellano and Bonhomme (2016) introduce a class of QR estimators for short panels, where the conditional quantile response function of the unobserved heterogeneity is specified as a function of observables. They develop a model general model for nonlinear panel data that covers static and dynamic autoregressive models, models with general predetermined regressors, and models with multiple individual effects. However, the correlated RE requires a specification of the individuals specific effects as a known function of the observables.

Extensions to the multi-dimensional case allow for different covariates to be used for each specific components, and as such, to be able to produce a more accurate model of unobserved effects. Let z_i, w_j and b_j be covariate sets that only vary on a given dimension, i, j, and t, possibly nested within x_{ijt}, and let (a_i, g_j, r_t) be unobserved components such that

$$\begin{aligned}\alpha_i &= \lambda_1' z_i + a_i,\\ \gamma_j &= \lambda_2' w_j + g_j,\\ \lambda_t &= \lambda_3' b_t + r_t.\end{aligned}$$

Following Abrevaya and Dahl (2008), the strategy is to replace them into equation (8.5) to obtain an explicit model of the quantiles as

$$Q_\tau(y_{ijt}|x_{ijt}) = \beta(\tau)'x_{ijt} + \lambda_1(\tau)'z_i + \lambda_2(\tau)'w_j + \lambda_3' b_t.$$

Arellano and Bonhomme (2016) develop an estimation strategy for general nonlinear panel models. They specify outcomes y_{ijt} as a function of covariates x_{ijt} and latent heterogeneity π_{ij} as

$$y_{ijt} = \sum_{k=1}^{K_1} \theta_k(U_{ijt}) g_k(x_{ijt}, \pi_{ij}),$$

and similarly specify the dependence of π_{ij} on covariates

$$\pi_{ij} = \sum_{k=1}^{K_2} \delta_k(V_{ij}) h_k(x_{ijt}),$$

where $U_{ij1}, ..., U_{ijT}, V_{ij}$ are independent uniform random variables, and g(.) and h(.) belong to some family of functions. Outcomes and heterogeneity are monotone in U_{ijt} and V_{ij}, respectively, so the above models correspond to conditional quantile functions. This is a correlated RE model that can become arbitrarily flexible as K_2 increases. For the multi-dimensional case, this could be made as

$$\alpha_i = \sum_{k=1}^{K_2} \delta_h^{\alpha}(V_{\alpha i}) h_k^{\alpha}(x_{ijt}),$$

$$\gamma_j = \sum_{k=1}^{K_3} \delta_h^{\gamma}(V_{\gamma j}) h_k^{\gamma}(x_{ijt}),$$

where heterogeneity is modeled in a different way for each dimension.

8.5 Specific Guidelines for Practitioners

QR has attracted considerable interest in econometrics and statistics. It offers an easy-to-implement method to estimate conditional quantiles. Recently, there has been a growing literature on estimation and testing using QR for panel data models. Panel QR has provided a valuable method of statistical analysis of the heterogeneous effects of policy variables.

Nevertheless, as discussed above, one particular difficulty in QR panel data models, both for fixed and random effects models, is that the asymptotic variance of QR estimators depends on the density of the innovation term, and it is not easy to compute in practice. We have presented several procedures for estimating the variance-covariance matrix in their corresponding models, all of them with a kernel implementation whenever the density is involved. This in turn depends on the specific multi-dimensional setting. Additional research is needed to evaluate the relative performance of each procedure. By selecting a specific model depending on the dimension to be considered as fixed or random, we are in effect modeling different quantiles, that is, different models to analyze the heterogeneity of the effects of

covariates on an outcome variable. This should be guided by the specific interest of the empirical analysis, in terms of why we are studying quantile heterogeneity.

Inference procedures and confidence interval construction can be greatly simplified by using bootstrap methods. Specific designs for different QR problems may be guided by the bootstrap results developed in the mean regression case. In particular, different bootstrapping procedures for panel data models, as in Kapetanios (2008), can be easily adapted to the multi-dimensional setting. Galvao and Montes-Rojas (2015) argue that bootstrapping techniques greatly simplify the variance-covariance estimation.[4] They propose to construct confidence intervals for the parameters of interest using percentile bootstrap with pairwise resampling. In practice, FE and RE QR parameters' point estimates can thus be implemented using standard QR codes available in econometric softwares, that work for both balanced and unbalanced panel data, and different bootstrapping techniques could be adapted for either FE or RE.

As discussed above, panel data QR estimators' consistency and other asymptotic properties rely on the dimension of the heterogeneity being described, either as FE or RE. In the multi-dimensional setting, this may exponentially grow depending on the researcher's choice. If for mean-based models this is a serious issue for efficiency reasons (i.e., degrees of freedom), the asymptotic results above should suggest caution in QR models with large dimensions. In particular, smoothed and/or mixed models should be considered to reduce potential asymptotic bias. As a practical example, if the researcher can choose which dimension is potentially correlated with covariates of interest and which one is not, then the former could be considered as a fixed-parameter to be estimated, with the corresponding incidental parameter problem, and the other could be modeled as a random effect. Galvao and Poirier (2015) test for RE vs. FE models could help in this direction.

References

Abrevaya, J. and Dahl, C. M. (2008). The effects of birth inputs on birthweight: Evidence from quantile estimation on panel data. *Journal of Business and Economic Statistics*, 26:379–397.

[4] Bootstrapping techniques have been extensively used to construct confidence intervals for QR in the cross-sectional context. Buchinsky (1995) uses Monte Carlo simulation to study several estimation procedures of the asymptotic covariance matrix in quantile regression models, and the results favor the bootstrap design. Hahn (1995) shows that the construction of confidence intervals based on the QR estimators can be greatly simplified by using bootstrapping. Moreover, the confidence intervals constructed by the bootstrap percentile method have asymptotically correct coverage probabilities. Horowitz (1998) proposes bootstrap methods for median regression models. Feng et al. (2011) propose an adaptation wild bootstrap methods for QR. Wang and He (2007) develop inference procedures based on rank-score tests with RE. In the panel data FE context, Abrevaya and Dahl (2008) use bootstrapping for constructing confidence intervals in the QR panel data.

Amemiya, T. (1982). Two stage least absolute deviations estimators. *Econometrica*, 50:689–711.

Arcones, M. A. (1998). Second order representations of the least absolute deviation regression estimator. *Annals of the Institute of Statistical Mathematics*, 50:87–117.

Arellano, M. and Bonhomme, S. (2016). Nonlinear panel data estimation via quantile regressions. *Econometrics Journal*, forthcoming.

Buchinsky, M. (1995). Estimating the asymptotic covariance matrix for quantile regression models: A Monte Carlo study. *Journal of Econometrics*, 68:303–338.

Canay, I. A. (2011). A simple approach to quantile regression for panel data. *Econometrics Journal*, 14:368–386.

Chamberlain, G. (1982). Multivariate regression models for panel data. *Journal of Econometrics*, 18:5–46.

Chamberlain, G. (1984). Panel data. In Griliches, Z. and Intriligator, M. D., editors, *Handbook of Econometrics*, pages 1248–1313. Amsterdam: North-Holland.

Chernozhukov, V., Fernandez-Val, I., Hahn, J., and Newey, W. (2013). Average and quantile effects in nonseparable panel models. *Econometrica*, 81:535–580.

Dhaene, G. and Jochmans, K. (2015). Split-panel jacknife estimation of fixed-effect models. *Review of Economic Studies*. forthcoming.

Feng, X., He, X., and Hu, J. (2011). Wild bootstrap for quantile regression. *Biometrika*, 98(4):995–999.

Fernandez-Val, I. (2005). Bias correction in panel data models with individual specific parameters. Mimeo.

Galvao, A. F. (2011). Quantile regression for dynamic panel data with fixed effects. *Journal of Econometrics*, 164:142–157.

Galvao, A. F. and Kato, K. (2016). Smoothed quantile regression for panel data. *Journal of Econometrics*, 193:92–112.

Galvao, A. F., Lamarche, C., and Lima, L. (2013). Estimation of censored quantile regression for panel data with fixed effects. *Journal of the American Statistical Association*, 108:1075–1089.

Galvao, A. F. and Montes-Rojas, G. V. (2015). On bootstrap inference for quantile regression panel data: A Monte Carlo study. *Econometrics*, 3:654–666.

Galvao, A. F. and Poirier, A. (2015). Quantile regression random effects. Mimeo.

Galvao, A. F. and Wang, L. (2015). Efficient minimum distance estimator for quantile regression fixed effects panel data. *Journal of Multivariate Analysis*, 133:1–26.

Geraci, M. and Bottai, M. (2007). Quantile regression for longitudinal data using the asymmetric laplace distribution. *Biostatistics*, 8(1):140–154.

Graham, B. S., Hahn, J., Poirier, A., and Powell, J. L. (2015). A quantile correlated random coefficients panel data model. Mimeo.

Graham, B. S., Hahn, J., and Powell, J. L. (2009). The incidental parameter problem in a non-differentiable panel data model. *Economics Letters*, 105:181–182.

Hahn, J. (1995). Bootstrapping quantile regression estimators. *Econometric Theory*, 11:105–121.

Hahn, J. and Kuersteiner, G. M. (2004). Asymptotically unbiased inference for a dynamic panel model with fixed effects when both n and T are large. Mimeo.

Hahn, J. and Newey, W. (2004). Jackknife and analytical bias reduction for nonlinear panel models. *Econometrica*, 72:1295–1319.

Horowitz, J. L. (1992). A smoothed maximum score estimator for the binary response model. *Econometrica*, 60:505–531.

Horowitz, J. L. (1998). Bootstrap methods for median regression models. *Econometrica*, 66:1327–1351.

Imbens, G. W. and Newey, W. K. (2009). Identification and estimation of triangular simultaneous equation models without additivity. *Econometrica*, 77:1481–1512.

Kapetanios, G. (2008). A bootstrap procedure for panel datasets with many cross-sectional units. *Econometrics Journal*, 11:377–395.

Kato, K., Galvao, A. F., and Montes-Rojas, G. (2012). Asymptotics for panel quantile regression models with individual effects. *Journal of Econometrics*, 170:76–91.

Knight, K. (1998). Limiting distributions for l_1 regression estimators under general conditions. *Annals of Statistics*, 26:755–770.

Koenker, R. (2004). Quantile regression for longitudinal data. *Journal of Multivariate Analysis*, 91:74–89.

Koenker, R. and Bassett, G. W. (1978). Regression quantiles. *Econometrica*, 46:33–49.

Koenker, R. and Hallock, K. (2000). Quantile regression: An introduction. Manuscript, University of Illinois at Urbana-Champaign.

Lancaster, T. (2000). The incidental parameter problem since 1948. *Journal of Econometrics*, 95:391–413.

Neyman, J. and Scott, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrica*, 16:1–32.

Powell, J. L. (1986). Censored regression quantiles. *Journal of Econometrics*, 32:143–155.

Rosen, A. (2012). Set identification via quantile restrictions in short panels. *Journal of Econometrics*, 166(1):127–167.

Wang, H. and He, X. (2007). Detecting differential expressions in GeneChip microarray studies. *Journal of the American Statistical Association*, 102:104–112.

Chapter 9
Models for Spatial Panels

Julie Le Gallo and Alain Pirotte

Abstract Economic interactions in space and other forms of peer effects now receive considerable attention both from a theoretical as well as from an applied perspective, especially on panel data. Until recently, the methodologies and specifications developed are related mainly to two-dimensional approaches that refer to observations on a cross-section of households, firms, countries, etc. over several time periods. However, lots of data exhibit more complex multi-dimensional structures that could be non-hierarchical or hierarchical. The multi-dimensional models that are not necessarily connected to a hierarchical structure are described in Chaps. 11, 13 and 14. Therefore, this chapter considers the case of hierarchical multi-dimensional spatial panels. We organize all the recent literature and emphasize a range of issues pertaining to the specification, estimation, testing procedures and prediction for these models. These issues include a mixture of usual topics on panel data, i.e., the form taken by individual and temporal heterogeneity, or topics more specific to spatial econometrics, i.e., dependence among observations across space, structures of the spatial matrix, Maximum Likelihood (ML) and Generalized Method of Moments (GMM) approaches, the determination and inference of direct and indirect (or spillover) effects. Only static panel data models will be considered.

9.1 Introduction

Spatial econometrics has now reached a stage of maturity (see Anselin, 2010). While originally most of the work in this field was inspired by research questions arising in regional science and economic geography, its definition and scope have moved

Julie Le Gallo
Université Bourgogne Franche-Comté, CESAER, AgroSup Dijon and INRA, Dijon, France

Alain Pirotte
CRED, Université Paris II Panthéon-Assas and TEPP (CNRS), Paris, France

L. Matyas (ed.), *The Econometrics of Multi-dimensional Panels*, Advanced Studies in Theoretical and Applied Econometrics 50, DOI 10.1007/978-3-319-60783-2_9

to mainstream economics and other social sciences, such as sociology or political science. Economic interactions in space and other forms of peer effects now receive considerable attention both from a theoretical as well as from an applied perspective. The applications making use of spatial econometrics have risen exponentially in the last two decades.

The first generation of spatial models was derived for cross-sectional data (Elhorst, 2014, p. 2). It is only since the 2000s that an important literature (second generation models) has developed to deal with spatial panel data; see, among others, Anselin et al. (2008); LeSage and Pace (2009); Lee and Yu (2010a); Elhorst (2010b, 2014); Lee and Yu (2015) for literature reviews. In this context, the methodologies and specifications developed are mainly related to two-dimensional approaches that refer to observations on a cross-section of households, firms, countries, etc. over several time periods. However, lots of data exhibit more complex multi-dimensional structures.

Examples of such multi-dimensional structures are gravity models, where economic flows (such as trade, foreign direct investment, etc.) between spatial objects (typically countries or regions) are modelled through three-dimensional panel data models with individual and time-specific fixed effects and also possibly bilateral interaction effects or other forms of composite fixed effects. Chapter 1 provides a full account of estimation and inference issues for multi-dimensional fixed effects panel data models. An alternative approach is to focus on the covariance structures using random effects rather than fixed effects, following the philosophy set out in so-called multi-level models in statistics, extensively used in education science (see, for instance, Goldstein, 1995). Multi-dimensional panel data models with random effects are presented in Chap. 2 (see also Mátyás et al., 2012 and Pus et al., 2013 for examples). The way spatial autocorrelation and spillovers can be introduced in such models is detailed in Chap. 11. Other cases of multi-dimensional data can be found in the international economics literature, with one dimension pertaining to countries, one dimension to industries or sectors, and the last dimension to time. Unobserved heterogeneity is in these cases typically modelled with fixed effects. Chapter 14 provides an example of such a structure, focusing mainly on the impacts of information and communication technologies, R&D and market regulations on productivity growth. None of the multi-dimensional models mentioned above necessarily connected to a hierarchical structure, and are described in Chaps. 11, 13 and 14. Therefore, we will not discuss them further here.

Another interesting topic to consider arises when the multi-dimensional structure may be hierarchical. For example, house price data naturally exhibit a nested structure, typically grouped by district within counties. Other examples concern regional data that can be observed at several nested spatial scales (NUTS3 regions nested in NUTS2 regions nested in NUTS1 regions in Europe), data on firms that may be grouped by industry, or data on air pollution that may be grouped by observation station within a city, a city within a country, and by country. Corrado and Fingleton (2012) also give another example of property taxes. These examples emphasize the need to account for the nested hierarchical structure of the data.

Early work on hierarchical non-spatial panels was carried out by Fuller and Battese (1973); Montmarquette and Mahseredjian (1989); Antweiler (2001); Baltagi et al. (2001, 2002); Davis (2002), and more recently by Baltagi and Pirotte (2013). The nested structure is modelled using covariance structures with random effects. Spatial interdependence and spatial heterogeneity, which are quite naturally a feature of multi-dimensional data containing a cross-sectional dimension, complicates the specifications and estimation procedures (see, for example, Baltagi et al. (2014b), Baltagi and Pirotte (2014) or Fingleton et al. (2017), which derive spatial nested panel data models with random effects concerning only one level, the most disaggregated). Overall, with the increasing availability of spatial data observed at different levels, the field dealing with multi-dimensional nested spatial panel data is becoming an important theoretical and applied topic.

In this chapter, we indeed focus on hierarchical spatial panels. We organize all this recent literature and emphasize a range of issues related to the specification, estimation, testing procedures and prediction for these models. These issues include a mixture of usual topics on panel data, i.e., the form taken by individual and temporal heterogeneity, or topics more specific to spatial econometrics, i.e., dependence among observations across space, structures of the spatial matrix, Maximum Likelihood (ML) and Generalized Method of Moments (GMM) approaches, the determination and inference of direct and indirect (or spillover) effects. Only static panel data models will be considered. This chapter contains six sections. Sect. 9.2 presents how traditional spatial cross-section models and two-dimensional panel data models can be extended to multi-dimensional nested spatial panels, along the lines set out previously. Next, we also tackle, in particular, extensions of estimation (Sect. 9.3), testing (Sect. 9.4) and prediction (Sect. 9.5). Sect. 9.6 presents a range of special topics, namely how coefficient heterogeneity can be accounted for. Lastly, Sect. 9.7 ends with some concluding remarks.

9.2 Spatial Models

9.2.1 The Baseline Model

In this section, we focus on the most relevant three-dimensional model that combines three different types of spatial interactions effects, i.e., endogenous interaction effects (spatial lag on the dependent variable y), exogenous interaction effects (spatial lags on the explanatory variables X) and interaction effects among the disturbances (for example, using a Spatial AutoRegressive (SAR) process or a Spatial Moving Average (SMA) process on ε). Using similar notations to that in Chap. 1, we consider three-dimensional panel data where the dependent variable is observed along three indices, y_{ijt} with $i = 1,2,\ldots,N$, $j = 1,2,\ldots,M_i$ and $t = 1,2,\ldots,T$. N denotes the number of groups. M_i denotes the number of individuals in group i, so in total there are $S = \sum_{i=1}^{N} M_i$ individuals. T represents the number of periods. Since

we allow for an unequal number of individuals across the N groups, the panel may therefore be unbalanced in the individual dimension, although it is balanced in the time dimension. This kind of structure is common in practice, for example, if we group data on firms by industry over a balanced time dimension.[1] The number of observations is therefore TS.

In the case of a hierarchical structure, suppose that the index j pertains to individuals that are nested with the N groups. Assuming that spatial autocorrelation only takes place at the individual level and that the slope coefficients are homogenous, the spatial specification takes the form

$$y_{ijt} = \rho \sum_{g=1}^{N} \sum_{h=1}^{M_g} w_{ij,gh} y_{ght} + x_{ijt}\beta + \sum_{g=1}^{N} \sum_{h=1}^{M_g} w_{ij,gh} x_{ght}\theta + \varepsilon_{ijt}, \tag{9.1}$$

where y_{ijt} is the dependent variable, in which the subscript denotes the individual j in group i at time period t. x_{ijt} is a $(1 \times K)$ vector of explanatory (exogenous) variables, while β, like θ, represents a $(K \times 1)$ vector of parameters to be estimated. ε_{ijt} is the disturbance, the properties of which will be discussed below. The weight $w_{ij,gh} = w_{k,l}$ is the $(k = ij; l = gh)$ element of the spatial matrix W_S with ij denoting individual j within group i, and similarly for gh. Thus $k, l = 1, \ldots, S$ and W_S is a $(S \times S)$ known spatial weights matrix which has zero diagonal elements and is usually row-normalized so that for row k, $\sum_{g=1}^{N} \sum_{h=1}^{M_g} w_{k,gh} = 1$. Overall, W_S is assumed non-stochastic, and its row and column sums are required to be uniformly bounded in absolute value. ρ is the spatial lag parameter to be estimated. This coefficient is bounded numerically to ensure spatial stationarity, i.e., $e_{min}^{-1} < \rho < 1$ where e_{min} is the minimum real characteristic root of W_S.

Model (9.1) implies that unobservable heterogeneity is only captured through the regression constant or a composite disturbance term, see Sect. 9.2.2. Nevertheless, the heterogeneity structure could be more complex, especially on multi-dimensional panels. Sect. 9.6.1 tackles this, mainly considering simultaneously the unobserved and observed heterogeneity, i.e., relaxing the assumption of a constant regression coefficient in (9.1).

Moreover, in contrast with the usual panel data framework, the disturbance ε_{ijt} could be contemporaneously correlated. A simple and widely used approach to modelling spatial error dependence is to assume a SAR process at the individual level, such as

$$\varepsilon_{ijt} = \lambda \sum_{g=1}^{N} \sum_{h=1}^{M_g} m_{ij,gh} \varepsilon_{ght} + u_{ijt}. \tag{9.2}$$

The weight $m_{ij,gh}$ is an element of the spatial matrix M_S which satisfies the same assumptions as the one of W_S. For simplicity, we assume $M_S = W_S$. λ is the spatial autoregressive parameter to be estimated. u_{ijt} is assumed to be i.i.d.$(0, \sigma_u^2)$. For a cross-section t, Eqs. (9.1) and (9.2) can be written as

[1] This presentation is mainly connected to the existing literature. A possible further extension is to consider an unbalanced time dimension T_{ij}.

$$y_t = \rho W_S y_t + x_t \beta + W_S x_t \theta + \varepsilon_t, \tag{9.3}$$

$$\varepsilon_t = \lambda W_S \varepsilon_t + u_t. \tag{9.4}$$

Stacking the T cross-sections gives

$$y = \rho W y + X\beta + WX\theta + \varepsilon, \tag{9.5}$$

$$\varepsilon = \lambda W \varepsilon + u, \tag{9.6}$$

with y and X being the vector and matrix of the dependent and explanatory variables (covariates), respectively of size $(TS \times 1)$ and $(TS \times K)$, β, just as θ, being the vector of the slope parameters of size $(K \times 1)$, and finally, ε the vector of the disturbance terms is of size $(TS \times 1)$. Given that I_T is an identity matrix of dimension $(T \times T)$, then $W = (I_T \otimes W_S)$ of size $(TS \times TS)$. Adopting Elhorst's (2014) taxonomy, Eqs. (9.5) and (9.6), namely the General Nesting Spatial (GNS) model, includes a family of nine linear spatial econometric models. For example, if $\theta = 0$, we obtain the Spatial Autoregressive Combined (SAC) model. If the restriction $\lambda = 0$ is imposed, the specification corresponds to the Spatial Durbin (SD) model, whereas if $\rho = 0$, it is referred to as the Spatial Durbin Error (SDE) model.

Among the family of GNS models, the Spatial Lag Model is the specification assuming that $\theta = 0$ and $\lambda = 0$, and this leads to specific direct and spillover effects. To understand this, it is necessary to consider the reduced form of the Spatial Lag Model, which is given by

$$y = D^{-1}[X\beta + \varepsilon], \tag{9.7}$$

with $D^{-1} = (I_T \otimes D_S^{-1})$, where

$$D_S^{-1} = (I_S - \rho W_S)^{-1} = I_S + \rho W_S + \rho^2 W_S^2 + \rho^3 W_S^3 + \dots, \tag{9.8}$$

where $I_S = \text{diag}\,(I_{M_i})$ is an identity matrix of dimension $(S \times S)$. By $\text{diag}\,(I_{M_i})$, we mean $\text{diag}\,(I_{M_1}, \dots, I_{M_N})$, where I_{M_i} is an identity matrix of dimension $(M_i \times M_i)$. For each cross-section $(S \times 1)$ at time period t, we have

$$y_t = D_S^{-1}[x_t \beta + \varepsilon_t]. \tag{9.9}$$

This means that the spatial distribution of the dependent variable y_t in each cross-section is determined not only by the explanatory variables and their parameters at each location, but also by those at neighboring locations. A change in x at any location will affect all other locations following (9.8), even if two locations are not connected in W_S. A strong assumption induced by the Spatial Lag Model is that the ratio between the spillover (indirect) and direct effects is the same for each explanatory variable. Moreover, it is difficult to interpret a $(S \times S)$ matrix for each of the K explanatory variables of direct and spillover effects, and so it is standard practice to compute the scalar average measures as proposed by LeSage and Pace (2009). Similar derivations can be obtained for the Spatial Durbin model ($\lambda = 0$), in which case the derivation of direct and indirect effects should take into account the presence of the spatial lags of the explanatory variables.

Other spatial models (Spatial Error Model, Spatial lag of X Model, etc.) could be obtained as the result of different simultaneous restrictions on the parameters (see Elhorst (2014) for an exhaustive spatial taxonomy).

Concerning the disturbances ε and their full $(TS \times TS)$ covariance matrix, the latter takes the form

$$\Omega_\varepsilon = \mathrm{E}[\varepsilon\varepsilon'] = \sigma_u^2[(I_T \otimes B_S^{-1})(I_T \otimes B_S^{-1})'] = \sigma_u^2[I_T \otimes (B_S' B_S)^{-1}], \tag{9.10}$$

with $B_S = I_S - \lambda W_S$. The matrix B_S is nonsingular with $|\lambda| < 1$, with row and column sums assumed uniformly bounded in absolute value. The SAR process (9.6) implies complex interdependence between individuals, so that changes affecting an individual j in group i impact all other individuals regardless of the group they belong to. In other words, this process is known to transmit exogenous shocks *globally*.[2] Another spatial autocorrelation structure is possible for the error term ε, namely the Spatial Moving Average (SMA) process, which can be expressed as

$$\varepsilon = \gamma(I_T \otimes W_S)u + u, \tag{9.11}$$

or

$$\varepsilon = [I_{TS} + \gamma(I_T \otimes W_S)]u = (I_T \otimes G_S)u, \tag{9.12}$$

where $G_S = I_S + \gamma W_S$. The full $(TS \times TS)$ covariance matrix associated with this process is

$$\Omega_\varepsilon = \sigma_u^2(I_T \otimes G_S G_S') = \sigma_u^2[I_T \otimes (I_S + \gamma(W_S + W_S') + \gamma^2 W_S W_S')]. \tag{9.13}$$

In contrast with the SAR process, SMA errors transmit the shocks *locally* rather than *globally*, i.e., the covariance matrix (9.13) includes only the first two-order neighbors (see Anselin, 2003 for an extensive discussion on the interpretation of local versus global spillovers in spatial econometric models).

So far, we have introduced the spatial effects for multi-dimensional panels in the form of spatial lag on y and X and spatial processes (SAR or SMA) on the disturbances under the homogeneity assumption. The panel dimension enables us to control for time and spatial heterogeneity. The main question is how to formalize this heterogeneity. We limit our attention to spatial models with unobserved heterogeneity that could be treated as random effects or as fixed effects, see also Chaps. 1 and 2.

9.2.2 Unobserved Heterogeneity

In our context, unobserved heterogeneity can pertain either to time heterogeneity or to spatial heterogeneity.

[2] As in Eq. (9.8), the inverse of B_S is related to the infinite series $I_S + \lambda W_S + \lambda^2 W_S^2 + \lambda^3 W_S^3 + \cdots$

With respect to unobservable *time heterogeneity*, it is usual to consider it using time fixed effects (time-specific intercepts) or time random effects (time random component or a common factor). In the case of time-specific fixed in (9.1), we have

$$y_{ijt} = \rho \sum_{g=1}^{N} \sum_{h=1}^{M_g} w_{ij,gh} y_{ght} + x_{ijt}\beta + \sum_{g=1}^{N} \sum_{h=1}^{M_g} w_{ij,gh} x_{ght}\theta + \delta_t + \varepsilon_{ijt}, \tag{9.14}$$

where δ_t are the time-specific fixed effects. If T is small compared to S, they can be estimated alongside the other coefficients using the estimation methods outlined in section 9.3. These effects can also be considered as random assuming that $\mathrm{E}(\delta_t) = 0$, $\mathrm{E}(\delta_t \delta_s) = \sigma_\delta^2$ if $t = s$ and 0 otherwise. This is a special case of those considered in Chap. 2, where the corresponding variance-covariance matrices are detailed. Random effects can also appear as common factors (see Pesaran, 2015a) in the disturbances

$$u_{ijt} = \zeta_{j1} f_{1t} + \zeta_{j2} f_{2t} + \cdots + \zeta_{jm} f_{mt} + e_{ijt}, \tag{9.15}$$

in which $f_t = (f_{1t}, f_{2t}, \ldots, f_{mt})'$ is an m-dimensional vector of unobservable common factors and $\zeta_j = (\zeta_{j1}, \zeta_{j2}, \ldots, \zeta_{jm})'$ is the associated $(m, 1)$ vector or factor loadings, with m assumed to be fixed relative to S and $m << S$. The idiosyncratic errors e_{ijt} can be cross-sectionally weakly dependent, the factor loadings can be considered as draws from a random distribution, or fixed unknown coefficients. This setting can be further extended to allow for heterogenous coefficients (see Sect. 9.6).

We now consider in further detail the problem of unobserved *spatial heterogeneity*. Following Baltagi and Pirotte (2013, 2014), we assume that the random effect structure of the errors u_{ijt} contains an unobserved permanent group-specific error component α_i, a nested permanent individual-group-specific error component μ_{ij}, together with a remainder error component v_{ijt}. Hence, we envisage a time-invariant group-effect applying equally to all individuals nested within a group, time-invariant individual-group-specific effects, and transient effects that vary at random across groups, individuals and time. More formally, the disturbance term is decomposed as

$$u_{ijt} = \alpha_i + \mu_{ij} + v_{ijt}, \tag{9.16}$$

with the following assumptions: (i) α_i is an unobservable group specific time-invariant effect which is assumed to be i.i.d.$N\left(0, \sigma_\alpha^2\right)$; (ii) μ_{ij} is the nested effect of individual j within the ith group, which is assumed to be i.i.d.$N\left(0, \sigma_\mu^2\right)$; (iii) v_{ijt} is the remainder term, which is also assumed to be i.i.d.$N\left(0, \sigma_v^2\right)$; (iv) The α_i's, μ_{ij}'s and v_{ijt}'s are independent of each other and among themselves.

For a cross-section t, the standard specification of the Nested Random Effects (NRE) u_t is given by

$$u_t = \mathrm{diag}\left(\iota_{M_i}\right) \alpha + \mu + v_t, \tag{9.17}$$

where u_t is $(S \times 1)$, α is the vector of group effects of dimension $(N \times 1)$, $\mu' = (\mu_1', \ldots, \mu_N')$, a vector of dimension $(1 \times S)$, $\mu_i' = (\mu_{i1}, \ldots, \mu_{iM_i})$, a vector of dimension

$(1 \times M_i)$. By $\text{diag}(\iota_{M_i})$, we mean $\text{diag}(\iota_{M_1}, \ldots, \iota_{M_N})$, where ι_{M_i} is a vector of ones of format $(M_i \times 1)$. v_t is of dimension $(S \times 1)$. The covariance matrix of u_t is

$$\text{E}\left[u_t u_t'\right] = \sigma_\alpha^2 \text{diag}\left(J_{M_i}\right) + \left(\sigma_\mu^2 + \sigma_v^2\right) I_S, \tag{9.18}$$

where $J_{M_i} = \left(\iota_{M_i} \iota_{M_i}'\right)$ is a matrix of ones of dimension $(M_i \times M_i)$. For the full $(TS \times 1)$ vector of disturbances u, we have

$$u = \left(\iota_T \otimes \text{diag}\left(\iota_{M_i}\right)\right) \alpha + \left(\iota_T \otimes I_S\right) \mu + v, \tag{9.19}$$

where ι_T is a vector of ones of dimension $(T \times 1)$. The covariance matrix of u corresponds to

$$\begin{aligned} \Omega_u &= \sigma_\alpha^2 \left(Z_\alpha Z_\alpha'\right) + \sigma_\mu^2 \left(Z_\mu Z_\mu'\right) + \sigma_v^2 \left(I_T \otimes I_S\right) \\ &= \sigma_\alpha^2 \left(J_T \otimes \text{diag}\left(J_{M_i}\right)\right) + \left(\left(\sigma_\mu^2 J_T + \sigma_v^2 I_T\right) \otimes I_S\right), \end{aligned} \tag{9.20}$$

where $Z_\alpha = \left(\iota_T \otimes \text{diag}\left(\iota_{M_i}\right)\right)$, $Z_\mu = \left(\iota_T \otimes I_S\right)$ and $J_T = \left(\iota_T \iota_T'\right)$ is a matrix of ones of dimension $(T \times T)$. Replace J_T by its idempotent counterpart $T\overline{J}_T$, J_{M_i} by $M_i \overline{J}_{M_i}$. In addition, define $Q_T = I_T - \overline{J}_T$, and $Q_{M_i} = I_{M_i} - \overline{J}_{M_i}$, and replace I_T by $\left(Q_T + \overline{J}_T\right)$, I_{M_i} by $\left(Q_{M_i} + \overline{J}_{M_i}\right)$. Collecting terms, one gets the spectral decomposition of Ω_u, which is

$$\Omega_u = \lambda_1 \widetilde{Q}_1 + \lambda_2 \widetilde{Q}_2 + \left(I_T \otimes \text{diag}\left(\lambda_{3i} I_{M_i}\right)\right) \widetilde{Q}_3, \tag{9.21}$$

with

$$\lambda_{1i} = \lambda_1 = \sigma_v^2,\ \lambda_{2i} = \lambda_2 = T\sigma_\mu^2 + \sigma_v^2,\ \lambda_{3i} = M_i T \sigma_\alpha^2 + T\sigma_\mu^2 + \sigma_v^2, \tag{9.22}$$

$$\widetilde{Q}_1 = \left(Q_T \otimes I_S\right),\ \widetilde{Q}_2 = \left(\overline{J}_T \otimes \text{diag}\left(Q_{M_i}\right)\right), \tag{9.23}$$

$$\widetilde{Q}_3 = \left(\overline{J}_T \otimes \text{diag}\left(\overline{J}_{M_i}\right)\right), \tag{9.24}$$

and $\overline{J}_T = J_T/T$, $\overline{J}_{M_i} = J_{M_i}/M_i$. The operators $\widetilde{Q}_1$, $\widetilde{Q}_2$ and $\widetilde{Q}_3$ are symmetric and idempotent, with their rank equal to their trace. Moreover, they are pairwise orthogonal and add up to the identity matrix. From (9.21), we can easily obtain Ω_u^{-1} as

$$\Omega_u^{-1} = \lambda_1^{-1} \widetilde{Q}_1 + \lambda_2^{-1} \widetilde{Q}_2 + \left(I_T \otimes \text{diag}\left(\lambda_{3i}^{-1} I_{M_i}\right)\right) \widetilde{Q}_3. \tag{9.25}$$

Considering the SAR process (9.6) of the vector disturbances ε, we get

$$\varepsilon = \left(I_T \otimes B_S^{-1}\right) u, \tag{9.26}$$

and the corresponding $(TS \times TS)$ covariance matrix is given by

$$\Omega_\varepsilon = A \Omega_u A', \tag{9.27}$$

where A is a block-diagonal matrix equal to $\left(I_T \otimes B_S^{-1}\right)$. Following the properties of the matrices Ω_u and A, we obtain the inverse covariance matrix of ε defined as

$$\Omega_{\varepsilon}^{-1} = \left(A'\right)^{-1} \Omega_{u}^{-1} A^{-1}. \tag{9.28}$$

If an SMA process (9.11) instead of a SAR process is considered for the vector disturbances ε, the matrix A corresponds to $(I_T \otimes G_S)$ instead of $(I_T \otimes B_S^{-1})$. Combining (9.19) and (9.26) give what we call a SAR-NRE process, and the association of equations (9.19) and (9.11) corresponds to the SMA-NRE process. This means for the latter process that the covariance matrix is given by (9.27), where $A = (I_T \otimes G_S)$. An alternative approach is to assume, first, that the disturbance ε_t has a NRE structure and, second, that the error v_t follows a SAR or an SMA process, namely NRE-SAR and NRE-SMA, respectively. We will not consider this approach further in this chapter.[3]

9.3 Spatial Estimation Methods

The estimation methods of multi-dimensional spatial panel models are direct extensions of the ones that have been done for the standard spatial panel data econometrics. This means that two main approaches are used to estimate these models, one based on the Maximum Likelihood (ML) principle, the other linked to the method of moments procedures.

9.3.1 Maximum Likelihood Estimation

Upton and Fingleton (1985), Anselin (1988), LeSage and Pace (2009) and Elhorst (2014) provide the general framework for the ML estimation of spatial models. Under normality of the disturbances, the log-likelihood function is proportional to

$$\begin{aligned} \ln L = & -\frac{TS}{2}\ln(2\pi) - \frac{1}{2}\ln|\Omega_{\varepsilon}| + T\ln|D_S| \\ & -\frac{1}{2}(Dy - X\beta - WX\theta)'\,\Omega_{\varepsilon}^{-1}\,(Dy - X\beta - WX\theta). \end{aligned} \tag{9.29}$$

If we consider a SAR-NRE process for the disturbances ε, after some mathematical manipulations, we obtain

$$\begin{aligned} \ln L = & -\frac{TS}{2}\ln(2\pi) - \frac{1}{2}\ln|\Omega_{u}| + T\ln|B_S| + T\ln|D_S| \\ & -\frac{1}{2}(Dy - X\beta - WX\theta)'\,\Omega_{\varepsilon}^{-1}\,(Dy - X\beta - WX\theta). \end{aligned} \tag{9.30}$$

[3] See Baltagi et al. (2013) and Baltagi and Liu (2016), for a discussion on a generalized spatial model that encompasses the NRE-SAR and the SAR-NRE models.

Let $\gamma_1 = \sigma_\alpha^2/\sigma_v^2$, $\gamma_2 = \sigma_\mu^2/\sigma_v^2$ and $\Omega_\varepsilon = \sigma_v^2\Sigma$, then the log-likelihood function (9.30) can be written as[4]

$$\begin{aligned}\ln L = &-\frac{TS}{2}\ln(2\pi) - \frac{TS}{2}\ln\sigma_v^2 - \frac{1}{2}\sum_{i=1}^{N}\ln\left(T\left(M_i\gamma_1+\gamma_2\right)+1\right)\\ &-\frac{1}{2}\sum_{i=1}^{N}\left(M_i-1\right)\ln\left(T\gamma_2+1\right)\\ &+T\sum_{i=1}^{N}\sum_{j=1}^{M_i}\ln\left(1-\omega_{ij}\lambda\right)+T\sum_{i=1}^{N}\sum_{j=1}^{M_i}\ln\left(1-\eta_{ij}\rho\right)\\ &-\frac{1}{2\sigma_v^2}\left(Dy-X\beta-WX\theta\right)'\Sigma^{-1}\left(Dy-X\beta-WX\theta\right), \end{aligned} \tag{9.31}$$

where ω_{ij}'s and η_{ij}'s are the eigenvalues of M_S and W_S respectively.

For a SMA-NRE process for the disturbances ε, the log-likelihood takes the following form

$$\begin{aligned}\ln L = &-\frac{TS}{2}\ln(2\pi) - \frac{1}{2}\ln|\Omega_u| - T\ln|G_S| + T\ln|D_S|\\ &-\frac{1}{2}\left(Dy-X\beta-WX\theta\right)'\Omega_\varepsilon^{-1}\left(Dy-X\beta-WX\theta\right). \end{aligned} \tag{9.32}$$

Using the same notations as before, this can be written as

$$\begin{aligned}\ln L = &-\frac{TS}{2}\ln(2\pi) - \frac{TS}{2}\ln\sigma_v^2 - \frac{1}{2}\sum_{i=1}^{N}\ln\left(T\left(M_i\gamma_1+\gamma_2\right)+1\right)\\ &-\frac{1}{2}\sum_{i=1}^{N}\left(M_i-1\right)\ln\left(T\gamma_2+1\right)\\ &-T\sum_{i=1}^{N}\sum_{j=1}^{M_i}\ln\left(1+\omega_{ij}\gamma\right)+T\sum_{i=1}^{N}\sum_{j=1}^{M_i}\ln\left(1-\eta_{ij}\rho\right)\\ &-\frac{1}{2\sigma_v^2}\left(Dy-X\beta-WX\theta\right)'\Sigma^{-1}\left(Dy-X\beta-WX\theta\right). \end{aligned} \tag{9.33}$$

The first-order conditions for the parameters in (9.31) and (9.33) are intertwined, which means that they are non-linear, i.e., the equations cannot be solved analytically. Therefore, a numerical solution by means of an iterative procedure is needed in the spirit of Anselin (1988). For a SAR-NRE process with $\rho \neq 0$, $\theta = 0$ and $\lambda = 0$, Baltagi et al. (2014b) give the general ML framework approach and propose

[4] Grouping the data by units rather than periods, Baltagi et al. (2001) have shown that the covariance matrix of u is given by $\Omega_u^p = \mathrm{diag}\left(\Lambda_i^p\right) = \mathrm{diag}\left(\lambda_{1i}^p\widetilde{Q}_{1i} + \lambda_{2i}^p\widetilde{Q}_{2i} + \lambda_{3i}^p\widetilde{Q}_{3i}\right)$, where $\widetilde{Q}_{1i} = (I_{M_i}\otimes Q_T)$, $\widetilde{Q}_{2i} = \left(Q_{M_i}\otimes\overline{J}_T\right)$, $\widetilde{Q}_{3i} = \left(\overline{J}_{M_i}\otimes\overline{J}_T\right)$ and $\lambda_{pi,}$, for $p = 1,2,3$, are the distinct characteristic roots of $\Lambda_i = \lambda_{1i}\widetilde{Q}_{1i} + \lambda_{2i}\widetilde{Q}_{2i} + \lambda_{3i}\widetilde{Q}_{3i}$ then $|\Lambda_i| = (\lambda_{3i})\left(\lambda_{2i}^{M_i-1}\right)\left(\lambda_{1i}^{M_i(T-1)}\right)$.

an Instrumental Variables (IV) approach due to the correlation between the spatial lag of the dependent variable and disturbances (see Sect. 9.3.2). If $\rho = 0$, $\theta = 0$ and $\lambda \neq 0$, Baltagi and Pirotte (2014) describe an iterative procedure to obtain the ML estimates.

9.3.2 GMM, FGLS and Instrumental Variables Approaches

One method of estimating panel data models with spatially dependent nested random effects is by means of ML, as we have seen above. However, with limited computing power, which has often been the case in the past and may still be a problem in the future for some researchers, ML procedures are likely to be impractical when the individual sample size is exceptionally large, and sample sizes are increasing, perhaps exponentially, as we enter an era of "big data". Finally, ML calls for explicit distributional assumptions, which may be difficult to satisfy, although Quasi-ML (QML) approaches may to some extent allay this problem, and specifying and maximizing likelihood functions appropriate to extensions to more complex models may be problematic. In view of the desirability of estimation approaches that avoid some of the challenges posed by ML, Kelejian and Prucha (1998, 1999) suggested an alternative instrumental variable estimation procedure for the spatial lag model, also including a SAR process for the disturbances. This approach is based on a Generalized Method of Moments (GMM) estimator of the parameter in the spatial autoregressive process. The procedures suggested in Kelejian and Prucha (1998, 1999) are computationally feasible even for very large sample sizes.[5] As in most of the spatial literature, they consider the case where a single cross section of data is available. Monte Carlo results in Das et al. (2003) suggest that both the GMM and the instrumental variable estimators are as efficient as the corresponding ML estimators in small samples. Alternatively, in a panel data context, Fingleton et al. (2017) take advantage of the Kapoor et al. (2007) (hereafter KKP) Generalized Method of Moments (GMM) estimator, which is computationally feasible even for large sample sizes, extending this procedure to capture nested spatial random effects.

Assuming that $\rho = 0$ and $\theta = 0$ in a SAR-NRE model, Fingleton et al. (2017) develop a GMM approach leading to estimators for λ, σ_α^2, σ_μ^2, σ_v^2, or equivalently of λ, σ_α^2, $\lambda_2 \left(= T\sigma_\mu^2 + \sigma_v^2\right)$ and σ_v^2, relying on appropriate moment conditions. For notational convenience, we have

$$\overline{\varepsilon} = (I_T \otimes W_S)\,\varepsilon, \tag{9.34}$$

$$\overline{\overline{\varepsilon}} = (I_T \otimes W_S)\,\overline{\varepsilon}, \tag{9.35}$$

$$\overline{u} = (I_T \otimes W_S)\,u. \tag{9.36}$$

[5] Kelejian and Prucha (1998, 1999) use the terminology Generalized Moments (GM) for GMM, we however, will stick to GMM throughout the chapter.

The GMM estimators are defined in terms of nine moment conditions. Fingleton et al. (2017) demonstrate that we have

$$\mathrm{E}\begin{bmatrix} \frac{1}{S(T-1)}u'\widetilde{Q}_1u \\ \frac{1}{S(T-1)}\overline{u}'\widetilde{Q}_1\overline{u} \\ \frac{1}{S(T-1)}\overline{u}'\widetilde{Q}_1u \\ \frac{1}{(S-N)}u'\widetilde{Q}_2u \\ \frac{1}{(S-N)}\overline{u}'\widetilde{Q}_2\overline{u} \\ \frac{1}{(S-N)}\overline{u}'\widetilde{Q}_2u \\ \frac{1}{T}u'\widetilde{Q}_3u \\ \frac{1}{T}\overline{u}'\widetilde{Q}_3\overline{u} \\ \frac{1}{T}\overline{u}'\widetilde{Q}_3u \end{bmatrix} = \begin{bmatrix} \sigma_v^2 \\ \sigma_v^2\frac{1}{S}\mathrm{tr}\left(W_S'W_S\right) \\ 0 \\ \lambda_2 \\ \lambda_2\frac{1}{(S-N)}\mathrm{tr}\left(W_S^{\bullet\prime}W_S^{\bullet}\right)+\frac{T}{(S-N)}\sigma_\alpha^2\mathrm{tr}\left(\Gamma W_S^{\bullet\prime}W_S^{\bullet}\right) \\ \lambda_2\frac{1}{(S-N)}\mathrm{tr}\left(W_S^{\bullet\prime}\right) \\ \frac{N}{T}\lambda_2+S\sigma_\alpha^2 \\ \lambda_2\frac{1}{T}\mathrm{tr}\left(W_S^{*\prime}W_S^{*}\right)+\sigma_\alpha^2\mathrm{tr}\left(\Gamma W_S^{*\prime}W_S^{*}\right) \\ \lambda_2\frac{1}{T}\mathrm{tr}\left(W_S^{*\prime}\right)+\sigma_\alpha^2\mathrm{tr}\left(\Gamma W_S'\right) \end{bmatrix}, \quad (9.37)$$

where $W_S^* = \mathrm{diag}\left(\overline{J}_{M_i}\right)W_S$, $W_S^\bullet = \mathrm{diag}\left(Q_{M_i}\right)W_S$ and $\Gamma = \mathrm{diag}\left(J_{M_i}\right)$. The GMM estimators of λ, σ_α^2, λ_2, σ_v^2 are based on these moments. Stacking the T cross-sections (9.4) and using (9.29), (9.35), (9.36), we obtain

$$u = \varepsilon - \lambda\overline{\varepsilon}, \quad (9.38)$$

$$\overline{u} = \varepsilon - \lambda\overline{\overline{\varepsilon}}. \quad (9.39)$$

Replacing u and $\overline{u}$ with their expressions (9.38), (9.39) into (9.37), we obtain a system of nine equations involving the second moments of ε, $\overline{\varepsilon}$ and $\overline{\overline{\varepsilon}}$. This system includes λ, σ_α^2, λ_2 and σ_v^2 and can be written as

$$\Lambda\left[\lambda, \lambda^2, \sigma_v^2, \lambda_2, \sigma_\alpha^2\right]' - \gamma = 0, \quad (9.40)$$

where

$$\Lambda = \begin{bmatrix} \gamma_{11}^1 & \gamma_{12}^1 & \gamma_{13}^1 & 0 & 0 \\ \gamma_{21}^1 & \gamma_{22}^1 & \gamma_{23}^1 & 0 & 0 \\ \gamma_{31}^1 & \gamma_{32}^1 & \gamma_{33}^1 & 0 & 0 \\ \gamma_{11}^2 & \gamma_{12}^2 & 0 & \gamma_{14}^2 & 0 \\ \gamma_{21}^2 & \gamma_{22}^2 & 0 & \gamma_{24}^2 & \gamma_{25}^2 \\ \gamma_{31}^2 & \gamma_{32}^2 & 0 & \gamma_{34}^2 & 0 \\ \gamma_{11}^3 & \gamma_{12}^3 & 0 & \gamma_{14}^3 & \gamma_{15}^3 \\ \gamma_{21}^3 & \gamma_{22}^3 & 0 & \gamma_{24}^3 & \gamma_{25}^3 \\ \gamma_{31}^3 & \gamma_{32}^3 & 0 & \gamma_{34}^3 & \gamma_{35}^3 \end{bmatrix}, \gamma = \begin{bmatrix} \gamma_1^1 \\ \gamma_2^1 \\ \gamma_3^1 \\ \gamma_1^2 \\ \gamma_2^2 \\ \gamma_3^2 \\ \gamma_1^3 \\ \gamma_2^3 \\ \gamma_3^3 \end{bmatrix}, \quad (9.41)$$

and

$$\gamma_{11}^1 = \frac{2}{S(T-1)}\mathrm{E}\left[\overline{\varepsilon}'\widetilde{Q}_1\varepsilon\right], \gamma_{12}^1 = -\frac{1}{S(T-1)}\mathrm{E}\left[\overline{\varepsilon}'\widetilde{Q}_1\overline{\varepsilon}\right], \gamma_{13}^1 = 1,$$

$$\gamma_{21}^1 = \frac{2}{S(T-1)}\mathrm{E}\left[\overline{\overline{\varepsilon}}'\widetilde{Q}_1\overline{\varepsilon}\right], \gamma_{22}^1 = -\frac{1}{S(T-1)}\mathrm{E}\left[\overline{\overline{\varepsilon}}'\widetilde{Q}_1\overline{\overline{\varepsilon}}\right],$$

$$\gamma_{23}^1 = \frac{1}{S}\mathrm{tr}\left(W_S'W_S\right),$$

$$\gamma_{31}^1 = \frac{1}{S(T-1)}\mathrm{E}\left[\overline{\overline{\varepsilon}}'\widetilde{Q}_1\varepsilon+\overline{\varepsilon}'\widetilde{Q}_1\overline{\varepsilon}\right], \gamma_{32}^1 = -\frac{1}{S(T-1)}\mathrm{E}\left[\overline{\overline{\varepsilon}}'\widetilde{Q}_1\overline{\varepsilon}\right], \gamma_{33}^1 = 0,$$

$$\gamma_{11}^2 = \frac{2}{(S-N)}\mathrm{E}\left[\overline{\varepsilon}'\widetilde{Q}_2\varepsilon\right], \gamma_{12}^2 = -\frac{1}{(S-N)}\mathrm{E}\left[\overline{\varepsilon}'\widetilde{Q}_2\overline{\varepsilon}\right], \gamma_{14}^2 = 1,$$

$$\gamma_{21}^2 = \frac{2}{(S-N)}\mathrm{E}\left[\overline{\overline{\varepsilon}}'\widetilde{Q}_2\overline{\varepsilon}\right], \gamma_{22}^2 = -\frac{1}{(S-N)}\mathrm{E}\left[\overline{\overline{\varepsilon}}'\widetilde{Q}_2\overline{\overline{\varepsilon}}\right],$$

$$\gamma_{24}^2 = \frac{1}{(S-N)}\mathrm{tr}\left(W_S^{\bullet\prime}W_S^{\bullet}\right), \gamma_{25}^2 = \frac{T}{(S-N)}\mathrm{tr}\left(\Gamma W_S^{\bullet\prime}W_S^{\bullet}\right),$$

$$\gamma_{31}^2 = \frac{1}{(S-N)}\mathrm{E}\left[\overline{\overline{\varepsilon}}'\widetilde{Q}_2\varepsilon+\overline{\varepsilon}'\widetilde{Q}_2\overline{\varepsilon}\right], \gamma_{32}^2 = -\frac{1}{(S-N)}\mathrm{E}\left[\overline{\overline{\varepsilon}}'\widetilde{Q}_2\overline{\varepsilon}\right],$$

$$\gamma_{34}^2 = \frac{1}{(S-N)}\mathrm{tr}\left(W_S^{\bullet\prime}\right),$$

$$\gamma_{11}^3 = \frac{2}{T}\mathrm{E}\left[\overline{\varepsilon}'\widetilde{Q}_3\varepsilon\right], \gamma_{12}^3 = -\frac{1}{T}\mathrm{E}\left[\overline{\varepsilon}'\widetilde{Q}_3\overline{\varepsilon}\right], \gamma_{14}^3 = \frac{N}{T}, \gamma_{15}^3 = S,$$

$$\gamma_{21}^3 = \frac{2}{T}\mathrm{E}\left[\overline{\overline{\varepsilon}}'\widetilde{Q}_3\overline{\varepsilon}\right], \gamma_{22}^3 = -\frac{1}{T}\mathrm{E}\left[\overline{\overline{\varepsilon}}'\widetilde{Q}_3\overline{\overline{\varepsilon}}\right],$$

$$\gamma_{24}^3 = \frac{1}{T}\mathrm{tr}\left(W_S^{*\prime}W_S^{*}\right), \gamma_{25}^3 = \mathrm{tr}\left(\Gamma W_S^{*\prime}W_S^{*}\right),$$

$$\gamma_{31}^3 = \frac{1}{T}\mathrm{E}\left[\overline{\overline{\varepsilon}}'\widetilde{Q}_3\varepsilon+\overline{\varepsilon}'\widetilde{Q}_3\overline{\varepsilon}\right], \gamma_{32}^3 = -\frac{1}{T}\mathrm{E}\left[\overline{\overline{\varepsilon}}'\widetilde{Q}_3\overline{\varepsilon}\right],$$

$$\gamma_{34}^3 = \frac{1}{T}\mathrm{tr}\left(W_S^{*\prime}\right), \gamma_{35}^3 = \mathrm{tr}\left(\Gamma W_S'\right),$$

$$\gamma_1^1 = \frac{1}{S(T-1)}\mathrm{E}\left[\varepsilon'\widetilde{Q}_1\varepsilon\right], \gamma_2^1 = \frac{1}{S(T-1)}\mathrm{E}\left[\overline{\varepsilon}'\widetilde{Q}_1\overline{\varepsilon}\right],$$

$$\gamma_3^1 = \frac{1}{S(T-1)}\mathrm{E}\left[\varepsilon'\widetilde{Q}_1\overline{\varepsilon}\right],$$

$$\gamma_1^2 = \frac{1}{(S-N)}\mathrm{E}\left[\varepsilon'\widetilde{Q}_2\varepsilon\right], \gamma_2^2 = \frac{1}{(S-N)}\mathrm{E}\left[\overline{\varepsilon}'\widetilde{Q}_2\overline{\varepsilon}\right],$$

$$\gamma_3^2 = \frac{1}{(S-N)}\mathrm{E}\left[\varepsilon'\widetilde{Q}_2\overline{\varepsilon}\right],$$

$$\gamma_1^3 = \frac{1}{T}\mathrm{E}\left[\varepsilon'\widetilde{Q}_3\varepsilon\right], \gamma_2^3 = \frac{1}{T}\mathrm{E}\left[\overline{\varepsilon}'\widetilde{Q}_3\overline{\varepsilon}\right] \text{ and } \gamma_3^3 = \frac{1}{T}\mathrm{E}\left[\varepsilon'\widetilde{Q}_3\overline{\varepsilon}\right].$$

The equations underlying these GMM procedures are the sample counterparts to the nine equations in (9.40) based on the appropriate residuals. We can observe that the first three equations in (9.40) do not include the parameters λ_2 and σ_α^2, while the last six do not include σ_v^2. This means that the GMM estimators of λ and σ_v^2 can be obtained using only the first three moments. These are given by

$$\Lambda^{\bullet}\left[\lambda, \lambda^2, \sigma_v^2\right]' - \gamma^{\bullet} = 0, \tag{9.42}$$

where

$$\Lambda^{\bullet}=\begin{bmatrix}\gamma_{11}^{1} & \gamma_{12}^{1} & \gamma_{13}^{1}\\ \gamma_{21}^{1} & \gamma_{22}^{1} & \gamma_{23}^{1}\\ \gamma_{31}^{1} & \gamma_{32}^{1} & \gamma_{33}^{1}\end{bmatrix} \text{ and } \gamma^{\bullet}=\begin{bmatrix}\gamma_{1}^{1}\\ \gamma_{2}^{1}\\ \gamma_{3}^{1}\end{bmatrix}. \tag{9.43}$$

Given $\widehat{\lambda}$ and $\widehat{\sigma}_v^2$, it is possible to estimate λ_2 and σ_α^2 from the fourth and seventh moment conditions respectively. In the case of a spatially non-nested random effects model, KKP (2007, p. 108 and Appendix A) established the consistency of $\widehat{\lambda}$, $\widehat{\sigma}_v^2$ and $\widehat{\lambda}_2$. They also showed that their RMSEs are close to those of the weighted and partially weighted GMM approaches. This suggests that this consistency remains valid in our case. Using Monte Carlo simulations, Fingleton et al. (2017) obtain similar results to those of KKP suggesting that this GMM estimator performs well. This estimator is called the unweighted GMM estimator.

However, the literature on GMM estimators indicates that it is optimal to use the inverse of the variance-covariance matrix of the sample moments at the true parameter values as a weighting matrix to obtain asymptotic efficiency. Ξ is a function of the variances σ_α^2, σ_μ^2 and σ_v^2, which are unobserved. The consistent unweighted GMM estimators $\widehat{\sigma}_\alpha^2$, $\widehat{\sigma}_\mu^2$ and $\widehat{\sigma}_v^2$ are used to obtain a consistent estimator of Ξ, i.e., $\widehat{\Xi}$. Thus, our second GMM estimator, called the weighted GMM estimator, is defined as the nonlinear least squares estimator based on the sample counterparts of (9.40) with the sample moments weighted by $\widehat{\Xi}^{-1}$

$$\left(\widetilde{\lambda},\widetilde{\sigma}_v^2,\widetilde{\lambda}_2,\widetilde{\sigma}_\alpha^2\right)=\arg\min\left\{\xi\left(\lambda,\sigma_v^2,\lambda_2,\sigma_\alpha^2\right)'\widehat{\Xi}^{-1}\xi\left(\lambda,\sigma_v^2,\lambda_2,\sigma_\alpha^2\right)\right\}. \tag{9.44}$$

Following KKP (2007, p. 109), it is apparent that the consistency of this estimator remains valid in the case of the nested random effects model.

9.3.2.1 The GMM Spatial FGLS Estimator

To estimate β, it is necessary to use a Feasible Generalized Least Squares (FGLS), namely GMM-S-FGLS estimator. One first calculates the unweighted GMM estimates of $\lambda,\sigma_v^2,\lambda_2$ and σ_α^2, following a two-stage procedure:

- First, the GMM estimators of λ, σ_α^2, λ_2, σ_v^2 are computed from the Ordinary Least Squares (OLS) residuals. Given exogenous regressors, the OLS estimator defined by $\widehat{\beta}_{OLS}\left(=(X'X)^{-1}X'y\right)$ is consistent, and thus the OLS estimated errors $\widehat{\varepsilon}_t=y_t-x_t\widehat{\beta}_{OLS}$ are consistent estimates. Then the sample counterpart of equation (9.42) in terms of $\widetilde{\varepsilon}$, $\widetilde{\widetilde{\varepsilon}}$ and $\widetilde{\widetilde{\widetilde{\varepsilon}}}$ is

$$\widetilde{\Lambda}^{\bullet}\left[\lambda,\lambda^2,\sigma_v^2\right]'-\widetilde{\gamma}^{\bullet}=\xi\left(\lambda,\sigma_v^2\right), \tag{9.45}$$

where $\widetilde{\Lambda}^{\bullet}$, $\widetilde{\gamma}^{\bullet}$ are the sample counterparts of (9.43), and $\xi\left(\lambda,\sigma_v^2\right)$ is a vector of residuals. The unweighted GMM estimators of λ and σ_v^2 are the nonlinear least squares estimators based on (9.45)

$$\left(\widehat{\lambda},\widehat{\sigma}_v^2\right)=\arg\min\left\{\xi\left(\lambda,\sigma_v^2\right)'\xi\left(\lambda,\sigma_v^2\right)\right\}. \tag{9.46}$$

Given $\widehat{\lambda}$ and $\widehat{\sigma}_v^2$, λ_2 is estimated using

$$\begin{aligned}\widehat{\lambda}_2&=\frac{1}{(S-N)}\left(\widetilde{\varepsilon}-\widehat{\lambda}\widetilde{\overline{\varepsilon}}\right)'\widetilde{Q}_2\left(\widetilde{\varepsilon}-\widehat{\lambda}\widetilde{\overline{\varepsilon}}\right)\\&=\widetilde{\gamma}_1^2-\widehat{\lambda}\widetilde{\gamma}_{11}^2-\widehat{\lambda}^2\widetilde{\gamma}_{12}^2,\end{aligned} \tag{9.47}$$

and $\widehat{\sigma}_\alpha^2$ by

$$\begin{aligned}\widehat{\sigma}_\alpha^2&=\frac{1}{ST}\left(\widetilde{\varepsilon}-\widehat{\lambda}\widetilde{\overline{\varepsilon}}\right)'\widetilde{Q}_3\left(\widetilde{\varepsilon}-\widehat{\lambda}\widetilde{\overline{\varepsilon}}\right)-\frac{N}{ST}\widehat{\lambda}_2\\&=\frac{1}{S}\left[\left(\widetilde{\gamma}_1^3-\widehat{\lambda}\widetilde{\gamma}_{11}^3-\widehat{\lambda}^2\widetilde{\gamma}_{12}^3\right)-\frac{N}{T}\left(\widetilde{\gamma}_1^2-\widehat{\lambda}\widetilde{\gamma}_{11}^2-\widehat{\lambda}^2\widetilde{\gamma}_{12}^2\right)\right].\end{aligned} \tag{9.48}$$

Using $\widehat{\sigma}_v^2$ and (9.48), we obtain

$$\widehat{\sigma}_\mu^2=\frac{1}{T}\left(\widehat{\lambda}_2-\widehat{\sigma}_v^2\right). \tag{9.49}$$

- In a second stage, we need the estimated variance-covariance matrix $\widehat{\Omega}_u$ obtained from the first stage estimates $\widehat{\sigma}_v^2$, $\widehat{\sigma}_\mu^2$, $\widehat{\sigma}_\alpha^2$ using (9.25). In order to obtain an equation in terms of u, from which spatial autocorrelation is absent, rather than considering the SAR error process of ε, we can purge the equation of spatial dependence by pre-multiplication by $\left[I_T\otimes\left(I_S-\widehat{\lambda}W_S\right)\right]$. This can be seen to be a type of Cochrane-Orcutt transformation appropriate to spatially dependent data. Hence, pre-multiplication of the model by $\left[I_T\otimes\left(I_S-\widehat{\lambda}W_S\right)\right]$ yields

$$y^*\left(\widehat{\lambda}\right)=X^*\left(\widehat{\lambda}\right)\beta+u, \tag{9.50}$$

where

$$X^*\left(\widehat{\lambda}\right)=\left[I_T\otimes\left(I_S-\widehat{\lambda}W_S\right)\right]X, \tag{9.51}$$

$$y^*\left(\widehat{\lambda}\right)=\left[I_T\otimes\left(I_S-\widehat{\lambda}W_S\right)\right]y. \tag{9.52}$$

However, a convenient way of computing $\widehat{\beta}_{GMM-S-FGLS}$, the GMM-S-FGLS estimator, is to use $\widehat{\Omega}_\varepsilon^{-1}$, which is derived as a function of $\widehat{\Omega}_u^{-1}$, as shown by (9.54). If we are guided by the classical panel data random effects literature (see Baltagi, 2013), and transform the model in (9.50) by premultiplying it by $\widehat{\Omega}_u^{-1/2}$, the OLS

estimator of β computed from the resulting transformed model is identical to the GMM-S-FGLS estimator $\widehat{\beta}_{GMM-S-FGLS}$. This latter estimator is given by

$$\begin{aligned}\widehat{\beta}_{GMM-S-FGLS} &= \left(X'\widehat{\Omega}_{\varepsilon}^{-1}X\right)^{-1}X'\widehat{\Omega}_{\varepsilon}^{-1}y \\ &= \left(X^*\left(\widehat{\lambda}\right)'\widehat{\Omega}_{u}^{-1}X^*\left(\widehat{\lambda}\right)\right)^{-1}X^*\left(\widehat{\lambda}\right)'\widehat{\Omega}_{u}^{-1}y^*\left(\widehat{\lambda}\right), \quad (9.53)\end{aligned}$$

in which

$$\widehat{\Omega}_{\varepsilon}^{-1} = \left(\widehat{A}'\right)^{-1}\widehat{\Omega}_{u}^{-1}\widehat{A}^{-1}, \quad (9.54)$$

where $\widehat{A} = \left(I_T \otimes \left(I_S - \widehat{\lambda}W_S\right)^{-1}\right)$.

Applying the weighted GMM estimators, we initially have to use the unweighted GMM estimates of step 1 above to construct the variance-covariance matrix $\widehat{\Xi}$. Then, nonlinear least squares is used to solve

$$\left(\widetilde{\lambda}, \widetilde{\sigma}_{\nu}^2, \widetilde{\lambda}_2, \widetilde{\sigma}_{\alpha}^2\right) = \arg\min\left\{\xi\left(\lambda, \sigma_{\nu}^2, \lambda_2, \sigma_{\alpha}^2\right)'\widehat{\Xi}^{-1}\xi\left(\lambda, \sigma_{\nu}^2, \lambda_2, \sigma_{\alpha}^2\right)\right\}. \quad (9.55)$$

They provide the weighted GMM estimators of $\lambda, \sigma_{\nu}^2, \lambda_2$ and σ_{α}^2, which are denoted by $\widetilde{\lambda}$, $\widetilde{\sigma}_{\nu}^2$, $\widetilde{\lambda}_2$, $\widetilde{\sigma}_{\alpha}^2$. First one replaces λ, σ_{ν}^2, λ_2, σ_{α}^2 with their weighted GMM counterpart estimates $\widetilde{\lambda}$, $\widetilde{\sigma}_{\nu}^2$, $\widetilde{\lambda}_2$, $\widetilde{\sigma}_{\alpha}^2$ to give $\widehat{\Omega}_u$ and hence $\widehat{\Omega}_{\varepsilon}^{-1}$. Then, one applies (9.53) and (9.54) to obtain estimates of $\widehat{\beta}_{GMM-S-FGLS}$.

If $\rho \neq 0$ and $\theta = 0$, the approach proposed by Fingleton et al. (2017) could be applied using the IV estimator in combination with GMM and GLS. In the spirit of Fingleton (2008), instead of using in the first step procedure the OLS residuals, it is necessary to use the IV residuals. This is due to the spatially lag dependent variable Wy, which is always correlated with the disturbances and which implies that neither OLS nor FGLS estimators will be consistent. For a spatial cross-section, Kelejian and Prucha (1998) suggest a Two-Stage Least Squares spatial estimator (S2SLS) for the spatial lag model. They propose that the instrument set should be kept at a low order to avoid linear dependence and retain full column rank for the matrix of instruments, and thus recommend using $[X, WX]$, if the number of regressors is large. However, inclusion of further spatial lags of the explanatory variables ($\theta \neq 0$) could have a major impact on the performance of the estimation procedures set forth. Pace et al. (2012) show that instrumental variables estimation suffers greatly in situations where spatial lags of the explanatory variables (WX) are included in the model specification. The reason is that this requires the use of $\left[W^2X, W^3X, W^4X, \ldots\right]$ as instruments, in place of the conventional instruments that rely on WX, and this appears to result in a weak instruments problem.[6] However, invoking an SMA error process potentially avoids this problem. If this embodies the same W matrix, then

[6] Alternatively, in cross-sectional models, one can use the optimal instruments proposed by Lee (2003) and extended to a panel spatial lag model with random effects by Baltagi and Liu (2011).

we might assume that omitted spatial lags of explanatory variables, so-called local spillovers, are implicitly embodied within the (local) error process. This means we can use the recommended instrument set without having exogenous spatial lags among the set of regressors. Instead, we assume that the disturbances are characterized by a SMA-RE structure which purposefully captures these local spillovers. In the context of a panel nested model, Baltagi et al. (2014b) present an IV approach that could be used to obtain a correct estimation of residuals. The latter are used to obtain estimates for λ, σ_α^2, σ_μ^2 and σ_v^2. In a second step, an IV estimator under non-spherical disturbances (Cochrane-Orcutt-type transformation with an IV approach) is used to obtain consistent estimates of the parameters.

9.4 Testing for Spatial Dependence

Lagrange Multiplier (LM) tests have a long tradition in spatial econometrics. They are typically used to help specify the model in a specific-to-general procedure and are convenient to implement as they are based on the residuals of the model under the null. He and Lin (2015) consider a multi-dimensional spatial model where the disturbances are spatially autocorrelated (9.2), whereas the remainder term has a nested error component structure (9.16) assuming that $\theta = 0$, proposing various standard LM statistics linked to joint and conditional LM tests. For example, consider the joint LM test: $H_0^1 : \rho = \lambda = \sigma_\alpha^2 = \sigma_\mu^2 = 0$ vs. H_1^1 : At least one of them is not zero. The associated LM_1 statistic is

$$\begin{aligned} LM_1 = {} & \widehat{\eta}_{\lambda\lambda}\widehat{s}_\lambda^2 + 2\widehat{\eta}_{\rho\lambda}\widehat{s}_\rho\widehat{s}_\lambda + 2\widehat{\eta}_{\lambda\sigma_\alpha^2}\widehat{s}_\lambda\widehat{s}_{\sigma_\alpha^2} + 2\widehat{\eta}_{\lambda\sigma_\mu^2}\widehat{s}_\lambda\widehat{s}_{\sigma_\mu^2} \\ & + \widehat{\eta}_{\rho\rho}\widehat{s}_\rho^2 + 2\widehat{\eta}_{\rho\sigma_\alpha^2}\widehat{s}_\rho\widehat{s}_{\sigma_\alpha^2} + 2\widehat{\eta}_{\rho\sigma_\mu^2}\widehat{s}_\rho\widehat{s}_{\sigma_\mu^2} \\ & + \widehat{\eta}_{\sigma_\alpha^2\sigma_\alpha^2}\widehat{s}_{\sigma_\alpha^2}^2 + 2\widehat{\eta}_{\sigma_\mu^2\sigma_\alpha^2}\widehat{s}_{\sigma_\mu^2}\widehat{s}_{\sigma_\alpha^2} + \widehat{\eta}_{\sigma_\mu^2\sigma_\mu^2}\widehat{s}_{\sigma_\mu^2}^2, \end{aligned} \tag{9.56}$$

where $\widehat{s}_\lambda$, $\widehat{s}_\rho$, $\widehat{s}_{\sigma_\alpha^2}$, $\widehat{s}_{\sigma_\mu^2}$ and $\widehat{s}_{\sigma_v^2}$ are defined as

$$s_\lambda = -\mathrm{tr}(WA) + \varepsilon' W' \Omega_\varepsilon^{-1} \varepsilon, \tag{9.57}$$

$$s_\rho = -\mathrm{tr}(WD^{-1}) + y' W' \Omega_\varepsilon^{-1} \varepsilon, \tag{9.58}$$

$$s_{\sigma_\alpha^2} = -\frac{T}{2}\left[\sum_{i=1}^{N} \frac{M_i}{\lambda_{3i}} - \varepsilon'(A')^{-1}\left[\overline{J}_T \otimes \mathrm{diag}(M_i \lambda_{3i}^{-2} \overline{J}_{M_i})\right] A^{-1}\varepsilon\right], \tag{9.59}$$

$$s_{\sigma_\mu^2} = -\frac{T}{2}\left[\frac{S}{\lambda_2} - \frac{T\sigma_\alpha^2}{\lambda_2}\sum_{i=1}^{N}\frac{M_i}{\lambda_{3i}} - \varepsilon'(A')^{-1}\left[(\Omega_u^{-1})^2 - \lambda_1^{-2}\widetilde{Q}_1\right] A^{-1}\varepsilon\right], \tag{9.60}$$

$$s_{\sigma_v^2} = -\frac{1}{2}\left[\frac{S(T-1)}{\lambda_1} + \frac{(S-N)}{\lambda_2} + \sum_{i=1}^{N}\lambda_{3i}^{-1} - \varepsilon'(A')^{-1}(\Omega_u^{-1})^2 A^{-1}\varepsilon\right], \tag{9.61}$$

evaluated under the null hypothesis. In this case, the restricted ML estimator is the OLS estimator. The terms $\widehat{\eta}_{\lambda\lambda}$, $\widehat{\eta}_{\rho\rho}$, $\widehat{\eta}_{\lambda\rho}$, $\widehat{\eta}_{\lambda\sigma_\alpha^2}$, $\widehat{\eta}_{\lambda\sigma_\mu^2}$, $\widehat{\eta}_{\rho\sigma_\alpha^2}$, $\widehat{\eta}_{\rho\sigma_\mu^2}$, $\widehat{\eta}_{\sigma_\alpha^2\sigma_\alpha^2}$, $\widehat{\eta}_{\sigma_\mu^2\sigma_\alpha^2}$ and $\widehat{\eta}_{\sigma_\mu^2\sigma_\mu^2}$ are those of the inverse information matrix also evaluated under H_0^1 (see He and Lin (2015)). Under the null hypothesis, the LM_1 statistic is asymptotically distributed as a chi-squared distribution with four degrees of freedom, i.e., χ_4^2. If the null hypothesis cannot be rejected, this corresponds to the usual pooled panel data model. Otherwise, more investigation is needed and conditional tests assuming all or part of spatial autocorrelation of the disturbances, spatial lag dependence, group-specific time-invariant effects, and nested individual effects, should be considered. For instance, if we are interested in testing spatial effects, i.e., spatial lag dependence and spatial autocorrelation of the disturbances, $H_0^2 : \rho = \lambda = 0 | \sigma_\alpha^2 > 0, \sigma_\mu^2 > 0$ vs. H_1^2 : At least one of them is not zero, the LM_2 statistic is given by

$$LM_2 = \widehat{\eta}_{\lambda\lambda}\widehat{s}_\lambda^2 + 2\widehat{\eta}_{\rho\lambda}\widehat{s}_\rho\widehat{s}_\lambda + \widehat{\eta}_{\rho\rho}\widehat{s}_\rho^2, \tag{9.62}$$

where $\widehat{\eta}_{\lambda\lambda}$, $\widehat{\eta}_{\rho\rho}$, $\widehat{\eta}_{\lambda\rho}$, $\widehat{s}_\rho^2$, $\widehat{s}_\mu^2$ and $\widehat{s}_{\rho\mu}$ are evaluated under H_0^2. The model is reduced to the nested random effects model suggested by Baltagi et al. (2001). Under the null hypothesis, the LM_2 statistic is asymptotically distributed as a chi-squared distribution with two degrees of freedom, i.e., χ_2^2 (see He and Lin (2015) for a description of all conditional tests).

Overall, using Monte Carlo simulations, He and Lin (2015) show the good finite sample performance of the LM tests that they have developed. To take into account possible distributional misspecification in the finite sample and spatial layout sensitivity, modified versions of these LM tests can be obtained along the lines of Yang (2010) and Baltagi and Yang (2013).

9.5 Prediction with Spatial Models

Predicting with spatial (two-dimensional) panels has recently become an integral part of the empirical work in economics (see Baltagi and Li (2004, 2006); Longhi and Nijkamp (2007); Kholodilin et al. (2008); Fingleton (2009); Schanne et al. (2010); Girardin and Kholodilin (2011); Baltagi et al. (2014a) among others). Nevertheless, prediction still remains in its infancy with regard to spatial multi-dimensional panels (see Baltagi and Pirotte (2014)).

Following Goldberger (1962), assuming $\rho = 0$ and $\theta = 0$, the BLUP for the dependent variable $y_{ij,T+\tau}$, denoted by $\widehat{y}_{ij,T+\tau}$, can be written in a general form as

$$\widehat{y}_{ij,T+\tau} = x_{ij,T+\tau}\widehat{\beta}_{GLS} + \omega'\Omega_\varepsilon^{-1}\widehat{\varepsilon}_{GLS}, \tag{9.63}$$

where $\omega = \mathrm{E}[\varepsilon_{ij,T+\tau}\varepsilon]$ is the covariance between the future disturbance $\varepsilon_{ij,T+\tau}$ and the sample disturbances ε. $\widehat{\beta}_{GLS}$ is the GLS estimator of β based on true Ω_ε, while $\widehat{\varepsilon}_{GLS}$ denotes the corresponding GLS residual vector. For the nested error compo-

nents model *without* spatial autocorrelation ($\lambda = 0$), assuming that the time dimension is unbalanced, Baltagi and Pirotte (2013) derive the BLUP where ω' is reduced to

$$\omega' = \left(0', \left[\left[\sigma_\alpha^2 \iota'_{M_i} + \sigma_\mu^2 l'_j\right] \otimes \iota'_{T_i}\right], 0'\right), \tag{9.64}$$

where l_j is the jth column of I_{M_i} and $0'$ is a row vector of zeros of appropriate length, and

$$\omega' \Omega_\varepsilon^{-1} \widehat{\varepsilon}_{GLS} = \left(\frac{T_i \sigma_\mu^2}{\lambda_{2i}}\right) \widehat{\varepsilon}_{ij.,GLS} + \left[\frac{M_i T_i \sigma_\alpha^2 \sigma_\nu^2}{\lambda_{2i} \lambda_{3i}}\right] \widehat{\varepsilon}_{i..,GLS}, \tag{9.65}$$

with $\widehat{\varepsilon}_{ij.,GLS} = \sum_{t=1}^{T_i} \widehat{\varepsilon}_{ijt,GLS}/T_i$ and $\widehat{\varepsilon}_{i..,GLS} = \sum_{j=1}^{M_i} \sum_{t=1}^{T_i} \widehat{\varepsilon}_{ijt,GLS}/M_i T_i$. Thus, if we transfer (9.65) into (9.63), the BLUP of $y_{ij,T_i+\tau}$ is given by

$$\widehat{y}_{ij,T_i+\tau} = x_{ij,T+\tau} \widehat{\beta}_{GLS} + \left(\frac{T_i \sigma_\mu^2}{\lambda_2}\right) \widehat{\varepsilon}_{ij.,GLS} + \left[\frac{M_i T_i \sigma_\alpha^2 \sigma_\nu^2}{\lambda_2 \lambda_{3i}}\right] \widehat{\varepsilon}_{i..,GLS}. \tag{9.66}$$

Therefore, the BLUP of $y_{ij,T_i+\tau}$ for the nested error components model modifies the usual GLS forecast by adding two terms. The first is a fraction of the average of the GLS residuals (over time) corresponding to the individual j in group i.The second term adds a fraction of the average GLS residual (over time as well as individual j) corresponding to group i. In order to make (9.66) operational, $\widehat{\beta}_{GLS}$ and the variance components are replaced by their feasible estimates proposed by Baltagi et al. (2001).

Now, if *both* spatial autoregressive and nested error components are present in the model (SAR-NRE model), after some algebra and assuming that the time dimension is balanced (see Baltagi and Pirotte (2014)), one can show that

$$\begin{aligned} \omega' = {} & \sigma_\alpha^2 b_{ij} \text{diag}\left(\iota_{M_i}\right) \left(\iota'_T \otimes \text{diag}\left(\iota_{M_i}\right)' \left(B_S^{-1}\right)'\right) \\ & + \sigma_\mu^2 b_{ij} \left(\iota'_T \otimes \left(B_S^{-1}\right)'\right), \end{aligned} \tag{9.67}$$

where b_{ij} is the ijth row of the matrix B_S^{-1}, and the second term of (9.63) is given by

$$\begin{aligned} \omega' \Omega_\varepsilon^{-1} \widehat{\varepsilon}_{GLS} = {} & \frac{\sigma_\alpha^2}{\sigma_\nu^2} b_{ij} \text{diag}\left(\iota_{M_i}\right) \left[\iota'_T \otimes \text{diag}\left(\theta_{2i}^{-1} \iota'_{M_i}\right)\right] \left[I_T \otimes B_S\right] \widehat{\varepsilon}_{GLS} \\ & + \frac{\sigma_\mu^2}{\sigma_\nu^2} b_{ij} \left[\iota'_T \otimes \left[\theta_1^{-1} \text{diag}\left(Q_{M_i}\right) + \text{diag}\left(\theta_{2i}^{-1} \overline{J}_{M_i}\right)\right]\right] \\ & \left[I_T \otimes B_S\right] \widehat{\varepsilon}_{GLS}, \end{aligned} \tag{9.68}$$

where $\theta_1 = (T\delta_2 + 1)$, $\theta_{2i} = (M_i T \delta_1 + T\delta_2 + 1)$, $\delta_1 = \sigma_\alpha^2/\sigma_\nu^2$ and $\delta_2 = \sigma_\mu^2/\sigma_\nu^2$. Thus, if we transfer (9.68) into (9.63), the BLUP of $y_{ij,T+\tau}$ is given by

$$\begin{aligned}\widehat{y}_{ij,T+\tau} = {} & x_{ij,T+\tau}\widehat{\beta}_{GLS} \\ & + \frac{\sigma_\alpha^2}{\sigma_v^2} b_{ij} \mathrm{diag}\left(\iota_{M_i}\right) \left[\iota_T' \otimes \mathrm{diag}\left(\theta_{2i}^{-1}\iota_{M_i}'\right)\right] \left[I_T \otimes B_S\right] \widehat{\varepsilon}_{GLS} \\ & + \frac{\sigma_\mu^2}{\sigma_v^2} b_{ij} \left[\iota_T' \otimes \left[\theta_1^{-1}\mathrm{diag}\left(Q_{M_i}\right) + \mathrm{diag}\left(\theta_{2i}^{-1}\overline{J}_{M_i}\right)\right]\right] \\ & \left[I_T \otimes B_S\right] \widehat{\varepsilon}_{GLS}. \end{aligned} \tag{9.69}$$

To compute (9.69), we use the ML estimates obtained from the iterative procedure or the GMM-S-FGLS approach described in Sect. 9.3.

If $\rho \neq 0$ and $\theta = 0$, it is possible to obtain a BLUP using the reduced form (9.7). The derivation is similar to the one developed by Baltagi et al. (2014a) in the case of the two-dimensional panel data model with spatially correlated error component disturbances. To operationalize this BLUP, an IV estimator has to be used, see Sect. 9.3.2. This approach remains valid when spatial lags on explanatory variables ($\theta \neq 0$) are introduced in the multi-dimensional spatial panel model.

9.6 Some Further Topics

9.6.1 Heterogenous Coefficients Spatial Models

As mentioned is Sect. 9.2, model (9.1) implies that unobservable heterogeneity is only captured through the regression constant or a composite disturbance term. If heterogeneity is more complex (slope heterogeneity) and neglected, the consistency of the estimates and the inference will be affected. Then, it is necessary to implement spatial models that explicitly allow for slope heterogeneity in the temporal or in the spatial dimension. In this section we derive a number of possibilities.

In the case where T is small compared to S, time slope heterogeneity can be allowed using a Seemingly Unrelated Regression (SUR) framework. Following Anselin (1988), we can specify one equation for each time period, which is estimated for a cross section of spatial units, which are organised hierarchically. For instance, we can allow for time-varying coefficients in equations (9.3) and (9.4)

$$y_t = \rho_t W_S y_t + x_t \beta_t + W_S x_t \theta_t + \varepsilon_t, \tag{9.70}$$

$$\varepsilon_t = \lambda_t W_S \varepsilon_t + u_t. \tag{9.71}$$

Additional assumptions allowing for cross-equation correlation in the u_t can be specified. In non-hierarchical panels, the estimation of such a model using ML has been derived by Anselin (1988). Mur et al. (2010) and Lopez et al. (2014) further extended it to deal with a higher number of equations, say Γ, for each time period. The case of a SUR model with spatial error autocorrelation and unobserved indi-

vidual random effects in a two-way error component model has been analyzed by Baltagi and Pirotte (2011) with a GMM-FGLS procedure.

Conversely, to deal with spatial slope heterogeneity, it is necessary to assume that the number of time observations, T, is large, see Pesaran (2015b) and Chudik and Pesaran (2015). Then, it becomes possible to specify a spatially heterogenous version of (9.1)

$$y_{ijt} = \rho_{ij} \sum_{g=1}^{N} \sum_{h=1}^{M_g} w_{ij,gh} y_{ght} + x_{ijt} \beta_{ij} + \sum_{g=1}^{N} \sum_{h=1}^{M_g} w_{ij,gh} x_{ght} \theta_{ij} + \varepsilon_{ijt}. \tag{9.72}$$

This specification assumes that the coefficients are allowed to vary across individuals which could be nested in groups.

For the non-hierarchical case, Aquaro et al. (2015) propose a QML estimation procedure for this category of spatial panel models with heterogenous coefficients, while LeSage and Chih (2016) derive the partial derivatives and the associated impacts. The extension of such procedures to the hierarchical case would then be an important issue for practical purposes.

Finally, an important question is the nature and the degree of dependencies between individuals (or spatial units). Are the observed dependencies between different individuals due to common factors that affect different units, rather than the result of local interactions that generate spatial spillover effects? The factor and spatial econometric approaches tend to complement each other, with the factor approach more suited to modelling strong cross-sectional dependence (e.g., aggregate shocks), while the spatial approach (connected to a spatial weighted matrix) generally requires the spatial dependence to be weak. This is an important point because most large panel datasets are subject to a combination of strong and weak cross-dependencies.

Bailey et al. (2016) adopt a two-stage estimation strategy. They apply this framework to real house price changes of 363 U.S. Metropolitan Statistical Areas (MSAs), excluding three MSAs located in Alaska and Hawaii, over the period 1974:Q1 to 2010:Q4 ($T = 144$ quarters). These data exhibit are typically grouped by MSA within states (49), thus forming a nested structure. Nevertheless, instead of focusing on states as the upper level of the nested hierarchy, they use the Bureau of Economic Analysis regional classification, which comprises 8 regions (New England, Mid East, South East, Great Lakes, Plains, South West, Rocky Mountains and Far West, $N = 8$), which are homogenous with regard to economic and social factors across states. Each region contains an average of around 45 MSAs. In a first step, they consider the hierarchical model

$$y_{ijt} = \beta_{0ij} + \beta_{1ij} y_{.it} + \beta_{2ij} y_{..t} + \varepsilon_{ijt}, \tag{9.73}$$

where $y_{.it} = M_i^{-1} \sum_{j=1}^{M_i} y_{ijt}$ and $y_{..t} = S^{-1} \sum_{i=1}^{N} \sum_{j=1}^{M_i} y_{ijt}$. $y_{.it}$ corresponds to the region mean at time t, $y_{..t}$, the national mean at time t. They examine the degree of cross-sectional dependence, using the Cross-sectional Dependence (CD) statistic, see Pesaran (2015a), to test if it is appropriate to apply standard spatial modelling

methods directly (if the null of weak cross-dependence is not rejected). If the null of weak dependence is rejected, spatial dependence is considered to be strong and factor models become appropriate. They show that the cross-sectional averages approach, as well as principal components, perform reasonably well in purging factor loadings from price changes (checked through a CD test). Then, pair-wise correlations of de-factored price changes are used to built positive and negative spatial matrices. In a second step, a time-space spatial model without any exogenous regressors is estimated using a QML estimator.

9.6.2 Time-Space Models

Panel data with spatial interactions are of great interest, not only to control for the observable/unobservable heterogeneities, but also to take into account the dynamics. A dynamic version of model (9.5) can be obtained by adding one or more of the following variables: a dependent variable lagged in time (y_{-1}), a dependent variable lagged in both space and time (Wy_{-1}), explanatory variables lagged in time (X_{-1}) and explanatory variables lagged in both space and time (WX_{-1}). In matrix form, we have

$$y = \phi y_{-1} + \rho Wy + \pi Wy_{-1} + X\beta + X_{-1}\beta^* + WX\theta + WX_{-1}\theta^* + \varepsilon. \qquad (9.74)$$

The disturbance terms ε could also be contemporaneously correlated (9.6) including a nested error components structure (9.19). An alternative to (9.19) is to introduce fixed effects to capture unobservable heterogeneity. In the two-dimensional spatial panel data literature, many articles focus on the estimation procedures: QML (Lee et al., 2008; Qu and Lee, 2015), ML (Lee and Yu, 2010c,b, 2016), Generalized Method of Moments (Lee and Yu, 2014; Baltagi et al., 2014a), bias-corrected estimators (Lee et al., 2008; Elhorst, 2010a; Korniotis, 2010) and Bayesian procedures (Parent and LeSage, 2010, 2012; Debarsy et al., 2012), assuming that all or some of the above-mentioned variables are present. Moreover, considering the time-space dynamic spatial model (9.74), several spatial multiplier matrices are at work at the same time, which complicate the parameter interpretations (direct, indirect and total short-run and long-run effects). More empirical research is needed to clarify interpretation. Despite this significant literature on two-dimensional dynamic spatial panel models, multi-dimensional hierarchical spatial panel models remain unexplored.

9.7 Conclusion

This chapter has focused on hierarchical spatial panels, i.e., panel data models where data present a nested structure together with spatial autocorrelation at the individ-

ual level. We started with a general specification with balanced data in the time dimension, slope coefficient homogeneity and spatial autocorrelation in the form of a spatial lag of the endogenous variable, spatial lags of the exogenous variables and spatial error autocorrelation (either in autoregressive or in moving average form). Then we showed how these models can be estimated with ML of GMM methods and some LM specification tests. Finally, some possible extensions dealing with parameter heterogeneity in one of the dimensions were presented.

Obviously, this literature is still in its infancy and much remains to be done. In particular, these models need to be extended to deal with the usual problems in panel data, such as allowing for both time and spatial observed and unobserved heterogeneity, unbalanced or incomplete panels in several dimensions or regressors that are invariant over one of the dimensions. Endogeneity other than the spatial lag of the explained variable remains to be explored in these settings (Chap. 3 examines this in non-spatial multi-dimensional panels). This aspect presents more of a challenge to single equation approaches involving ML estimation compared to other methods involving instrumental variables, as ML estimation is not possible when endogeneity is in implicit form, unlike the endogenous spatial lag variable. Finally, large-sample theory available for spatial panels also has to be extended to hierarchical spatial panels.

Acknowledgements

We would particularly like to thank Raja Chakir, Nicolas Debarsy, J. Paul Elhorst, Bernard Fingleton, Alban Thomas and the editor of this handbook, László Mátyás, for their useful comments and suggestions.

References

Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Springer-Science, Dordrecht.

Anselin, L. (2003). Spatial externalities. *International Regional Science Review*, 26:147–152.

Anselin, L. (2010). Thirty years of spatial econometrics. *Papers in Regional Science*, 89:3–25.

Anselin, L., Le Gallo, J., and Jayet, H. (2008). Spatial panel econometrics. In Mátyás, L. and Sevestre, P., editors, *The Econometrics of Panel Data, 3rd edition*, pages 625–660. Springer Verlag.

Antweiler, W. (2001). Nested random effects estimation in unbalanced panel data. *Journal of Econometrics*, 101:295–313.

Aquaro, M., Bailey, N., and Pesaran, H. (2015). *Quasi Maximum Likelihood Estimation of Spatial Models with Heterogeneous Coefficients*. mimeo.

Bailey, N., Holly, S., and Pesaran, M. H. (2016). A two-stage approach to spatio-temporal analysis with strong and weak cross-sectional dependence. *Journal of Applied Econometrics*, 31:249–280.

Baltagi, B. H. (2013). *Econometric Analysis of Panel Data*. John Wiley & Sons., 5 edition.

Baltagi, B. H., Egger, P., and Pfaffermayr, M. (2013). A generalized spatial panel data model with random effects. *Econometric Reviews*, 32:650–685.

Baltagi, B. H., Fingleton, B., and Pirotte, A. (2014a). Estimating and forecasting with a dynamic spatial panel data model. *Oxford Bulletin of Economics and Statistics*, 76:112–136.

Baltagi, B. H., Fingleton, B., and Pirotte, A. (2014b). Spatial lag models with nested random effects: An instrumental variable procedure with an application to english house prices. *Journal of Urban Economics*, 80:76–86.

Baltagi, B. H. and Li, D. (2004). Prediction in the panel data model with spatial correlation. In L. Anselin, R. F. and Rey, S., editors, *Advances in Spatial Econometrics: Methodology, Tools and Applications*, pages 283–295. Springer, Berlin.

Baltagi, B. H. and Li, D. (2006). Prediction in the panel data model with spatial correlation: the case of liquor. *Spatial Economic Analysis*, 1:175–185.

Baltagi, B. H. and Liu, L. (2011). Instrumental variable estimation of a spatial autoregressive panel model with random effects. *Economics Letters*, 111:135–137.

Baltagi, B. H. and Liu, L. (2016). Random effects, fixed effects and Hausman's test for the generalized mixed regressive spatial autoregressive panel data model. *Econometric Reviews*, 35:638–658.

Baltagi, B. H. and Pirotte, A. (2011). Seemingly unrelated regressions with spatial error components. *Empirical Economics*, 40:5–49.

Baltagi, B. H. and Pirotte, A. (2013). Prediction in an unbalanced nested error components panel data model. *Journal of Forecasting*, 32:755–768.

Baltagi, B. H. and Pirotte, A. (2014). Prediction in a spatial nested error components panel data model. *International Journal of Forecasting*, 30:407–414.

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). Simple LM tests for the unbalanced nested error component regression model. *Econometric Reviews*, 21:167–187.

Baltagi, B. H. and Yang, Z. L. (2013). Standardized LM tests for spatial error dependence in linear panel regressions. *Econometrics Journal*, 16:103–134.

Chudik, A. and Pesaran, H. M. (2015). Large panel data models with cross-sectional dependence: A survey. In Baltagi, B. H., editor, *The Oxford Handbook of Panel Data*, pages 3–45. Oxford University Press.

Corrado, L. and Fingleton, B. (2012). Where is the economics in spatial econometrics? *Journal of Regional Science*, 52:210–239.

Das, D., Kelejian, H. H., and Prucha., I. R. (2003). Small sample properties of estimators of spatial autoregressive models with autoregressive disturbances. *Papers in Regional Science*, 82:1–26.

Davis, P. (2002). Estimating multi-way error components models with unbalanced data structures. *Journal of Econometrics*, 106:67–95.

Debarsy, N., Ertur, C., and LeSage, J. P. (2012). Interpreting dynamic space-time panel data models. *Regional Science and Urban Economics*, 9:158–171.

Elhorst, J. (2014). *Spatial Econometrics: From Cross-Sectional Data to Spatial Panels*. Springer Verlag.

Elhorst, J. P. (2010a). Dynamic panels with endogenous interaction effects when T is small. *Regional Science and Urban Economics*, 40:272–282.

Elhorst, J. P. (2010b). Spatial panel data models. In Fischer, M. M. and Getis, A., editors, *Handbook of Applied Spatial Analysis*, pages 377–405. Springer Verlag.

Fingleton, B. (2008). A generalized method of moments estimator for a spatial panel model with an endogenous spatial lag and spatial moving average errors. *Spatial Economic Analysis*, 3:27–44.

Fingleton, B. (2009). Prediction using panel data regression with spatial random effects. *International Regional Science Review*, 32:195–220.

Fingleton, B., Le Gallo, J., and Pirotte, A. (2017). *Panel Data Models with Spatially Nested Random Effects*. Journal of Regional Science, forthcoming.

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.

Girardin, E. and Kholodilin, K. A. (2011). How helpful are spatial effects in forecasting the growth of Chinese provinces? *Journal of Forecasting*, 30:622–643.

Goldberger, A. S. (1962). Best linear unbiased prediction in the generalized linear regression model. *Journal of the American Statistical Association*, 57:369–375.

Goldstein, H. (1995). *Multilevel Statistical Models*. John Wiley & Sons.

He, M. and Lin, K.-P. (2015). Testing spatial effects and random effects in a nested panel data model. *Economics Letters*, 135:85–91.

Kapoor, M., Kelejian, H. H., and Prucha, I. R. (2007). Panel data models with spatially correlated error components. *Journal of Econometrics*, 140:97–130.

Kelejian, H. H. and Prucha, I. R. (1998). A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *Journal of Real Estate Finance and Economics*, 17:99–121.

Kelejian, H. H. and Prucha, I. R. (1999). A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review*, 40:509–533.

Kholodilin, K. A., Siliverstovs, B., and Kooths, S. (2008). A dynamic panel data approach to the forecasting of the GDP of German lander. *Spatial Economic Analysis*, 3:195–207.

Korniotis, G. M. (2010). Estimating panel models with internal and external habit formation. *Journal of Business Economic and Statistics*, 28:145–158.

Lee, L. F. (2003). Best spatial two-stage least square estimators for a spatial autoregressive model with autoregressive disturbances. *Econometric Reviews*, 22:307–335.

Lee, L. F., de Jong, R., and Yu, J. (2008). Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both N and T are large. *Journal of Econometrics*, 146:118–134.

Lee, L. F. and Yu, J. (2010a). Estimation of spatial panels. *Foundations and Trends in Econometrics*, 4:1–164.

Lee, L. F. and Yu, J. (2010b). Some recent developments in spatial panel data models. *Regional Science and Urban Economics*, 40:255–271.

Lee, L. F. and Yu, J. (2010c). A spatial dynamic panel data model with both time and individual fixed effects. *Econometric Theory*, 26:564–597.

Lee, L. F. and Yu, J. (2014). Efficient GMM estimation of spatial dynamic panel data models with fixed effects. *Journal of Econometrics*, 180:174–197.

Lee, L. F. and Yu, J. (2015). Spatial panel data models. In Baltagi, B., editor, *The Oxford Handbook of Panel Data*, pages 363–401. Oxford.

Lee, L. F. and Yu, J. (2016). Identification of spatial Durbin panel models. *Journal of Applied Econometrics*, 31:133–162.

LeSage, J. P. and Chih, Y.-Y. (2016). Interpreting heterogeneous coefficient spatial autoregressive panel models. *Economics Letters*, 142:1–5.

LeSage, J. P. and Pace, R. K. (2009). *Introduction to Spatial Econometrics*. CRC Press.

Longhi, S. and Nijkamp, P. (2007). Forecasting regional labor market developments under spatial heterogeneity and spatial correlation. *International Regional Science Review*, 30:109–119.

Lopez, F., Mur, J., and Angulo, A. (2014). Spatial model selection strategies in a SUR framework. the case of regional productivity in the EU. *The Annals of Regional Science*, 53:197–220.

Mátyás, L., Hornok, C., and Pus, D. (2012). The formulation and estimation of random effects panel data models of trade. In *CEU-Economics Working Paper* $n^o 2$.

Montmarquette, C. and Mahseredjian, S. (1989). Does school matter for educational achievement? A two-way nested-error components analysis. *Journal of Applied Econometrics*, 4:181–193.

Mur, J., Lopez, F., and Herrera, M. (2010). Testing for spatial effects in seemingly unrelated regressions. *Spatial Economic Analysis*, 5:399–440.

Pace, R. K., LeSage, J. P., and Zhu, S. (2012). Spatial dependence in regressors. In Terrell, D. and Millimet, D., editors, *Advances in Econometrics Volume 30, Thomas B. Fomby, R. Carter Hill, Ivan Jeliazkov, Juan Carlos Escanciano and Eric Hillebrand*, pages 257–295. Emerald Group Publishing Limited.

Parent, O. and LeSage, J. P. (2010). A spatial dynamic panel model with random effects applied to commuting times. *Transportation Research, Part B*, 44:633–645.

Parent, O. and LeSage, J. P. (2012). Spatial dynamic panel data models with random effects. *Regional Science and Urban Economics*, 42:727–738.

Pesaran, M. H. (2015a). Testing weak cross-sectional dependence in large panels. *Econometric Reviews*, 34:1089–1117.

Pesaran, M. H. (2015b). *Time Series and Panel Data Econometrics*. Oxford University Press.

Pus, D., Mátyás, L., and Hornok, C. (2013). Modelling firm-product level trade: A multi-dimensional random effects panel data approach. In *CEU-Economics Working Paper n^{o}2*.

Qu, X. and Lee, L. F. (2015). Estimating a spatial autoregresssive model with an endogenous spatial weight matrix. *Journal of Econometrics*, 184:209–232.

Schanne, N., Wapler, R., and Weyh, A. (2010). Regional unemployment forecasts with spatial interdependencies. *International Journal of Forecasting*, 26:908–926.

Upton, G. and Fingleton, B. (1985). *Spatial Data Analysis by Example, Volume 1: Point Pattern and Quantitative Data*. John Wiley & Sons.

Yang, Z. L. (2010). A robust LM test for spatial error components. *Regional Science of Urban Economics*, 40:299–310.

Chapter 10
Modelling in the Presence of Cross-sectional Error Dependence

George Kapetanios, Camilla Mastromarco, Laura Serlenga, and Yongcheol Shin

Abstract Given the growing availability of big datasets which contain information on multiple dimensions and following the recent research trend on multi-dimensional modelling, we develop three-dimensional panel data models with three-way error components that allow for strong cross-sectional dependence (CSD) through unobserved heterogeneous global factors, and propose appropriate consistent estimation procedures. We also discuss the extent of CSD in 3D models and provide a diagnostic test for cross-sectional dependence. We provide the extensions to unbalanced panels and 4D models. The validity of the proposed approach is confirmed by the Monte Carlo simulation results. We also demonstrate the empirical usefulness through the application to the 3D panel gravity model of the intra-EU trade flows.

10.1 Introduction

Given the growing availability of big datasets containing information on multiple dimensions, the recent literature on panel data has focused more on extending the two-way error components models to a multi-dimensional setting. Chapter 1 (BMW hereafter, from the initials of the authors) introduces the appropriate Within estimators for the most frequently used three-dimensional (3D) fixed effects panel data

George Kapetanios
King's College London, London, United Kingdom

Camilla Mastromarco
University of Salento, Lecce, Italy

Laura Serlenga
University of Bari, Bari, Italy

Yongcheol Shin
University of York, York, United Kingdom

L. Matyas (ed.), *The Econometrics of Multi-dimensional Panels*, Advanced Studies in Theoretical and Applied Econometrics 50, DOI 10.1007/978-3-319-60783-2_10

models, and Chap. 2 (BBMP hereafter, from the initials of the authors) considers the random effects approach and proposes a sequence of optimal GLS estimators. This multi-dimensional approach is expected to become an essential tool for the analysis of the complex interconnectedness of big datasets, and can be applied not only to bilateral (origin-destination) flows such as trade, FDI, capital or migration (see, e.g., Feenstra, 2004; Bertoli and Moraga, 2013; Gunnella et al., 2015), but also to a variety of matched datasets which may link, for example, employers-employees, pupils-teachers, (see, e.g., Abowd et al., 1999; Kramarz et al., 2008).

However, no study has attempted to address the important issue of explicitly controlling cross-sectional error dependence in 3D or higher-dimensional panel data, even though the cross-sectional dependence (CSD) seems pervasive even in 2D panels. This is because it seems rare that the cross-sectional covariance of the errors is zero (see, e.g., Pesaran, 2015). Recently, there has been much progress in modelling CSD in 2D panels by two main approaches, the factor-based approach (see, e.g., Pesaran, 2006; Bai, 2009) and the spatial econometrics techniques (see, e.g., Behrens et al., 2012; Mastromarco et al., 2016b). Chudik et al. (2011) show that factor-based models exhibit the strong CSD, whilst the spatial-based models can deal with weak CSD only (see also Bailey et al. (2016b) for a more general discussion).

Chapter 9 reviews the current state-of-art in the analysis of multi-dimensional nested spatial panels, highlighting a range of issues related to the specification, estimation, testing procedures and predictions. Chapter 11 provides a survey of empirical issues in the analysis of the gravity-model estimation of international trade flows, then proceeds with the modelling of the multi-dimensional stochastic structure, focusing on the fixed-effects estimation, and describes how spatial autocorrelation and spillovers can be introduced into such models. Chapter 12 surveys hedonic housing models and discrete choice models using multi-dimensional panels, also focussing on the spatial econometrics approach.

Following this research trend, we develop 3D panel data models with strong CSD. In particular, we generalise the multi-dimensional error components specification by modelling residual CSD via unobserved heterogeneous global factors. The multi-dimensional country-time fixed (CTFE) and random effects (CTRE) estimators proposed by BMW and BBMP fail to remove heterogenous global factors, suggesting that they are biased in the presence of nonzero correlation between the regressors and the unobserved global factors. In this regard, we develop a two-step consistent estimation procedure. First, we follow Pesaran (2006) and augment the 3D model with the cross-sectional averages of the dependent variable and regressors over double cross-sectional units, which are shown to provide valid proxies for unobserved heterogenous global factors. Next, we apply the 3D-Within transformation to the augmented specification and obtain consistent estimators, called the 3D-PCCE estimator. Our approach is the first attempt to accommodate strong CSD in multi-dimensional panels, and is expected to a make timely contribution to the growing literature.

We discuss the extent of CSD within the 3D panel data models under three different error components specifications: the CTFE, the two-way heterogeneous factor, and both components. We also distinguish between three types of CSD under

the hierarchical multi-factor error components specification recently advanced by Kapetanios and Shin (2017). First, the global factor tends to display strong CSD as it influences the (ij) pairwise interactions for $i = 1,...,N_1$ and $j = 1,...,N_2$ (of N_1N_2 dimension). Next, the local factors show semi-strong or semi-weak CSD, as they influence origin and destination countries separately (each of N_1 or N_2 dimension). Finally, idiosyncratic errors are characterised with weak or no CSD.

We then develop a diagnostic test for the null hypothesis of (pairwise) residual cross-sectional independence or weak dependence in the 3D panels, which is a modified counterpart of an existing CD test in the 2D panels proposed by Pesaran (2015) and we describe how to consistently estimate the exponent of cross-sectional dependence by extending Bailey et al. (2016b). Furthermore, we provide a couple of extensions into unbalanced panels and 4D or higher dimensional models.

We have conducted a Monte Carlo studies to investigate the small sample properties of the 3D-PCCE estimators relative to the CTFE estimator. We find strong evidence that the 3D-PCCE estimators perform well when the 3D panel data is subject to the strong CSD through heterogeneous global factors. In contrast, the CTFE estimator tends to display severe biases and size distortions.

We apply our proposed 3D PCCE estimation techniques, together with the two-way fixed effects and the CTFE estimators, to a dataset over the period 1960–2008 (49 years) for 182 country-pairs amongst 14 EU countries. Based on the CD test results, estimates of CSD exponent, and the predicted signs and statistical significance of the coefficients, we come to the conclusion that the 3D PCCE estimation results are mostly satisfactory and reliable. In particular, when we explicitly control for strong CSD in the 3D panels, we find that the trade effect of currency union is rather modest. It seems that this evidence provides strong support for the thesis that the trade increase within the Euro area may reflect a continuation of a long-run historical trend linked to the broader set of the EU's economic integration policies.

This chapter proceeds in seven sections. Section 10.2 introduces 3D models with three-way error components that allow for strong cross-sectional dependence, and develops a consistent estimation procedure. Section 10.3 discusses the nature of CSD in 3D models and provide a diagnostic test for cross-sectional dependence. Section 10.4 presents the extension to unbalanced panels and 4D models. Section 10.5 discusses the Monte Carlo simulation results. The empirical results for the gravity model of EU export flows are presented in Sect. 10.6, and Sect. 10.7 concludes.

Throughout the chapter, we adopt the following standard notations. $\mathbf{I}_N$ is an $(N \times N)$ identity matrix, $\mathbf{J}_N$ the $(N \times N)$ identity matrix of ones, and ι_N the $(N \times 1)$ vector of ones, respectively. $\mathbf{M}_A$ projects the $(N \times N)$ matrix $\mathbf{A}$ into its null-space, i.e., $\mathbf{M}_A = \mathbf{I}_N - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$. Finally, $\bar{y}_{.jt} = N_1^{-1}\sum_{i=1}^{N_1} y_{ijt}$, $\bar{y}_{i.t} = N_2^{-1}\sum_{j=1}^{N_2} y_{ijt}$ and $\bar{y}_{ij.} = T^{-1}\sum_{t=1}^{T} y_{ijt}$ denote the average of y over the index i, j and t, respectively, with the definition extending to other quantities, such as $\bar{y}_{..t}$, $\bar{y}_{.j.}$, $\bar{y}_{i..}$ and $\bar{y}_{...}$. This notational convention extends naturally to the higher dimensional cases and unbalanced panels. For example, in the 4D extensions covered in Section 10.4.2, $\bar{y}_{...t} = N_1^{-1}\sum_{i=1}^{N_1} N_2^{-1}\sum_{j=1}^{N_2} N_3^{-1}\sum_{s=1}^{N_3} y_{ijst}$, denotes the average of y over the indices

i, j, s, with the definition extending to other quantities such as $\bar{y}_{.jst}$, $\bar{y}_{i.st.}$, $\bar{y}_{ij.t}$, $\bar{y}_{..st}$, $\bar{y}_{.j.t.}$, $\bar{y}_{.js.}$, $\bar{y}_{i..t}$, $\bar{y}_{i.s..}$, $\bar{y}_{ij..}$, $\bar{y}_{...t}$, $\bar{y}_{..s.}$, $\bar{y}_{.j..}$, $\bar{y}_{i...}$ and $\bar{y}_{...t}$.

10.2 3D Models with Cross-sectional Error Dependence

Following Chaps. 1 and 2, we consider the following three-dimensional country-time fixed effects panel data model

$$y_{ijt} = \beta' \mathbf{x}_{ijt} + \gamma' \mathbf{s}_{it} + \delta' \mathbf{d}_{jt} + \kappa' \mathbf{q}_t + \varphi' \mathbf{z}_{ij} + u_{ijt}, i = 1,...,N_1, j = 1,...,N_2, t = 1,...,T, \tag{10.1}$$

with the error components

$$u_{ijt} = \mu_{ij} + \upsilon_{it} + \zeta_{jt} + \varepsilon_{ijt}\ , \tag{10.2}$$

where y_{ijt} is the dependent variable observed across three indices (e.g., the import of country j from country i at period t), $\mathbf{x}_{ijt}$, $\mathbf{s}_{it}$, $\mathbf{d}_{jt}$, $\mathbf{q}_t$, $\mathbf{z}_{ij}$ are the $(k_x \times 1)$, $(k_s \times 1)$, $(k_d \times 1)$, $(k_q \times 1)$, $(k_z \times 1)$ vectors of covariates covering all possible measurements observed across three indices, and β, γ, δ, κ, φ, are the associated vectors of the parameters. The multiple error components in (10.2) contain bilateral pair-fixed effects (μ_{ij}), as well as origin and destination country-time fixed effects (CTFE), υ_{it} and ζ_{jt}, respectively.[1]

To remove all unobserved fixed effects, μ_{ij}, υ_{it} and ζ_{jt}, BMW derive the following 3D Within transformation[2]

$$\tilde{y}_{ijt} = y_{ijt} - \bar{y}_{ij.} - \bar{y}_{.jt} - \bar{y}_{i.t} + \bar{y}_{..t} + \bar{y}_{.j.} + \bar{y}_{i..} - \bar{y}_{...}\ . \tag{10.3}$$

Applying the 3D Within transformation to (10.1), we can estimate consistently β only from the following regression

$$\tilde{y}_{ijt} = \beta' \tilde{\mathbf{x}}_{ijt} + \tilde{\varepsilon}_{ijt},\ \ i = 1,...,N_1,\ j = 1,...,N_2,\ t = 1,...,T, \tag{10.4}$$

where $\tilde{\mathbf{x}}_{ijt} = \mathbf{x}_{ijt} - \bar{\mathbf{x}}_{ij.} - \bar{\mathbf{x}}_{.jt} - \bar{\mathbf{x}}_{i.t} + \bar{\mathbf{x}}_{..t} + \bar{\mathbf{x}}_{.j.} + \bar{\mathbf{x}}_{i..} - \bar{\mathbf{x}}_{...}$ and similarly for $\tilde{\varepsilon}_{ijt}$. We write (10.4) compactly as

$$\tilde{\mathbf{Y}}_{ij} = \tilde{\mathbf{X}}_{ij}\beta + \tilde{\mathbf{E}}_{ij}\ , \tag{10.5}$$

where

$$\underset{(T\times 1)}{\tilde{\mathbf{Y}}_{ij}} = \begin{bmatrix} \tilde{y}_{ij1} \\ \vdots \\ \tilde{y}_{ijT} \end{bmatrix},\ \underset{(T\times k_x)}{\tilde{\mathbf{X}}_{ij}} = \begin{bmatrix} \tilde{\mathbf{x}}'_{ij1} \\ \vdots \\ \tilde{\mathbf{x}}'_{ijT} \end{bmatrix},\ \underset{(T\times 1)}{\tilde{\mathbf{E}}_{ij}} = \begin{bmatrix} \tilde{\varepsilon}_{ij1} \\ \vdots \\ \tilde{\varepsilon}_{ijT} \end{bmatrix}.$$

The 3D-Within estimator of β is obtained by

[1] Note that the error component specification (10.2) is proposed by Baltagi et al. (2003).

[2] Baltagi et al. (2015) also derive the same projection by applying Davis's (2002) Lemma twice (see Corollary 1).

$$\hat{\beta}_W = \left(\sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \tilde{\mathbf{X}}_{ij}' \tilde{\mathbf{X}}_{ij} \right)^{-1} \left(\sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \tilde{\mathbf{X}}_{ij}' \tilde{\mathbf{Y}}_{ij} \right). \tag{10.6}$$

Then, it follows that, as $(N_1, N_2, T) \to \infty$ (see also BBMP),

$$\sqrt{N_1 N_2 T} \left(\hat{\beta}_W - \beta \right) \overset{a}{\sim} N \left(\mathbf{0}, \sigma_\varepsilon^2 \lim_{(N_1,N_2,T)\to\infty} \left(\frac{1}{N_1 N_2 T} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \tilde{\mathbf{X}}_{ij}' \tilde{\mathbf{X}}_{ij} \right)^{-1} \right).$$

By construction, the Within transformation in (10.3) wipes out all other covariates, $\mathbf{x}_{it}$, $\mathbf{x}_{jt}$, $\mathbf{x}_t$, and $\mathbf{x}_{ij}$ in (10.1). However, we may be interested in uncovering the effects of those covariates (e.g., the impacts of measured trade costs in the structural gravity model). In order to recover them, it would be worthwhile developing an extension of the Hausman and Taylor (1981) estimation, which has been popular in the two-way panel data models even in the presence of cross-sectionally correlated errors (see, e.g., Serlenga and Shin, 2007). Chapter 3 in fact develops an extended Hausman-Taylor estimator for multi-dimensional panel data models.

BMW also show that the CTFE error components in (10.2) nests a number of special cases by applying suitable restrictions to (10.2).[3] Note, however, that model (10.1) with (10.2) does not address the important issue of cross-sectional error dependence. In the presence of such cross-sectional dependence (CSD), the 3D-Within estimator would likely be biased. In this regard, we consider a couple of alternative three-way error components specifications that can accommodate CSD, and develop the appropriate estimation techniques.

Given that υ_{it} and ζ_{jt} are supposed to measure the (local) origin and destination country-time fixed effects, it is natural to add the global factor λ_t to (10.2)

$$u_{ijt} = \mu_{ij} + \upsilon_{it} + \zeta_{jt} + \lambda_t + \varepsilon_{ijt} \ .$$

However, the 3D-Within transformation (10.3) removes λ_t together with μ_{ij}, υ_{it} and ζ_{jt}, because λ_t is shown to be proportional to $\sum_{i=1}^{N_1} \upsilon_{it}$ or $\sum_{j=1}^{N_2} \zeta_{jt}$.

To introduce strong CSD explicitly in the 3D model, (10.1), we first consider the following error components specification

$$u_{ijt} = \mu_{ij} + \pi_{ij} \lambda_t + \varepsilon_{ijt} \ . \tag{10.7}$$

This is similar to the two-way heterogeneous factor model considered by Serlenga and Shin (2007). We follow Pesaran (2006) and apply the cross-sectional averages of (10.1) and (10.7) over i and j to obtain

[3] Baltagi et al. (2003), Baldwin and Taglioni (2006), and Baier and Bergstrand (2007) consider several forms of fixed effects, such as $u_{ijt} = \alpha_i + \gamma_j + \lambda_t + \varepsilon_{ijt}$, $u_{ijt} = \mu_{ij} + \lambda_t + \varepsilon_{ijt}$, $u_{ijt} = \zeta_{jt} + \varepsilon_{ijt}$, $u_{ijt} = \upsilon_{it} + \varepsilon_{ijt}$, and $u_{ijt} = \upsilon_{it} + \zeta_{jt} + \varepsilon_{ijt}$.

$$\bar{y}_{..t} = \frac{1}{N_1}\sum_{i=1}^{N_1}\frac{1}{N_2}\sum_{j=1}^{N_2}\left(\beta'\mathbf{x}_{ijt}+\gamma'\mathbf{s}_{it}+\delta'\mathbf{d}_{jt}+\kappa'\mathbf{q}_t+\varphi'\mathbf{z}_{ij}+\mu_{ij}+\pi_{ij}\lambda_t+\varepsilon_{ijt}\right)$$
$$= \beta'\bar{\mathbf{x}}_{..t}+\gamma'\bar{\mathbf{s}}_{.t}+\delta'\bar{\mathbf{d}}_{.t}+\kappa'\mathbf{q}_t+\varphi'\bar{\mathbf{z}}_{..}+\bar{\mu}_{..}+\bar{\pi}_{..}\lambda_t+\bar{\varepsilon}_{..t}\ , \tag{10.8}$$

where $\bar{\mathbf{s}}_{.t}=N_1^{-1}\sum_{i=1}^{N_1}\mathbf{s}_{it}$, $\bar{\mathbf{d}}_{.t}=N_2^{-1}\sum_{j=1}^{N_2}\mathbf{d}_{jt}$, $\bar{\mathbf{z}}_{..}=(N_1N_2)^{-1}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\mathbf{z}_{ij}$, $\bar{\mu}_{..}=(N_1N_2)^{-1}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\mu_{ij}$ and $\bar{\pi}_{..}=(N_1N_2)^{-1}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\pi_{ij}$. Hence,

$$\lambda_t=\frac{1}{\bar{\pi}_{..}}\left\{\bar{y}_{..t}-\left(\beta'\bar{\mathbf{x}}_{..t}+\gamma'\bar{\mathbf{s}}_{.t}+\delta'\bar{\mathbf{d}}_{.t}+\kappa'\mathbf{q}_t+\varphi'\bar{\mathbf{z}}_{..}+\bar{\mu}_{..}+\bar{\varepsilon}_{..t}\right)\right\}\ .$$

Using these results, we can augment the model (10.1) with the cross-sectional averages as follows

$$y_{ijt}=\beta'\mathbf{x}_{ijt}+\gamma'\mathbf{s}_{it}+\delta'\mathbf{d}_{jt}+\psi_{ij}'\mathbf{f}_t+\tau_{ij}+\mu_{ij}^*+\varepsilon_{ijt}^*\ , \tag{10.9}$$

where

$$\psi_{ij}'=\left(\psi_{0ij},\psi_{1ij}',\psi_{2ij}',\psi_{3ij}',\psi_{4ij}'\right)=\left(\frac{\pi_{ij}}{\bar{\pi}_{..}},\frac{-\pi_{ij}\beta'}{\bar{\pi}_{..}},\frac{-\pi_{ij}\gamma'}{\bar{\pi}_{..}},\frac{-\pi_{ij}\delta'}{\bar{\pi}_{..}},\left(1-\frac{\pi_{ij}}{\bar{\pi}_{..}}\right)\kappa'\right)$$

$$\mathbf{f}_t=\left(\bar{y}_{..t},\bar{\mathbf{x}}_{..t}',\bar{\mathbf{s}}_{.t}',\bar{\mathbf{d}}_{.t}',\mathbf{q}_t'\right)' \tag{10.10}$$

$$\tau_{ij}=\varphi'\mathbf{z}_{ij}-\frac{-\pi_{ij}}{\bar{\pi}_{..}}\varphi'\bar{\mathbf{z}}_{..},\ \mu_{ij}^*=\mu_{ij}-\frac{\pi_{ij}\bar{\mu}_{..}}{\bar{\pi}_{..}},\ \varepsilon_{ijt}^*=\varepsilon_{ijt}-\frac{\pi_{ij}}{\bar{\pi}_{..}}\bar{\varepsilon}_{..t}.$$

We write (10.9) compactly as

$$\mathbf{Y}_{ij}=\mathbf{X}_{ij}\beta+\mathbf{S}_i\gamma+\mathbf{D}_j\delta+\mathbf{F}\psi_{ij}+\tau_{ij}\iota_T+\mu_{ij}^*\iota_T+\mathbf{E}_{ij}^* \tag{10.11}$$
$$=\mathbf{W}_{ij}\theta+\mathbf{H}\psi_{ij}^*+\mathbf{E}_{ij}^*,\ i=1,...,N_1,j=1,...,N_2\ ,$$

where

$$\underset{(T\times 1)}{\mathbf{Y}_{ij}}=\begin{bmatrix}y_{ij1}\\ \vdots\\ y_{ijT}\end{bmatrix},\ \underset{(T\times k_x)}{\mathbf{X}_{ij}}=\begin{bmatrix}\mathbf{x}_{ij1}'\\ \vdots\\ \mathbf{x}_{ijT}'\end{bmatrix},\ \underset{(T\times k_s)}{\mathbf{S}_i}=\begin{bmatrix}\mathbf{s}_{i1}'\\ \vdots\\ \mathbf{s}_{iT}'\end{bmatrix},$$
$$\underset{(T\times k_d)}{\mathbf{D}_j}=\begin{bmatrix}\mathbf{d}_{j1}'\\ \vdots\\ \mathbf{d}_{jT}'\end{bmatrix},\ \underset{(T\times k_f)}{\mathbf{F}}=\begin{bmatrix}\mathbf{f}_1'\\ \vdots\\ \mathbf{f}_T'\end{bmatrix},\ \underset{(T\times 1)}{\mathbf{E}_{ij}^*}=\begin{bmatrix}\varepsilon_{ij1}^*\\ \vdots\\ \varepsilon_{ijT}^*\end{bmatrix},$$

$\mathbf{W}_{ij}=\left(\mathbf{X}_{ij},\mathbf{S}_i,\mathbf{D}_j\right)$, $\theta=\left(\beta'\ \gamma'\ \delta'\right)'$, $\psi_{ij}^*=\left(\psi_{ij}',\left(\tau_{ij}+\mu_{ij}^*\right)\right)'$ and $\mathbf{H}=[\mathbf{F},\iota_T]$. Then, we derive the consistent estimator of θ (called 3D-PCCE) by[4]

[4] κ and φ cannot be identified due to the factor approximations and the Within transformation.

$$\hat{\theta}_{PCCE} = \left(\sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \mathbf{W}_{ij}' \mathbf{M}_H \mathbf{W}_{ij} \right)^{-1} \left(\sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \mathbf{W}_{ij}' \mathbf{M}_H \mathbf{Y}_{ij} \right), \tag{10.12}$$

where $\mathbf{M}_H = \mathbf{I}_T - \mathbf{H}\left(\mathbf{H}'\mathbf{H}\right)^{-1}\mathbf{H}'$. Following Pesaran (2006), it is straightforward to show that as $(N_1, N_2, T) \to \infty$, the PCCE estimator, (10.12) follows the asymptotic normal distribution (see also Kapetanios and Shin, 2017)

$$\sqrt{N_1 N_2 T}\left(\hat{\theta}_{PCCE} - \theta\right) \overset{a}{\sim} N\left(\mathbf{0}, \mathbf{\Sigma}_\theta\right),$$

where the (robust) consistent estimator of $\mathbf{\Sigma}_\theta$ is given by

$$\hat{\mathbf{\Sigma}}_\theta = \frac{1}{N_1 N_2} \mathbf{S}_\theta^{-1} \mathbf{R}_\theta \mathbf{S}_\theta^{-1},$$

$$\mathbf{R}_\theta = \frac{1}{N_1 N_2 - 1} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \left(\frac{\mathbf{W}_{ij}' \mathbf{M}_H \mathbf{W}_{ij}}{T} \right) \left(\hat{\theta}_{ij} - \hat{\theta}_{MG} \right) \left(\hat{\theta}_{ij} - \hat{\theta}_{MG} \right)' \left(\frac{\mathbf{W}_{ij}' \mathbf{M}_H \mathbf{W}_{ij}}{T} \right),$$

$$\mathbf{S}_\theta = \frac{1}{N_1 N_2} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \left(\frac{\mathbf{W}_{ij}' \mathbf{M}_H \mathbf{W}_{ij}}{T} \right), \; \hat{\theta}_{MG} = \frac{1}{N_1 N_2} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \hat{\theta}_{ij},$$

where $\hat{\theta}_{ij}$ is the (ij) pairwise OLS estimator obtained from the individual regression of $\mathbf{Y}_{ij}$ on $(\mathbf{W}_{ij}, \mathbf{H})$ in (10.11) for $i = 1, ..., N_1$ and $j = 1, ..., N_2$.

Next, we consider the 3D model (10.1) with the following general error components by combining CTFEs and heterogeneous global factors

$$u_{ijt} = \mu_{ij} + \upsilon_{it} + \zeta_{jt} + \pi_{ij}\lambda_t + \varepsilon_{ijt} \, . \tag{10.13}$$

It is straightforward to show that the 3D-Within transformation (10.3) fails to remove heterogeneous factors $\pi_{ij}\lambda_t$, because it is easily seen that

$$\tilde{u}_{ijt} = \tilde{\pi}_{ij}\tilde{\lambda}_t + \tilde{\varepsilon}_{ijt} \, ,$$

where $\tilde{\lambda}_t = \lambda_t - \bar{\lambda}$ with $\bar{\lambda} = T^{-1}\sum_{t=1}^{T}\lambda_t$ and $\tilde{\pi}_{ij} = \pi_{ij} - \bar{\pi}_{.j} - \bar{\pi}_{i.} + \bar{\pi}_{..}$ with $\bar{\pi}_{.j} = N_1^{-1}\sum_{i=1}^{N_1}\pi_{ij}$ and $\bar{\pi}_{i.} = N_2^{-1}\sum_{j=1}^{N_2}\pi_{ij}$.[5] It is clear in the presence of the nonzero correlation between $\mathbf{x}_{ijt}$ and λ_t that the 3D-Within estimator of β is biased.

We develop a two-step consistent estimation procedure. First, taking the cross-sectional averages of (10.1) and (10.13) over i and j, we have

$$\bar{y}_{..t} = \beta'\bar{\mathbf{x}}_{..t} + \gamma'\bar{\mathbf{s}}_{.t} + \delta'\bar{\mathbf{d}}_{.t} + \kappa'\mathbf{q}_t + \varphi'\bar{\mathbf{z}}_{..} + \bar{\mu}_{..} + \bar{\upsilon}_{.t} + \bar{\zeta}_{.t} + \bar{\pi}_{..}\lambda_t + \bar{\varepsilon}_{..t} \, , \tag{10.14}$$

where $\bar{\upsilon}_{.t} = N_1^{-1}\sum_{i=1}^{N_1}\upsilon_{it}$, $\bar{\zeta}_{.t} = N_2^{-1}\sum_{j=1}^{N_2}\zeta_{jt}$ and see (10.8) for other definitions. Hence, we augment the model (10.1) with the cross-sectional averages as

[5] Unless $\tilde{\pi}_{ij} = 0, \tilde{u}_{ijt} \neq \tilde{\varepsilon}_{ijt}$. This holds only if factor loadings, π_{ij} are homogeneous for all (i, j) pairs.

$$y_{ijt} = \beta' \mathbf{x}_{ijt} + \gamma' \mathbf{s}_{it} + \delta' \mathbf{d}_{jt} + \psi'_{ij} \mathbf{f}_t + \tau_{ij} + \mu^*_{ij} + \upsilon^*_{ijt} + \zeta^*_{ijt} + \varepsilon^*_{ijt}, \tag{10.15}$$

where $\upsilon^*_{ijt} = \upsilon_{it} - \frac{\pi_{ij}\bar{v}_{.t}}{\bar{\pi}_{..}}$, $\zeta^*_{ijt} = \zeta_{jt} - \frac{\pi_{ij}\bar{\zeta}_{.t}}{\bar{\pi}_{..}}$, and see (10.11) for other definitions. We rewrite (10.15) as

$$y_{ijt} = \beta' \mathbf{x}_{ijt} + \gamma' \mathbf{s}_{it} + \delta' \mathbf{d}_{jt} + \psi'_{ij} \mathbf{f}_t + \tau_{ij} + \mu^*_{ij} + \upsilon_{it} + \zeta_{jt} + \varepsilon^{**}_{ijt}, \tag{10.16}$$

where $\varepsilon^{**}_{ijt} = \varepsilon_{ijt} - \frac{\pi_{ij}}{\bar{\pi}_{..}}\bar{\varepsilon}_{..t} - \frac{\pi_{ij}\bar{v}_{.t}}{\bar{\pi}_{..}} - \frac{\pi_{ij}\bar{\zeta}_{.t}}{\bar{\pi}_{..}}$. Note that as $N_1, N_2 \to \infty$, $\varepsilon^{**}_{ijt} \to_p \varepsilon_{ijt}$ since $\bar{v}_{.t} \to_p 0$, $\bar{\zeta}_{.t} \to_p 0$ and $\bar{\varepsilon}_{..t} \to_p 0$. Next, we apply the 3D-Within transformation (10.3) to (10.16), and obtain[6]

$$\tilde{y}_{ijt} = \beta' \tilde{\mathbf{x}}_{ijt} + \tilde{\psi}'_{ij} \tilde{\mathbf{f}}_t + \tilde{\varepsilon}^{**}_{ijt}, \tag{10.17}$$

where $\tilde{\psi}_{ij} = \psi_{ij} - \bar{\psi}_{.j} - \bar{\psi}_{j.} + \bar{\psi}_{..}$, $\tilde{\mathbf{f}}_t = \mathbf{f}_t - \bar{\mathbf{f}}$ with $\bar{\mathbf{f}} = T^{-1}\sum_{t=1}^{T} \mathbf{f}_t$, and $\mathbf{f}_t$ is defined in (10.10). Rewriting (10.17) compactly as

$$\tilde{\mathbf{Y}}_{ij} = \tilde{\mathbf{X}}_{ij}\beta + \tilde{\mathbf{F}}\tilde{\psi}_{ij} + \tilde{\mathbf{E}}^{**}_{ij},\ i = 1, ..., N_1, j = 1, ..., N_2\ , \tag{10.18}$$

where

$$\underset{(T\times 1)}{\tilde{\mathbf{Y}}_{ij}} = \begin{bmatrix} \tilde{y}_{ij1} \\ \vdots \\ \tilde{y}_{ijT} \end{bmatrix},\ \underset{(T\times k_x)}{\tilde{\mathbf{X}}_{ij}} = \begin{bmatrix} \tilde{\mathbf{x}}'_{ij1} \\ \vdots \\ \tilde{\mathbf{x}}'_{ijT} \end{bmatrix},\ \underset{(T\times k_f)}{\tilde{\mathbf{F}}} = \begin{bmatrix} \tilde{\mathbf{f}}'_1 \\ \vdots \\ \tilde{\mathbf{f}}'_T \end{bmatrix},\ \underset{(T\times 1)}{\tilde{\mathbf{E}}^{**}_{ij}} = \begin{bmatrix} \tilde{\varepsilon}^{**}_{ij1} \\ \vdots \\ \tilde{\varepsilon}^{**}_{ijT} \end{bmatrix}.$$

Then, the 3D-PCCE estimator of β is obtained by

$$\hat{\beta}_{PCCE} = \left(\sum_{i=1}^{N_1}\sum_{j=1}^{N_2} \tilde{\mathbf{X}}'_{ij} \mathbf{M}_{\tilde{F}} \tilde{\mathbf{X}}_{ij} \right)^{-1} \left(\sum_{i=1}^{N_1}\sum_{j=1}^{N_2} \tilde{\mathbf{X}}'_{ij} \mathbf{M}_{\tilde{F}} \tilde{\mathbf{Y}}_{ij} \right), \tag{10.19}$$

where $\mathbf{M}_{\tilde{F}} = \mathbf{I}_T - \tilde{\mathbf{F}}\left(\tilde{\mathbf{F}}'\tilde{\mathbf{F}}\right)^{-1}\tilde{\mathbf{F}}'$ is the $(T \times T)$ idempotent matrix. Following Pesaran (2006) and Kapetanios and Shin (2017), it is also straightforward to show that as $(N_1, N_2, T) \to \infty$, the PCCE estimator, (10.19), follows the asymptotic normal distribution

$$\sqrt{N_1 N_2 T}\left(\hat{\beta}_{PCCE} - \beta\right) \overset{a}{\sim} N\left(\mathbf{0}, \mathbf{\Sigma}_\beta\right),$$

where the (robust) consistent estimator of $\mathbf{\Sigma}_\beta$ is given by

$$\hat{\mathbf{\Sigma}}_\beta = \frac{1}{N_1 N_2}\mathbf{S}_\beta^{-1}\mathbf{R}_\beta\mathbf{S}_\beta^{-1},$$

[6] It is clear that γ, δ, κ, and φ cannot be identified due to the 3D-Within transformation and the factor approximation. Define $\theta_{ijt} = \psi'_{ij}\mathbf{f}_t$, then it is straightforward to show that

$$\tilde{\theta}_{ijt} = \theta_{ijt} - \left(\bar{\theta}_{ij.} + \bar{\theta}_{.jt} + \bar{\theta}_{i.t}\right) + \left(\bar{\theta}_{..t} + \bar{\theta}_{i..} + \bar{\theta}_{.j.}\right) - \bar{\theta}_{...} = \left(\psi_{ij} - \bar{\psi}_{.j} - \bar{\psi}_{j.} + \bar{\psi}_{..}\right)'\left(\mathbf{f}_t - \bar{\mathbf{f}}\right)\ .$$

$$\mathbf{R}_{\beta}=\frac{1}{N_1N_2-1}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\left(\frac{\tilde{\mathbf{X}}_{ij}'\mathbf{M}_{\tilde{F}}\tilde{\mathbf{X}}_{ij}}{T}\right)\left(\hat{\beta}_{ij}-\hat{\beta}_{MG}\right)\left(\hat{\beta}_{ij}-\hat{\beta}_{MG}\right)'\left(\frac{\tilde{\mathbf{X}}_{ij}'\mathbf{M}_{\tilde{F}}\tilde{\mathbf{X}}_{ij}}{T}\right),$$

$$\mathbf{S}_{\beta}=\frac{1}{N_1N_2}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\left(\frac{\tilde{\mathbf{X}}_{ij}'\mathbf{M}_{\tilde{F}}\tilde{\mathbf{X}}_{ij}}{T}\right),\ \hat{\beta}_{MG}=\frac{1}{N_1N_2}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\hat{\beta}_{ij},$$

where $\hat{\beta}_{ij}$ is the (ij) pairwise OLS estimator obtained from the individual regression of $\tilde{\mathbf{Y}}_{ij}$ on $\left(\tilde{\mathbf{X}}_{ij},\tilde{\mathbf{F}}\right)$ in (10.18) for $i=1,...,N_1$ and $j=1,...,N_2$.

We can extend the proposed approach to the 3D panels with heterogeneous slope parameters

$$y_{ijt}=\beta_{ij}'\mathbf{x}_{ijt}+\gamma_j'\mathbf{s}_{it}+\delta_i'\mathbf{d}_{jt}+\kappa_{ij}'\mathbf{q}_t+\varphi'\mathbf{z}_{ij}+u_{ijt} \tag{10.20}$$

with $i=1,...,N_1,\ j=1,...,N_2,\ t=1,...,T$. In this case, we can develop the mean group estimators for (10.2), (10.7) and (10.13) in a straightforward manner (see, e.g., Pesaran, 2006; Kapetanios and Shin, 2017)

$$\hat{\beta}_{W,MG}=\frac{1}{N_1N_2}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\left(\tilde{\mathbf{X}}_{ij}'\tilde{\mathbf{X}}_{ij}\right)^{-1}\left(\tilde{\mathbf{X}}_{ij}'\mathbf{Y}_{ij}\right)$$

$$\hat{\theta}_{MGCCE}=\frac{1}{N_1N_2}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\left(\mathbf{W}_{ij}'\mathbf{M}_H\mathbf{W}_{ij}\right)^{-1}\left(\mathbf{W}_{ij}'\mathbf{M}_H\mathbf{Y}_{ij}\right)$$

$$\hat{\beta}_{MGCCE}=\frac{1}{N_1N_2}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\left(\tilde{\mathbf{X}}_{ij}'\mathbf{M}_{\tilde{F}}\tilde{\mathbf{X}}_{ij}\right)^{-1}\left(\tilde{\mathbf{X}}_{ij}'\mathbf{M}_{\tilde{F}}\tilde{\mathbf{Y}}_{ij}\right),$$

10.3 Cross-sectional Dependence (CD) Test

We discuss next the extent of cross-sectional dependence in 3D panel data models. Following Pesaran (2015) and Bailey et al. (2016b) (hereafter BKP), we can show that the extent of CSD is captured by the non-zero covariance between u_{ijt} and $u_{i'j't}$ for $i\neq i'$ and $j\neq j'$, denoted as $\sigma_{ijt,u}$. Here, the extent of CSD involves both N_1 and N_2, and thus relates to the rate at which $\frac{1}{N_1N_2}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\sigma_{ijt,u}$ declines with the product, N_1N_2. First, we consider the 3D model (10.1) with country-time effects (10.2). Following BBMP, we make the following random effects assumptions

$$\mu_{ij}\sim\text{iid}\left(0,\sigma_{\mu}^2\right),\ \upsilon_{it}\sim\text{iid}\left(0,\sigma_{v}^2\right),\ \zeta_{jt}\sim\text{iid}\left(0,\sigma_{\zeta}^2\right),\ \varepsilon_{ijt}\sim\text{iid}\left(0,\sigma_{\varepsilon}^2\right)$$

and μ_{ij}, υ_{it}, ζ_{jt} and ε_{ijt} are pairwise uncorrelated. Rewrite (10.2) sequentially as

$$\underset{(T\times1)}{\mathbf{u}_{ij}}=\mu_{ij}\iota_T+\mathbf{v}_i+\zeta_j+\varepsilon_{ij},\ i=1,...,N_1,\ j=1,...,N_2,$$

$$\underset{(N_2T\times 1)}{\mathbf{u}_i} = (\mu_i \otimes \iota_T) + (\iota_{N_2} \otimes \mathbf{v}_i) + \zeta + \varepsilon_i, i = 1,...,N_1,$$

$$\underset{(N_1N_2T\times 1)}{\mathbf{u}} = (\mu \otimes \iota_T) + \mathbf{V} + (\iota_{N_1} \otimes \zeta) + \varepsilon\,, \tag{10.21}$$

where

$$\underset{(T\times 1)}{\mathbf{u}_{ij}} = \begin{bmatrix} u_{ij1} \\ \vdots \\ u_{ijT} \end{bmatrix}, \underset{(T\times 1)}{\mathbf{v}_i} = \begin{bmatrix} v_{i1} \\ \vdots \\ v_{iT} \end{bmatrix}, \underset{(T\times 1)}{\zeta_j} = \begin{bmatrix} \zeta_{j1} \\ \vdots \\ \zeta_{jT} \end{bmatrix}, \underset{(T\times 1)}{\varepsilon_{ij}} = \begin{bmatrix} \varepsilon_{ij1} \\ \vdots \\ \varepsilon_{ijT} \end{bmatrix},$$

$$\underset{(N_2T\times 1)}{\mathbf{u}_i} = \begin{bmatrix} \mathbf{u}_{i1} \\ \vdots \\ \mathbf{u}_{iN_2} \end{bmatrix}, \underset{(N_2\times 1)}{\mu_i} = \begin{bmatrix} \mu_{i1} \\ \vdots \\ \mu_{iN_2} \end{bmatrix}, \underset{(N_2T\times 1)}{\zeta} = \begin{bmatrix} \zeta_1 \\ \vdots \\ \zeta_{N_2} \end{bmatrix}, \underset{(N_2T\times 1)}{\varepsilon_i} = \begin{bmatrix} \varepsilon_{i1} \\ \vdots \\ \varepsilon_{iN_2} \end{bmatrix},$$

$$\begin{aligned} \underset{(N_1N_2T\times 1)}{\mathbf{u}} &= \begin{bmatrix} \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_{N_1} \end{bmatrix}, \quad \underset{(N_1N_2\times 1)}{\mu} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_{N_1} \end{bmatrix}, \\ \underset{(N_1N_2T\times 1)}{\mathbf{V}} &= \begin{bmatrix} (\iota_{N_2} \otimes \mathbf{v}_1) \\ \vdots \\ (\iota_{N_2} \otimes \mathbf{v}_{N_1}) \end{bmatrix}, \quad \underset{(N_1N_2T\times 1)}{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_{N_1} \end{bmatrix}. \end{aligned} \tag{10.22}$$

Then, it is easily seen that (see also BBMP)

$$\begin{aligned} \underset{(N_1N_2T\times N_1N_2T)}{\mathrm{Cov}(\mathbf{u})} &= \sigma_\mu^2 (\mathbf{I}_{N_1N_2} \otimes \mathbf{J}_T) + \sigma_v^2 (\mathbf{I}_{N_1} \otimes \mathbf{J}_{N_2} \otimes \mathbf{I}_T) \\ &\quad + \sigma_\zeta^2 (\mathbf{J}_{N_1} \otimes \mathbf{I}_{N_2T}) + \sigma_\varepsilon^2 \mathbf{I}_{N_1N_2T}\,. \end{aligned} \tag{10.23}$$

Note that the CTRE model imposes a very limited structure on the CSD, because for $i \neq i'$ and $j \neq j'$, we have

$$\mathrm{E}[u_{ijt}u_{ij't}] = \sigma_v^2,\ \mathrm{E}[u_{ijt}u_{ij't}] = \sigma_\zeta^2 \text{ and } \mathrm{E}[u_{ijt}u_{i'j't}] = 0. \tag{10.24}$$

This suggests that the covariance between u_{ijt} and $u_{i'jt}$ is common σ_v^2 for all $i = 1,...,N_1$, while the covariance between u_{ijt} and $u_{ij't}$ is common σ_ζ^2 for all $j = 1,...,N_2$. Further, it imposes zero covariance between u_{ijt} and $u_{ij't'}$.

Next, we consider the 3D model with the two-way heterogeneous factor specification, (10.7). In this case, it is straightforward to derive

$$\underset{(N_1N_2T\times 1)}{\mathbf{u}} = (\mu \otimes \iota_T) + (\pi \otimes \lambda_T) + \varepsilon\,, \tag{10.25}$$

where

$$\underset{(N_1N_2\times 1)}{\pi} = \begin{bmatrix} \pi_1 \\ \vdots \\ \pi_{N_1} \end{bmatrix}, \underset{(N_2\times 1)}{\pi_i} = \begin{bmatrix} \pi_{i1} \\ \vdots \\ \pi_{iN_2} \end{bmatrix}, \underset{(T\times 1)}{\lambda_T} = \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_T \end{bmatrix},$$

and see (10.22) for other definitions. The covariance matrix for $\mathbf{u}$ in (10.25) is

$$\underset{(N_1N_2T\times N_1N_2T)}{\mathrm{Cov}(\mathbf{u})} = \sigma_\mu^2(\mathbf{I}_{N_1N_2}\otimes\mathbf{J}_T)+\sigma_\lambda^2(\pi\pi'\otimes\mathbf{I}_T)+\sigma_\varepsilon^2\mathbf{I}_{N_1N_2T}\,. \tag{10.26}$$

Thus, the specification (10.25) can capture CSD by non-zero covariances between u_{ijt} and $u_{i'j't}$ for $i\neq i'$ and $j\neq j'$ by

$$\mathrm{E}[u_{ijt}u_{ij't}]=\pi_{ij}\pi_{ij'}\sigma_\lambda^2,\ \mathrm{E}[u_{ijt}u_{i'jt}]=\pi_{ij}\pi_{i'j}\sigma_\lambda^2,\ \mathrm{E}[u_{ijt}u_{i'j't}]=\pi_{ij}\pi_{i'j'}\sigma_\lambda^2. \tag{10.27}$$

Next, we consider the 3D model with more general error components, (10.13). Combining the above results, it is straightforward to derive

$$\underset{(N_1N_2T\times 1)}{\mathbf{u}} = (\mu\otimes\iota_T)+\mathbf{V}+(\iota_{N_1}\otimes\zeta)+(\pi\otimes\lambda_T)+\varepsilon\,. \tag{10.28}$$

Thus, the covariance matrix for $\mathbf{u}$ in (10.28) is given by

$$\begin{aligned}\underset{(N_1N_2T\times N_1N_2T)}{\mathrm{Cov}(\mathbf{u})} &= \sigma_\mu^2(\mathbf{I}_{N_1N_2}\otimes\mathbf{J}_T)+\sigma_v^2(\mathbf{I}_{N_1}\otimes\mathbf{J}_{N_2}\otimes\mathbf{I}_T) \\ &\quad+\sigma_\zeta^2(\mathbf{J}_{N_1}\otimes\mathbf{I}_{N_2T})+\sigma_\lambda^2(\pi\pi'\otimes\mathbf{I}_T)+\sigma_\varepsilon^2\mathbf{I}_{N_1N_2T}\,.\end{aligned} \tag{10.29}$$

This model can capture CSD by non-zero covariances between u_{ijt} and $u_{i'j't}$ for $i\neq i'$ and $j\neq j'$, given by

$$\mathrm{E}[u_{ijt}u_{ij't}]=\pi_{ij}\pi_{ij'}\sigma_\lambda^2+\sigma_v^2,\mathrm{E}[u_{ijt}u_{i'jt}]=\pi_{ij}\pi_{i'j}\sigma_\lambda^2+\sigma_\xi^2,\mathrm{E}[u_{ijt}u_{i'j't}]=\pi_{ij}\pi_{i'j'}\sigma_\lambda^2. \tag{10.30}$$

Comparing (10.24), (10.27) and (10.30), we find that the CTFE specification in (10.2) can only accommodate non-zero covariances locally, but it also imposes the same covariance for all $i=1,...,N_1$ and $j=1,...,N_2$, respectively. Such restrictions are too strong to hold in practice.[7]

In contrast, our proposed error components specification (10.13) can accommodate non-zero covariances both locally and globally.

Note that υ_{it} and ζ_{jt} are related to the local-time factors. In order to examine whether they exhibit weak or strong CSD, we consider the following heterogeneous local factors specifications

$$\upsilon_{it}=\upsilon_i\tau_t \text{ and } \zeta_{jt}=\zeta_j\tau_t^*\,,$$

where τ_t and τ_t^* are the origin and the destination-specific local common factors, respectively. Then, (10.13) can be replaced by

$$u_{ijt}=\mu_{ij}+\upsilon_i\tau_t+\zeta_j\tau_t^*+\varepsilon_{ijt}\,. \tag{10.31}$$

[7] In the two-way error components with individual effects and time effects, the cross-sectional correlation is the same for all cross-sectional pairs. Serlenga and Shin (2007) show that such specification would produce very misleading results in the presence of heterogeneous strong CSD in a 2D panel data model.

This specification implies that the exporter, i, reacts heterogeneously to the common import market conditions, τ_t and the importer, j, reacts heterogeneously to the common export market conditions, τ_t^*. Recently, Kapetanios and Shin (2017) proposed a more general hierarchical multi-factor error components specification

$$u_{ijt} = \mu_{ij} + \upsilon_i \tau_t + \zeta_j \tau_t^* + \pi_{ij} \lambda_t + \varepsilon_{ijt}. \tag{10.32}$$

Within this model, we can distinguish between three types of CSD: (i) the strong global factor, λ_t which influences the (ij) pairwise interactions (of $N_1 N_2$ dimension); (ii) the semi-strong local factors, τ_t and τ_t^*, which influence origin or destination countries separately (each of N_1 or N_2 dimension); and (iii) the weak CSD idiosyncratic errors, ε_{ijt}. We expect that this kind of generalisation would be most natural within the 3D panel data models. Following BBMP, we assume

$$\begin{array}{llll} \mu_{ij} \sim \text{iid}\left(0, \sigma_\mu^2\right), & \tau_t \sim \text{iid}\left(0, \sigma_\tau^2\right), & \tau_t^* \sim \text{iid}\left(0, \sigma_{\tau^*}^2\right), \\ \lambda_t \sim \text{iid}\left(0, \sigma_\lambda^2\right), & \varepsilon_{ijt} \sim \text{iid}\left(0, \sigma_\varepsilon^2\right) & \end{array}$$

and μ_{ij}, τ_t, τ_t^*, λ_t and ε_{ijt} are mutually independent.[8] It is clear that model (10.32) can capture CSD by non-zero covariances between u_{ijt} and $u_{i'j't}$ for $i \neq i'$ and $j \neq j'$, given by

$$\mathrm{E}[u_{ijt} u_{ij't}] = \upsilon_i^2 \sigma_\tau^2 + \pi_{ij} \pi_{ij'} \sigma_\lambda^2, \quad \mathrm{E}[u_{ijt} u_{i'jt}] = \zeta_j^2 \sigma_{\tau^*}^2 + \pi_{ij} \pi_{i'j} \sigma_\lambda^2$$

$$\mathrm{E}[u_{ijt} u_{i'j't}] = \pi_{ij} \pi_{i'j'} \sigma_\lambda^2. \tag{10.33}$$

The covariance structure in (10.33) is clearly more general than (10.30).

We now develop a diagnostic test for the null hypothesis of residual cross-sectional independence in the triple-index panel data models. These are evaluated using the residuals obtained respectively from (10.5), (10.11) and (10.18), which we denote as $\mathbf{e}_{ij} = \left(e_{ij1}, ..., e_{ijT}\right)'$. In particular, we have $\mathbf{e}_{ij} = \tilde{\mathbf{Y}}_{ij} - \tilde{\mathbf{X}}_{ij} \hat{\beta}_W$ for model (10.5), $\mathbf{e}_{ij} = \mathbf{M}_H \mathbf{Y}_{ij} - \mathbf{M}_H \mathbf{W}_{ij} \hat{\theta}_{PCCE}$ for model (10.11), and finally $\mathbf{e}_{ij} = \mathbf{M}_{\bar{F}} \tilde{\mathbf{Y}}_{ij} - \mathbf{M}_{\bar{F}} \tilde{\mathbf{X}}_{ij} \hat{\beta}_{PCCE}$ for the model (10.18).

The proposed cross-sectional dependence (CD) test is a modified counterpart of an existing CD test proposed by Pesaran (2015). For convenience, we represent $\mathbf{e}_{ij}$ as the (ij) pair using the single index $n = 1, ..., N_1 N_2$, and compute the pair-wise residual correlations between n and n' cross-sectional units by

$$\hat{\rho}_{nn'} = \frac{\mathbf{e}_n' \mathbf{e}_{n'}}{\sqrt{\left(\mathbf{e}_n' \mathbf{e}_n\right)\left(\mathbf{e}_{n'}' \mathbf{e}_{n'}\right)}}, \; n, n' = 1, ..., N_1 N_2 \text{ and } n \neq n'.$$

Then, we construct the CD statistic by

$$CD = \sqrt{\frac{2}{N_1 N_2 \left(N_1 N_2 - 1\right)}} \sum_{n=1}^{N_1 N_2 - 1} \sum_{n'=n+1}^{N_1 N_2} \sqrt{T} \hat{\rho}_{nn'}. \tag{10.34}$$

[8] This assumption is still more general than the random effects assumptions made in BBMP.

Pesaran derives that the CD test has the limiting $N(0,1)$ distribution under the null hypothesis of cross-sectional error independence, namely $H_0 : \hat{\rho}_{nn'} = 0$ for all $n, n' = 1, ..., N_1N_2$ and $n \neq n'$. Following BKP, Pesaran further shows that the CD statistic, (10.34), can also be applicable to testing the null hypothesis of weak cross-sectional error dependence. As an extension, one can also construct hierarchical CD tests based on a 2D sub-dataset out of the 3D dataset.

Following BKP, we introduce the exponent of cross-sectional dependence based on the double cross-sectional averages defined as $\bar{u}_{..t} = (N_1N_2)^{-1} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} u_{ijt}$. If u_{ijt}'s are cross-sectionally correlated across (i, j) pairs, $\text{Var}(\bar{u}_{..t})$ declines at a rate that is a function of α, where α is defined as

$$\lim_{N_1N_2 \to \infty} (N_1N_2)^{-\alpha} \lambda_{\max}(\boldsymbol{\Sigma}_u)$$

and $\boldsymbol{\Sigma}_u$ is the $(N_1N_2 \times N_1N_2)$ covariance matrix of $\mathbf{u}_t = (u_{11t}, ..., u_{N_1N_2t})'$ with $\lambda_{\max}(\boldsymbol{\Sigma}_u)$ denoting the largest eigenvalue. Clearly, $\text{Var}(\bar{u}_{..t})$ cannot decline at a rate faster than $(N_1N_2)^{-1}$, as well as it cannot decline at a rate slower than $(N_1N_2)^{\alpha-1}$ with $0 \leq \alpha \leq 1$. Given that

$$\text{Var}(\bar{u}_{..t}) \leq (N_1N_2)^{-1} \lambda_{\max}(\boldsymbol{\Sigma}_u),$$

we find that α defined by $(N_1N_2)^{-1} \lambda_{\max}(\boldsymbol{\Sigma}_u) = O\left((N_1N_2)^{\alpha-1}\right)$ will provide an upper bound for $\text{Var}(\bar{u}_{..t})$. The extent of CSD depends on the nature of the factor loadings in the following factor-based errors

$$u_{ijt} = \pi_{ij}\lambda_t + \varepsilon_{ijt}.$$

If the average of the heterogenous loading parameters, π_{ij}, denoted μ_π, is bounded away from zero, the cross-sectional dependence will be strong, in which case, $(N_1N_2)^{-1} \lambda_{\max}(\boldsymbol{\Sigma}_u)$ and $\text{Var}(\bar{u}_{..t})$ are both $O(1)$, which yields $\alpha = 1$.

Furthermore, for $1/2 < \alpha \leq 1$, BKP propose the following bias-adjusted estimator to consistently estimate α:

$$\mathring{\alpha} = 1 + \frac{1}{2} \frac{\ln\left(\hat{\sigma}^2_{\bar{u}_{..t}}\right)}{\ln(N_1N_2)} - \frac{\ln\left(\hat{\mu}^2_\pi\right)}{2\ln(N_1N_2)} - \frac{\hat{c}_{N_1N_2}}{2\left[N_1N_2 \ln(N_1N_2)\, \hat{\sigma}^2_{\bar{u}_{..t}}\right]}, \tag{10.35}$$

where

$$\hat{\sigma}^2_{\bar{u}_{..t}} = T^{-1} \Sigma_{t=1}^T \bar{u}^2_{..t}, \quad \hat{c}_{N_1N_2} = \widehat{\bar{\sigma}^2_{N_1N_2}} = (N_1N_2)^{-1} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \hat{\sigma}^2_{ij},$$

and $\hat{\sigma}^2_{ij} = T^{-1} \Sigma_{t=1}^T \hat{\varepsilon}^2_{ijt}$ is the ijth diagonal element of the estimated covariance matrix, $\hat{\Sigma}_\varepsilon$ with $\hat{\varepsilon}_{ijt} = u_{ijt} - \hat{\delta}_{ij}\bar{u}_{..t}$ and $\hat{\delta}_{ij}$ is the OLS estimator from the regression of u_{ijt} on $\bar{u}_{..t}$. BKP also argue that a suitable estimation of μ^2_π can be derived, noting that μ_π is the mean of the population regression coefficient of u_{ijt} on $\tilde{u}_{..t} = \bar{u}_{..t}/\hat{\sigma}_{\bar{u}_{..t}}$ for units u_{ijt}, that have at least one non-zero loading, and those units are selected using Holm's (1979) multiple testing approach.

In the empirical section, we apply the above 3D extension of the BKP estimation and testing techniques directly to the residuals e_{ijt} obtained respectively from (10.5), (10.11) and (10.18). We also evaluate the confidence band for the estimated CSD exponent by employing the test statistic defined in (B47) in BKP's Supplementary Appendix VI.

10.4 Extensions

We provide two extensions of the proposed estimation techniques into unbalanced panels and four dimensional (4D) models. Such extensions are challenging as they involve several layers of factor specifications.

10.4.1 Unbalanced Panels

In practice, we may be faced with unbalanced panels. Note, however, that the issue of unbalanced panels or missing data has been almost neglected even in the literature on 2D panels with unobserved factors or interactive effects. Kapetanios and Pesaran (2005) briefly deal with this issue in their Monte Carlo studies. Bai et al. (2015) investigate the unbalanced 2D panel data model with interactive effects, and propose the use of the functional principal components analysis and the EM algorithm. Via simulation studies, they find that the EM-type estimators are consistent for both smooth and stochastic factors, though no asymptotic analysis is provided. For the error components model (10.2), the 3D Within transformation fails to fully eliminate the fixed effects. BMW thus extend the Wansbeek and Kapteyn (1989) approach and derive a complex Within transformation, which is computationally quite demanding as it may involve an inversion of $(NT \times NT)$ matrices. Thus, we expect that this extension of our proposed 3D PCCE estimation into unbalanced panels will be more challenging.

We now introduce a vector of selection indicators for each pair (i,j), $\mathbf{s}_{ij} = \left(s_{ij,1},...,s_{ij,T}\right)'$, where $s_{ijt} = 1$ if time period t for pair (i,j) can be used in the estimation. We only use information on units where a full set of data is observed. Therefore, $s_{ijt} = 1$ if and only if (x_{ijt}, y_{ijt}) is fully observed; otherwise, $s_{ijt} = 0$. Following Wooldridge (2010), we assume that selection is ignorable conditional on $(\mathbf{x}_{ijt}, \mathbf{s}_{it}, \mathbf{d}_{jt}, \mathbf{q}_t, \mathbf{z}_{ij}, \mu_{ij}, \lambda_t)$:

$$\mathrm{E}\left(y_{it}|\mathbf{x}_{ijt}, \mathbf{s}_{it}, \mathbf{d}_{jt}, \mathbf{q}_t, \mathbf{z}_{ij}, \mu_{ij}, \lambda_t, \mathbf{s}_i\right) = \mathrm{E}\left(y_{it}|\mathbf{x}_{ijt}, \mathbf{s}_{it}, \mathbf{d}_{jt}, \mathbf{q}_t, \mathbf{z}_{ij}, \mu_{ij}, \lambda_t\right).$$

Let $n = \sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\sum_{t=1}^{T} s_{ijt}$ be the total number of observations. Also define $n_t = \sum_{i=1}^{N_1}\sum_{j=1}^{N_2} s_{ijt}$ and $n_{ij} = \sum_{t=1}^{T} s_{ijt}$ as the number of cross-sectional pairs observed for time period t and the number of time periods observed for pair (i,j). Similarly, define $n_i = \sum_{j=1}^{N_2}\sum_{t=1}^{T} s_{ijt}$, $n_j = \sum_{i=1}^{N_1}\sum_{t=1}^{T} s_{ijt}$, $n_{it} = \sum_{j=1}^{N_2} s_{ijt}$ and $n_{jt} = \sum_{i=1}^{N_1} s_{ijt}$,

respectively. To simplify further analysis, we maintain the assumption[9]

$$\left(\min_i n_i, \min_j n_j, \min_t n_t, \min_{(ij)} n_{ij}\right) \to \infty \quad \text{or} \quad \left(\min_t n_t, \min_{(ij)} n_{ij}\right) \to \infty .$$

Consider the 3D model (10.1) with the error components specification (10.7). We multiply the model by the selection indicator to get

$$y_{ijt}^s = \beta' \mathbf{x}_{ijt}^s + \gamma' \mathbf{s}_{it}^s + \delta' \mathbf{d}_{jt}^s + \kappa' \mathbf{q}_t^s + \varphi' \mathbf{z}_{ij}^s + \mu_{ij}^s + \pi_{ij}^s \lambda_t + \varepsilon_{ijt}^s, \tag{10.36}$$

where $y_{ijt}^s = s_{ijt} y_{ijt}$, $\mathbf{x}_{ijt}^s = s_{ijt}\mathbf{x}_{ijt}$, $\mathbf{s}_{it}^s = s_{ijt}\mathbf{s}_{it}$, $\mathbf{d}_{jt}^s = s_{ijt}\mathbf{d}_{jt}$, $\mathbf{q}_t^s = s_{ijt}\mathbf{q}_t$, $\mathbf{z}_{ij}^s = s_{ijt}\mathbf{z}_{ij}$, $\mu_{ij}^s = s_{ijt}\mu_{ij}$, $\pi_{ij}^s = s_{ijt}\pi_{ij}$, and $\varepsilon_{ijt}^s = s_{ijt}\varepsilon_{ijt}$. Applying the cross-sectional averages of (10.36) over i and j, we obtain

$$\bar{y}_{..t}^s = \beta' \bar{\mathbf{x}}_{..t}^s + \gamma' \bar{\mathbf{s}}_{.t}^s + \delta' \bar{\mathbf{d}}_{.t}^s + \kappa' \mathbf{q}_t + \varphi' \bar{\mathbf{z}}_{..t}^s + \bar{\mu}_{..t}^s + \bar{\pi}_{..t}^s \lambda_t + \bar{\varepsilon}_{..t}^s , \tag{10.37}$$

where $\bar{y}_{..t}^s = \frac{1}{n_t}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2} s_{ijt} y_{ijt} = \sum_{i=1}^{N_1} w_{it} \bar{y}_{i.t}^s$ is expressed as a weighted average with $w_{it} = n_{it}/n_t$ and $\bar{y}_{i.t}^s = n_{it}^{-1}\sum_{j=1}^{N_2} s_{ijt} y_{ijt}$.[10] Similarly for $\bar{\mathbf{x}}_{..t}^s$, $\bar{\mathbf{z}}_{..t}^s$, $\bar{\mu}_{..t}^s$, $\bar{\pi}_{..t}^s$ and $\bar{\varepsilon}_{..t}^s$. Further, $\bar{\mathbf{s}}_{.t}^s = \sum_{i=1}^{N_1} w_{it}\mathbf{s}_{it}$, $\bar{\mathbf{d}}_{.t}^s = \sum_{j=1}^{N_2} w_{jt}\mathbf{d}_{jt}$ with $w_{jt} = n_{jt}/n_t$, and $\bar{\mathbf{q}}_t^s = \mathbf{q}_t$. As $n_t \to \infty$,

$$\bar{\mathbf{z}}_{..t}^s = \bar{\mathbf{z}} + o_p(1),\ \bar{\mu}_{..t}^s = \bar{\mu} + o_p(1),\ \bar{\pi}_{..t}^s = \bar{\pi} + o_p(1) \text{ and } \bar{\varepsilon}_{..t}^s = \bar{\varepsilon}_{..t} + o_p(1) , \tag{10.38}$$

where $\bar{\mathbf{z}} = (N_1 N_2)^{-1}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2} \mathbf{z}_{ij} \to_p \mathrm{E}(\mathbf{z}_{ij})$, $\bar{\mu} = (N_1 N_2)^{-1}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2} \mu_{ij} \to_p 0$, $\bar{\pi} = (N_1 N_2)^{-1}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2} \pi_{ij} \to_p \mathrm{E}(\pi_{ij}) \neq 0$ and $\bar{\varepsilon}_{..t} = (N_1 N_2)^{-1}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2} \varepsilon_{ijt} \to_p 0$. Using (10.38), we rewrite (10.37) as

$$\bar{y}_{..t}^s = \beta' \bar{\mathbf{x}}_{..t}^s + \gamma' \bar{\mathbf{s}}_{.t}^s + \delta' \bar{\mathbf{d}}_{.t}^s + \kappa' \mathbf{q}_t + \varphi' \bar{\mathbf{z}} + \bar{\mu} + \bar{\pi}\lambda_t + \bar{\varepsilon}_{..t} + o_p(1) .$$

Hence, λ_t can be approximated by

$$\lambda_t \simeq \frac{1}{\bar{\pi}} \left\{ \bar{y}_{..t}^s - \left(\beta' \bar{\mathbf{x}}_{..t}^s + \gamma' \bar{\mathbf{s}}_{.t}^s + \delta' \bar{\mathbf{d}}_{.t}^s + \kappa' \mathbf{q}_t + \varphi' \bar{\mathbf{z}} + \bar{\mu} + \bar{\varepsilon}_{..t} \right) \right\} .$$

Using these results, we can augment the model (10.36) with the cross-sectional averages as follows

$$y_{ijt}^s = \beta' \mathbf{x}_{ijt}^s + \gamma' \mathbf{s}_{it}^s + \delta' \mathbf{d}_{jt}^s + \psi_{ij}' \mathring{\mathbf{f}}_t^s + \tau_{ij}^s + \mu_{ij}^s + \varepsilon_{ijt}^{*s}, \tag{10.39}$$

where $\tau_{ij}^s = s_{ijt}\tau_{ij}$, $\varepsilon_{ijt}^{*s} = s_{ijt}\varepsilon_{ijt}^*$ and

$$\mathring{\mathbf{f}}_t^s = s_{ijt}\mathbf{f}_t^s \text{ with } \mathbf{f}_t^s = \left(\bar{y}_{..t}^s, \bar{\mathbf{x}}_{..t}^{s\prime}, \bar{\mathbf{s}}_{.t}^{s\prime}, \bar{\mathbf{d}}_{.t}^{s\prime}, \mathbf{q}_t' \right)' . \tag{10.40}$$

[9] For factor approximation by cross-sectional averages or the principal components to be valid, we still require $\min_{(ij)} n_{ij} \to \infty$, see Pesaran (2006) and Bai (2009).

[10] $\bar{y}_{..t}^s$ can be expressed as a (column sum) weighted average $\sum_{j=1}^{N_2} w_{jt}\bar{y}_{.jt}^s$ with $w_{jt} = n_{jt}/n_t$ and $\bar{y}_{.jt}^s = n_{jt}^{-1}\sum_{i=1}^{N_1} s_{ijt} y_{ijt}$.

Collecting only the n_{ij} observations with $s_{ijt} = 1$ from (10.39), we have

$$\mathbf{Y}_{ij} = \mathbf{X}_{ij}\beta + \mathbf{S}_{ij}\gamma + \mathbf{D}_{ij}\delta + \mathbf{F}_{ij}\psi_{ij} + (\tau_{ij} + \mu_{ij})\,\iota_{n_{ij}} + \mathbf{E}^*_{ij} = \mathbf{W}_{ij}\theta + \mathbf{H}_{ij}\psi^*_{ij} + \mathbf{E}^*_{ij}\,, \tag{10.41}$$

where

$$\mathbf{W}_{ij} = \left(\mathbf{X}_{ij}, \mathbf{S}_{ij}, \mathbf{D}_{ij}\right)\,, \quad \theta = \left(\beta'\ \gamma'\ \delta'\right)'\,,$$
$$\psi^*_{ij} = \left(\psi'_{ij}, (\tau_{ij} + \mu_{ij})\right)'\,, \quad \mathbf{H}_{ij} = \left[\mathbf{F}_{ij}, \iota_{n_{ij}}\right]$$

and

$$\underset{(n_{ij}\times 1)}{\mathbf{Y}_{ij}} = \begin{bmatrix} y_{ij(1)} \\ \vdots \\ y_{ij(n_{ij})} \end{bmatrix}, \quad \underset{(T_{ij}\times k_x)}{\mathbf{X}_{ij}} = \begin{bmatrix} \mathbf{x}'_{ij(1)} \\ \vdots \\ \mathbf{x}'_{ij(n_{ij})} \end{bmatrix}, \quad \underset{(T_{ij}\times k_s)}{\mathbf{S}_{ij}} = \begin{bmatrix} \mathbf{s}'_{i(1)} \\ \vdots \\ \mathbf{s}'_{i(n_{ij})} \end{bmatrix},$$

$$\underset{(n_{ij}\times k_d)}{\mathbf{D}_{ij}} = \begin{bmatrix} \mathbf{d}'_{j(1)} \\ \vdots \\ \mathbf{d}'_{j(n_{ij})} \end{bmatrix}, \quad \underset{(n_{ij}\times k_f)}{\mathbf{F}_{ij}} = \begin{bmatrix} \mathbf{f}^{s\prime}_{(1)} \\ \vdots \\ \mathbf{f}^{s\prime}_{(n_{ij})} \end{bmatrix}, \quad \underset{(T_{ij}\times 1)}{\mathbf{E}_{ij}} = \begin{bmatrix} \varepsilon^*_{ij(1)} \\ \vdots \\ \varepsilon^*_{ij(n_{ij})} \end{bmatrix}.$$

Here we express the time index inside $(.)$ to highlight different initial and last time periods for each cross-sectional pair (ij) respectively. Then, the 3D-PCCE estimator of θ is obtained by

$$\hat{\theta}_{PCCE} = \left(\sum_{i=1}^{N_1}\sum_{j=1}^{N_2} \mathbf{W}'_{ij}\mathbf{M}_{H_{ij}}\mathbf{W}_{ij}\right)^{-1} \left(\sum_{i=1}^{N_1}\sum_{j=1}^{N_2} \mathbf{W}'_{ij}\mathbf{M}_{H_{ij}}\mathbf{Y}_{ij}\right)\,, \tag{10.42}$$

where $\mathbf{M}_{H_{ij}} = \mathbf{I}_{T_{ij}} - \mathbf{H}_{ij}\left(\mathbf{H}'_{ij}\mathbf{H}_{ij}\right)^{-1}\mathbf{H}'_{ij}$.

Next, we consider the 3D model (10.1) with the general error components specification (10.13). To develop the two-step consistent estimation procedure for unbalanced panels, we multiply the model by $s_{ijt} = 1$ to get

$$y^s_{ijt} = \beta'\mathbf{x}^s_{ijt} + \gamma'\mathbf{s}^s_{it} + \delta'\mathbf{d}^s_{jt} + \kappa'\mathbf{q}^s_t + \varphi'\mathbf{z}^s_{ij} + \mu^s_{ij} + \upsilon^s_{it} + \zeta^s_{jt} + \pi_{ij}\lambda^s_t + \varepsilon^s_{ijt}\,, \tag{10.43}$$

where $y^s_{ijt} = s_{ijt}y_{ijt}$ and similarly for others. Taking the cross-sectional averages of (10.43) over i and j, we have

$$\bar{y}^s_{..t} = \beta'\bar{\mathbf{x}}^s_{..t} + \gamma'\bar{\mathbf{s}}^s_{.t} + \delta'\bar{\mathbf{d}}^s_{.t} + \kappa'\mathbf{q}_t + \varphi'\bar{\mathbf{z}}^s_{..t} + \bar{\mu}^s_{..t} + \bar{v}^s_{.t} + \bar{\zeta}^s_{.t} + \bar{\pi}^s_{..t}\lambda_t + \bar{\varepsilon}^s_{..t}\,, \tag{10.44}$$

where $\bar{v}^s_{.t} = \sum_{i=1}^{N_1} w_{it}\upsilon_{it}$ with $w_{it} = n_{it}/n_t$, $\bar{\zeta}^s_{.t} = \sum_{j=1}^{N_2} w_{jt}\zeta_{jt}$ with $w_{jt} = n_{jt}/n_t$, and see (10.37) for other definitions. As $n_t \to \infty$**,**

$$\bar{v}^s_{.t} = \bar{v} + o_p(1) \text{ and } \bar{\zeta}^s_{.t} = \bar{\zeta} + o_p(1)\,, \tag{10.45}$$

where $\bar{v} = N_1^{-1}\sum_{i=1}^{N_1}\upsilon_{it} \to_p 0$ and $\bar{\zeta} = N_2^{-1}\sum_{j=1}^{N_2}\zeta_{jt} \to_p 0$. Using (10.38) and (10.45), we can approximate $\bar{y}^s_{..t}$ and λ_t by

$$\bar{y}^s_{..t} = \beta' \bar{\mathbf{x}}^s_{..t} + \gamma' \bar{\mathbf{s}}^s_{.t} + \delta' \bar{\mathbf{d}}^s_{.t} + \kappa' \mathbf{q}_t + \varphi' \bar{\mathbf{z}} + \bar{\mu} + \bar{v} + \bar{\zeta} + \bar{\pi} \lambda_t + \bar{\varepsilon}_{..t} + o_p(1)$$

$$\lambda_t = \frac{1}{\bar{\pi}} \left\{ \bar{y}^s_{..t} - \left(\beta' \bar{\mathbf{x}}^s_{..t} + \gamma' \bar{\mathbf{s}}^s_{.t} + \delta' \bar{\mathbf{d}}^s_{.t} + \kappa' \mathbf{q}_t + \varphi' \bar{\mathbf{z}} + \bar{\mu} + \bar{v} + \bar{\zeta} + \bar{\varepsilon}_{..t} \right) \right\} + o_p(1) \ .$$

Hence, we augment the model (10.43) with the cross-sectional averages as

$$y^s_{ijt} = \beta' \mathbf{x}^s_{ijt} + \gamma' \mathbf{s}^s_{it} + \delta' \mathbf{d}^s_{jt} + \psi'_{ij} \mathbf{f}^s_t + \tau^s_{ij} + \mu^s_{ij} + \upsilon^s_{it} + \zeta^s_{jt} + \varepsilon^{*s}_{ijt}, \tag{10.46}$$

where $\varepsilon^{*s}_{ijt} = s_{ijt} \varepsilon^*_{ijt}$ with $\varepsilon^*_{ijt} = \varepsilon_{ijt} - \frac{\pi_{ij}}{\bar{\pi}} \left(\bar{\varepsilon}_{..t} + \bar{\mu} + \bar{v} + \bar{\zeta} \right) \to_p \varepsilon_{ijt}$ and see (10.39) for other definitions.

To derive the appropriate 3D-Within transformation directly for unbalanced panels (10.46), we consider the simple specification

$$y^s_{ijt} = \mu^s_{ij} + \upsilon^s_{it} + \xi^s_{jt} + \varepsilon^s_{ijt}, \tag{10.47}$$

and examine the property of the transformed data given by

$$\tilde{y}^s_{ijt} = y^s_{ijt} + s_{ijt} \left(-\bar{y}^s_{ij\cdot} - \bar{y}^s_{\cdot jt} - \bar{y}^s_{i\cdot t} + \bar{y}^s_{\cdot\cdot t} + \bar{y}^s_{\cdot j\cdot} + \bar{y}^s_{i\cdot\cdot} - \bar{y}^s_{\cdots} \right) \ . \tag{10.48}$$

Then, it is straightforward to show

$$\begin{aligned} &\left(-\bar{y}^s_{ij\cdot} - \bar{y}^s_{\cdot jt} - \bar{y}^s_{i\cdot t} + \bar{y}^s_{\cdot\cdot t} + \bar{y}^s_{\cdot j\cdot} + \bar{y}^s_{i\cdot\cdot} - \bar{y}^s_{\cdots} \right) \\ &= -\left(\mu_{ij} + \upsilon_{it} + \xi_{jt} \right) + D_1 + D_2 + D_3 + D_4 + D_5 \ , \end{aligned}$$

where

$$D_1 = -\left(\bar{v}^s_{ij\cdot} - \sum_{j=1}^{N_2} \frac{n_{ij}}{n_i} \bar{v}^s_{ij\cdot} \right) + \left(\sum_{i=1}^{N_1} \frac{n_{ij}}{n_j} \bar{v}^s_{ij\cdot} - \sum_{i=1}^{N_1} \frac{n_i}{n} \sum_{j=1}^{N_2} \frac{n_{ij}}{n_i} \bar{v}^s_{ij\cdot} \right)$$

$$D_2 = -\left(\bar{\xi}^s_{ij\cdot} - \sum_{i=1}^{N_1} \frac{n_{ij}}{n_j} \bar{\xi}^s_{ij\cdot} \right) + \left(\sum_{j=1}^{N_2} \frac{n_{ij}}{n_i} \bar{\xi}^s_{ij\cdot} - \sum_{j=1}^{N_2} \frac{n_j}{n} \sum_{i=1}^{N_1} \frac{n_{ij}}{n_j} \bar{\xi}^s_{ij\cdot} \right)$$

$$D_3 = -\left(\bar{\mu}_{\cdot jt} - \sum_{t=1}^{T} \frac{n_{jt}}{n_j} \bar{\mu}_{\cdot jt} \right) - \left(\bar{\mu}_{i\cdot t} - \sum_{t=1}^{T} \frac{n_{it}}{n_i} \bar{\mu}_{i\cdot t} \right) + \left(\bar{\mu}_{\cdot\cdot t} - \sum_{t=1}^{T} \frac{n_t}{n} \bar{\mu}_{\cdot\cdot t} \right)$$

$$D_4 = -\left(\bar{v}^s_{\cdot jt} - \sum_{j=1}^{N_2} \frac{n_{jt}}{n_t} \bar{v}^s_{\cdot jt} \right), \ D_5 = -\left(\bar{\xi}_{i\cdot t} - \sum_{i=1}^{N_1} \frac{n_{it}}{n_t} \bar{\xi}_{i\cdot t} \right)$$

with

$$\begin{aligned}
\bar{v}^s_{ij.} &= \tfrac{1}{n_{ij}} \textstyle\sum_{t=1}^{T} s_{ijt} \upsilon_{it} \\
\bar{\xi}^s_{ij.} &= \tfrac{1}{n_{ij}} \textstyle\sum_{t=1}^{T} s_{ijt} \xi_{jt} \\
\bar{\mu}^s_{.jt} &= \tfrac{1}{n_{jt}} \textstyle\sum_{i=1}^{N_1} s_{ijt} \mu_{ij} \\
\bar{\mu}^s_{i.t} &= \tfrac{1}{n_{it}} \textstyle\sum_{j=1}^{N_2} s_{ijt} \mu_{ij} \\
\bar{\mu}^s_{..t} &= \tfrac{1}{n_t} \textstyle\sum_{i=1}^{N_1} \sum_{j=1}^{N_2} s_{ijt} \mu_{ij} \\
\bar{v}^s_{.jt} &= \tfrac{1}{n_{jt}} \textstyle\sum_{i=1}^{N_1} s_{ijt} \upsilon_{it} \quad \text{and} \\
\bar{\xi}^s_{i.t} &= \tfrac{1}{n_{it}} \textstyle\sum_{j=1}^{N_2} \xi_{jt} \, .
\end{aligned}$$

Notice in the balanced panels that $D_1 = D_2 = D_3 = D_4 = D_5 = 0$. In addition, as $\left(\min_i n_i, \min_j n_j, \min_t n_t, \min_{(ij)} n_{ij}\right) \to \infty$, $D_i \to_p 0$ for $i = 1, ..., 5$. Therefore, we obtain

$$\left(-\bar{y}^s_{ij.} - \bar{y}^s_{.jt} - \bar{y}^s_{i\cdot t} + \bar{y}^s_{..t} + \bar{y}^s_{.j.} + \bar{y}^s_{i..} - \bar{y}^s_{...}\right) = -\left(\mu_{ij} + \upsilon_{it} + \xi_{jt}\right) + o_p(1) \, . \quad (10.49)$$

Using (10.49) and applying (10.48) to (10.47), we obtain the desired result

$$\tilde{y}^s_{ijt} = \tilde{\varepsilon}^s_{ijt}, \quad (10.50)$$

where $\tilde{\varepsilon}^s_{ijt} = \varepsilon^s_{ijt} - s_{ijt}\left(\bar{\varepsilon}^s_{ij.} - \bar{\varepsilon}^s_{.jt} - \bar{\varepsilon}^s_{i\cdot t} + \bar{\varepsilon}^s_{..t} + \bar{\varepsilon}^s_{.j.} + \bar{\varepsilon}^s_{i..} - \bar{\varepsilon}^s_{...}\right)$.

We now apply the 3D-Within transformation (10.48) to (10.46) and obtain

$$\tilde{y}^s_{ijt} = \beta' \tilde{\mathbf{x}}^s_{ijt} + \tilde{\psi}'_{ij} \mathring{\mathbf{f}}^s_{ijt} + \tilde{\varepsilon}^{*s}_{ijt}, \quad (10.51)$$

where $\tilde{\psi}'_{ij} = \psi'_{ij} - \left(\frac{1}{n_{jt}} \sum_{i=1}^{N_1} \psi'_{ij}\right) - \left(\frac{1}{n_{it}} \sum_{j=1}^{N_2} \psi'_{ij}\right) + \left(\frac{1}{n_t} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \psi'_{ij}\right)$, $\mathring{\mathbf{f}}^s_{ijt} = s_{ijt} \tilde{\mathbf{f}}^s_{ij}$ with $\tilde{\mathbf{f}}^s_{ij} = \mathbf{f}^s_t - \bar{\mathbf{f}}^s_{ij}$ and $\bar{\mathbf{f}}^s_{ij} = n_{ij}^{-1} \sum_{t=1}^{T} s_{ijt} \mathbf{f}^s_t$, and $\mathbf{f}^s_t$ is defined in (10.40). Collecting only the n_{ij} observations with $s_{ijt} = 1$ from (10.51), we have

$$\tilde{\mathbf{Y}}_{ij} = \tilde{\mathbf{X}}_{ij} \beta + \tilde{\mathbf{F}}_{ij} \tilde{\psi}_{ij} + \tilde{\mathbf{E}}^*_{ij} \, , \quad (10.52)$$

where

$$\underset{(n_{ij} \times 1)}{\tilde{\mathbf{Y}}_{ij}} = \begin{bmatrix} \tilde{y}_{ij(1)} \\ \vdots \\ \tilde{y}_{ij(n_{ij})} \end{bmatrix}, \; \underset{(n_{ij} \times k_x)}{\tilde{\mathbf{X}}_{ij}} = \begin{bmatrix} \tilde{\mathbf{x}}'_{ij(1)} \\ \vdots \\ \tilde{\mathbf{x}}'_{ij(n_{ij})} \end{bmatrix},$$

$$\underset{(n_{ij} \times k_f)}{\tilde{\mathbf{F}}_{ij}} = \begin{bmatrix} \tilde{\mathbf{f}}^{s\prime}_{ij(1)} \\ \vdots \\ \tilde{\mathbf{f}}^{s\prime}_{ij(n_{ij})} \end{bmatrix}, \; \underset{(n_{ij} \times 1)}{\tilde{\mathbf{E}}^*_{ij}} = \begin{bmatrix} \tilde{\varepsilon}^*_{ij(1)} \\ \vdots \\ \tilde{\varepsilon}^*_{ij(n_{ij})} \end{bmatrix}.$$

As before, we use the time index inside $(.)$ to highlight different initial and last time periods for each cross-sectional pair (ij) respectively. Then, the 3D-PCCE estimators of β are obtained by

$$\tilde{\beta}_{PCCE} = \left(\sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \tilde{\mathbf{X}}'_{ij} \mathbf{M}_{ij} \tilde{\mathbf{X}}_{ij} \right)^{-1} \left(\sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \tilde{\mathbf{X}}'_{ij} \mathbf{M}_{ij} \tilde{\mathbf{Y}}_{ij} \right), \tag{10.53}$$

where $\mathbf{M}_{ij} = \mathbf{I}_T - \tilde{\mathbf{F}}_{ij} \left(\tilde{\mathbf{F}}'_{ij} \tilde{\mathbf{F}}_{ij} \right)^{-1} \tilde{\mathbf{F}}'_{ij}$. As $\left(\min_t n_t, \min_{(ij)} n_{(ij)} \right) \to \infty$, both PCCE estimators, (10.42) and (10.53), follow the asymptotic normal distribution.

10.4.2 4D Model Extensions

BMW propose the following baseline 4D model

$$y_{ijst} = \mathbf{x}'_{ijst} \beta + u_{ijst}, \tag{10.54}$$

$$u_{ijst} = \mu_{ijs} + \theta_{ijt} + \zeta_{jst} + \upsilon_{ist} + \varepsilon_{ijst} \tag{10.55}$$

for $i = 1, ..., N_1$, $j = 1, ..., N_2$, $s = 1, ..., N_3$, and $t = 1, ..., T$. BMW derive the following 4D Within transformation to eliminate pair-wise interaction effects, μ_{ijs}, υ_{ist}, ζ_{jst}, and λ_{ijt} from (10.54)

$$\begin{aligned} \tilde{y}_{ijst} = {} & y_{ijst} - \bar{y}_{.jst} - \bar{y}_{i.st} - \bar{y}_{ij.t} - \bar{y}_{ijs.} + \bar{y}_{..st} + \bar{y}_{.j.t} + \bar{y}_{.js.} + \bar{y}_{i..t} + \bar{y}_{i.s.} + \bar{y}_{ij..} \\ & - \bar{y}_{...t} - \bar{y}_{..s.} - \bar{y}_{.j..} - \bar{y}_{i...} - \bar{y}_{....} \end{aligned} \tag{10.56}$$

and estimate β consistently from the transformed specification

$$\tilde{y}_{ijst} = \tilde{\mathbf{x}}'_{ijst} \beta + u_{ijst}. \tag{10.57}$$

Further, BBMP propose the feasible GLS random effects estimator of β under the assumption that u_{ijst} and its components individually have a zero mean, and the error components are pairwise uncorrelated.

To introduce CSD explicitly into (10.54), we consider the following extension

$$y_{ijst} = \mathbf{x}'_{ijst} \beta + \mu_{ijs} + \theta_{ijt} + \zeta_{jst} + \upsilon_{ist} + \pi_{ijs} \lambda_t + \varepsilon_{ijst}, \tag{10.58}$$

where λ_t is the global factor with heterogeneous coefficients π_{ijs}. In the presence of such CSD, both the 4D-FE estimator and the 4D-RE estimator would likely be biased due to the correlation between $\mathbf{x}_{ijst}$ and λ_t. Thus, we develop the two-step consistent estimation procedure. Taking the cross-sectional averages of (10.58) over i, j and s, we obtain

$$\begin{aligned} \bar{y}_{...t} = {} & \frac{1}{N_1} \sum_{i=1}^{N_1} \frac{1}{N_2} \sum_{j=1}^{N_2} \frac{1}{N_3} \sum_{s=1}^{N_3} \left(\beta' \mathbf{x}_{ijst} + \mu_{ijs} + \theta_{ijt} + \zeta_{jst} + \upsilon_{ist} + \pi_{ijs} \lambda_t + \varepsilon_{ijst} \right) \\ = {} & \beta' \bar{\mathbf{x}}_{...t} + \bar{\mu}_{...} + \bar{\theta}_{..t} + \bar{\zeta}_{..t} + \bar{\upsilon}_{..t} + \bar{\pi}_{...} \lambda_t + \bar{\varepsilon}_{...t}, \end{aligned} \tag{10.59}$$

where

$$\begin{aligned}
\bar{\mathbf{x}}_{...t} &= \frac{1}{N_1}\sum_{i=1}^{N_1}\frac{1}{N_2}\sum_{j=1}^{N_2}\frac{1}{N_3}\sum_{s=1}^{N_3}\mathbf{x}_{ijst} \\
\bar{\varepsilon}_{...t} &= \frac{1}{N_1}\sum_{i=1}^{N_1}\frac{1}{N_2}\sum_{j=1}^{N_2}\frac{1}{N_3}\sum_{s=1}^{N_3}\varepsilon_{ijst} \\
\bar{\mu}_{...} &= \frac{1}{N_1}\sum_{i=1}^{N_1}\frac{1}{N_2}\sum_{j=1}^{N_2}\frac{1}{N_3}\sum_{s=1}^{N_3}\mu_{ijs} \\
\bar{\pi}_{...} &= \frac{1}{N_1}\sum_{i=1}^{N_1}\frac{1}{N_2}\sum_{j=1}^{N_2}\frac{1}{N_3}\sum_{s=1}^{N_3}\pi_{ijs} \\
\bar{\theta}_{..t} &= \frac{1}{N_1}\sum_{i=1}^{N_1}\frac{1}{N_2}\sum_{j=1}^{N_2}\theta_{ijt} \\
\bar{\zeta}_{..t} &= \frac{1}{N_2}\sum_{j=1}^{N_2}\frac{1}{N_3}\sum_{s=1}^{N_3}\zeta_{jst} \quad \text{and} \\
\bar{\upsilon}_{..t} &= \frac{1}{N_1}\sum_{i=1}^{N_1}\frac{1}{N_3}\sum_{s=1}^{N_3}\upsilon_{ist}\ .
\end{aligned}$$

From (10.59), we have

$$\lambda_t = \frac{1}{\bar{\pi}_{...}}\left\{\bar{y}_{...t} - \left(\beta'\bar{\mathbf{x}}_{...t} + \bar{\mu}_{...} + \bar{\theta}_{..t} + \bar{\zeta}_{..t} + \bar{\upsilon}_{..t} + \bar{\pi}_{...}\lambda_t + \bar{\varepsilon}_{...t}\right)\right\}.$$

Thus, we derive the cross-sectional augmented version of (10.58) by

$$y_{ijst} = \beta'\mathbf{x}_{ijst} + \psi'_{ijs}\mathbf{f}_t + \mu_{ijs} + \theta_{ijt} + \zeta_{jst} + \upsilon_{ist} + \varepsilon^*_{ijst}\ , \tag{10.60}$$

where $\mathbf{f}_t = (\bar{y}_{...t}, \bar{\mathbf{x}}'_{...t})'$, $\psi'_{ijs} = \left(\psi_{0,ijs}, \psi'_{ijs}\right) = \left(\frac{\pi_{ijs}}{\bar{\pi}_{...}}, -\frac{\pi_{ijs}}{\bar{\pi}_{...}}\beta'\right)$ and $\varepsilon^*_{ijst} = \varepsilon_{ijst} - \frac{\pi_{ijs}}{\bar{\pi}_{...}}\left(\bar{\varepsilon}_{...t} + \bar{\theta}_{..t} + \bar{\zeta}_{..t} + \bar{\upsilon}_{..t}\right)$. As $N_1, N_2, N_3 \to \infty$, $\varepsilon^*_{ijst} \to_p \varepsilon_{ijst}$ because of the following approximations: $\bar{\mu}_{...} \to_p 0$, $\theta_{..t} \to_p 0$, $\zeta_{..t} \to_p 0$, $\upsilon_{..t} \to_p 0$, and $\bar{\varepsilon}_{...t} \to_p 0$.

Next, we apply the 4D-Within transformation (10.56) to (10.60), and obtain

$$\tilde{y}_{ijst} = \beta'\tilde{\mathbf{x}}_{ijst} + \tilde{\psi}'_{ijs}\tilde{\mathbf{f}}_t + \tilde{\varepsilon}^*_{ijt}\ , \tag{10.61}$$

where

$$\tilde{\psi}'_{ijs} = \left(\psi_{ijs} - \psi_{.js} - \psi_{i.s} - \psi_{ij.} + \psi_{..s} + \psi_{.j.} + \psi_{i..} - \psi_{...}\right)' \text{ and } \tilde{\mathbf{f}}_t = \left(\mathbf{f}_t - \bar{\mathbf{f}}\right).$$

We rewrite (10.61) as

$$\tilde{\mathbf{Y}}_{ijs} = \tilde{\mathbf{X}}_{ijs}\beta + \tilde{\mathbf{F}}\tilde{\psi}_{ijs} + \tilde{\varepsilon}^*_{ijs}, \tag{10.62}$$

where

$$\underset{(T\times 1)}{\tilde{\mathbf{Y}}_{ijs}} = \begin{bmatrix} \tilde{y}_{ijs1} \\ \vdots \\ \tilde{y}_{ijsT} \end{bmatrix},\ \underset{(T\times k_x)}{\tilde{\mathbf{X}}_{ijs}} = \begin{bmatrix} \tilde{\mathbf{x}}'_{ijs1} \\ \vdots \\ \tilde{\mathbf{x}}'_{ijsT} \end{bmatrix},\ \underset{(T\times k_f)}{\tilde{\mathbf{F}}} = \begin{bmatrix} \tilde{\mathbf{f}}'_1 \\ \vdots \\ \tilde{\mathbf{f}}'_T \end{bmatrix}.$$

Then, it is straightforward to derive the PCCE estimator of β by

$$\hat{\beta}_{PCCE} = \left(\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\sum_{s=1}^{N_3}\tilde{\mathbf{X}}'_{ijs}\mathbf{M}_{\tilde{\mathbf{F}}}\tilde{\mathbf{X}}_{ijs}\right)^{-1}\left(\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\sum_{s=1}^{N_3}\tilde{\mathbf{X}}'_{ijs}\mathbf{M}_{\tilde{\mathbf{F}}}\tilde{\mathbf{Y}}_{ijs}\right), \tag{10.63}$$

where $\mathbf{M}_{\tilde{\mathbf{F}}} = \mathbf{I}_T - \tilde{\mathbf{F}}\left(\tilde{\mathbf{F}}'\tilde{\mathbf{F}}\right)^{-1}\tilde{\mathbf{F}}'$.

Further, we can follow Kapetanios and Shin (2017) and develop 4D models with the hierarchical multi-factor error structure. To this end define the global factor λ_t

which affects all (ijs) pairs, the regional factors τ_{it}, τ^{*}_{jt}, τ^{**}_{st}, and finally the local factors τ_{ijt}, τ^{*}_{ist} and τ^{**}_{jst}. This logic suggests the following model

$$y_{ijst} = \mathbf{x}'_{ijst}\beta + \mu_{ijs} + \upsilon_{js}\tau_{it} + \upsilon^{*}_{is}\tau_{jt} + \upsilon^{**}_{ij}\tau_{st} + \zeta_{s}\tau_{ijt} + \zeta^{*}_{j}\tau^{*}_{ist} + \zeta^{**}_{i}\tau^{**}_{jst} + \pi_{ij}\lambda_{t} + \varepsilon_{ijst}. \tag{10.64}$$

Such higher-dimensional setups involve several layers of factor specifications (a number that grows with the dimensions), rendering estimation and inference non-trivial and challenging.

10.5 Monte Carlo Analysis

In this section, we conduct Monte Carlo studies and investigate the small sample properties of the CTFE and two versions of the PCCE estimator for models, (10.4), (10.11) and (10.17), respectively. We consider the two data generating processes (DGP). We construct DGP1 by

$$y_{ijt} = \beta x_{ijt} + \mu_{ij} + \pi_{ij}\lambda_{t} + \varepsilon_{ijt}, \tag{10.65}$$

$$x_{ijt} = \mu^{x}_{ij} + \mu_{ij} + \pi^{x}_{ij}\lambda_{t} + \upsilon_{ijt}, \tag{10.66}$$

for $i = 1,...,N_1$, $j = 1,...,N_2$, and $t = 1,...,T$. The global common factor, λ_t and idiosyncratic errors, ε_{ijt} and υ_{ijt} are generated independently as *iid* processes with zero mean and unit variance: namely, $\lambda_t \sim$ iid $N(0,1)$, $\varepsilon_{ijt} \sim$ iid $N(0,1)$ and $\upsilon_{ijt} \sim$ iid $N(0,1)$. We generate pairwise individual effects independently as $\mu_{ij} \sim$ iid $N(0,1)$ and $\mu^{x}_{ij} \sim$ iid $N(0,1)$. The factor loadings, π_{ij} and π^{x}_{ij}, are then independently generated from $U[1,2]$ random variables.

Next, we construct DGP2 by

$$y_{ijt} = \beta x_{ijt} + \mu_{ij} + \upsilon_{it} + \zeta_{jt} + \pi_{ij}\lambda_{t} + \varepsilon_{ijt}\ , \tag{10.67}$$

$$x_{ijt} = \mu^{x}_{ij} + \mu_{ij} + \pi^{x}_{ij}\lambda_{t} + \upsilon_{ijt}, \tag{10.68}$$

for $i = 1,...,N_1$, $j = 1,...,N_2$, and $t = 1,...,T$. In addition, we follow BBMP and generate υ_{it} and ζ_{jt} independently as

$$\upsilon_{it} \sim U(-1,1) \text{ and } \zeta_{jt} \sim U(-1,1) \text{ for } i = 1,...,N_1,\ j = 1,...,N_2,\ t = 1,...T.$$

In both DGP1 and DGP2, we set $\beta = 1$.

For each experiment, we evaluate the following summary statistics:

- Bias: $\hat{\beta}_R - \beta_0$, where $\beta_0\,(=1)$ is the true parameter value, and $\hat{\beta}_R = R^{-1}\sum_{r=1}^{R}\hat{\beta}_r$ is the mean of $\hat{\beta}_r$.
- Root mean square error (RMSE):

$$\sqrt{R^{-1}\sum_{r=1}^{R}\left(\hat{\beta}_r-\beta_0\right)^2}.$$

- Size: the empirical rejection probability of the t-statistic for the null hypothesis $\beta=\beta_0$ against $\beta\neq\beta_0$ at the 5% significance.

We conduct each experiment 1,000 times for the (N_1,N_2,T) triples with $N_1,N_2=\{25,49,100\}$[11] and $T=\{50,100,200,400\}$. The simulation results are provided in Table 10.1 for DGP1 and Table 10.2 for DGP2 respectively.

In Table 10.1 we find that biases of the 2D and 3D PCCE estimators of β are mostly negligible even for the relatively small sample size at $(N_1,N_2,T)=(25,25,50)$. On the other hand, the CTFE estimator displays substantial biases for most cases. As N_1, N_2 or T increases, the biases become smaller but still non-negligible. RMSE results are qualitatively similar to the bias patterns. RMSEs of both PCCE estimators decrease sharply with $N_1\,(N_2)$ or T, whilst RMSEs of the CTFE estimator fall only with $N_1\,(N_2)$. Turning to the empirical sizes of the t-test for the null hypothesis, $\hat{\beta}=\beta_0\,(=1)$, we find that the CTFE over-rejects the null in all cases and tends to 1 as $N_1\,(N_2)$ or T rises. By contrast, the size of the 2D PCCE estimator is reasonably close to the nominal 5% level in all cases, while the 3D PCCE estimator tends to slightly over-reject the null when N_1 or N_2 is relatively small. As expected, the overall performance of the 2D PCCE estimator is the best under DGP1.

Simulation results in Table 10.2 are qualitatively similar to those in Table 10.1. Biases of both PCCE estimators are almost negligible in all cases, and their RMSEs decrease rapidly with $N_1\,(N_2)$ or T. Empirical sizes of the t-test for $\hat{\beta}=\beta_0\,(=1)$ are still close to the nominal 5% level in almost all cases for the 2D PCCE estimator. The 3D PCCE estimator tends to slightly over-reject the null, but its size performance improves quickly as $N_1\,(N_2)$ or T increases. In contrast, the CTFE estimator suffers from substantial biases and size distortions, and its performance does not improve in large samples. We note in passing that such good performance of the 2D PCCE estimator is rather surprising as we expect that the 3D PCCE estimator will dominate under DGP2. Overall simulation results support the simulation findings reported under the 2D panels by Kapetanios and Pesaran (2005) and Pesaran (2006).

10.6 Empirical Application to the Gravity Model of the Intra-EU Trade

Anderson and van Wincoop (2003) show that "the gravity equation tells us that bilateral trade, after controlling for size, depends on the bilateral trade barriers but relative to the product of their Multilateral Resistance Indices (MTR)." Bilateral barriers are relative to average trade barriers that both regions face with all their trading partners. Omitting MTR induces a potentially severe bias (see, e.g., Baldwin

[11] Namely, 25 are pairs of 5 units, 49 pairs of 7 units, and 100 pairs of 10 units.

Table 10.1 Simulation results for β under the DGP1

	CTFE							
	Bias							
(N_1N_2, T)	50	100	200	400				
25	0.0829	0.0832	0.0833	0.0822				
49	0.0347	0.0341	0.0338	0.0344				
100	-0.0307	-0.0315	-0.0313	-0.0316				
	RMSE							
(N_1N_2, T)	50	100	200	400				
25	0.0914	0.0871	0.0854	0.0832				
49	0.0420	0.0383	0.0357	0.0353				
100	0.0347	0.0336	0.0324	0.0322				
			Size					
(N_1N_2, T)	50	100	200	400				
25	0.7610	0.9590	0.9980	1.0000				
49	0.4020	0.6290	0.8810	0.9950				
100	0.5530	0.8410	0.9910	1.0000				
	2D PCCE				3D PCCE			
	Bias				Bias			
(N_1N_2, T)	50	100	200	400	50	100	200	400
25	0.0017	0.0011	0.0014	0.0003	0.0017	0.0008	0.0012	0.0002
49	0.0006	-0.0005	0.0002	0.0004	0.0000	-0.0001	0.0000	0.0006
100	0.0004	-0.0005	-0.0001	-0.0003	0.0009	-0.0002	0.0000	-0.0003
	RMSE				RMSE			
(N_1N_2, T)	50	100	200	400	50	100	200	400
25	0.0290	0.0202	0.0146	0.0101	0.0290	0.0202	0.0146	0.0101
49	0.0207	0.0156	0.0100	0.0071	0.0207	0.0156	0.0100	0.0071
100	0.0142	0.0103	0.0070	0.0051	0.0142	0.0103	0.0070	0.0051
	Size				Size			
(N_1N_2, T)	50	100	200	400	50	100	200	400
25	0.049	0.045	0.052	0.052	0.133	0.124	0.132	0.114
49	0.041	0.062	0.047	0.047	0.095	0.113	0.097	0.086
100	0.042	0.048	0.044	0.055	0.081	0.093	0.074	0.093

Notes: We report the simulation results for three estimators for DGP1, (10.65) and (10.66). CTFE refers to the 3D Within estimator given by (10.5), 2D PCCE is the PCCE estimator given by (10.11), and 3D PCCE is the PCCE estimator given by (10.18).

and Taglioni, 2006). Subsequent research has focused on estimating the model with directional country-specific fixed effects to control for unobservable MTRs (see, e.g., Feenstra, 2004).

A large number of studies have established the importance of taking into account multilateral resistance and bilateral heterogeneity, simultaneously, in the 2D panels. Serlenga and Shin (2007) is the first paper to develop the panel gravity model by incorporating observed and unobserved factors. Alternatively, Behrens

Table 10.2 Simulation results for β under the DGP2

	CTFE							
	Bias							
(N_1N_2, T)	50	100	200	400				
25	0.0835	0.0829	0.0830	0.0827				
49	0.0143	0.0144	0.0155	0.0156				
100	-0.0365	-0.0371	-0.0362	-0.0370				
	RMSE							
(N_1N_2, T)	50	100	200	400				
25	0.0921	0.0872	0.0850	0.0839				
49	0.0272	0.0220	0.0194	0.0177				
100	0.0400	0.0388	0.0371	0.0374				
			Size					
(N_1N_2, T)	50	100	200	400				
25	0.7780	0.9450	1.0000	1.0000				
49	0.1420	0.2060	0.3630	0.5650				
100	0.7120	0.9400	0.9940	1.0000				
	2D PCCE				3D PCCE			
	Bias				Bias			
(N_1N_2, T)	50	100	200	400	50	100	200	400
25	-0.0001	0.0008	0.0009	0.0001	0.0012	0.0006	0.0009	0.0005
49	-0.0002	0.0000	0.0006	0.0001	-0.0001	0.0000	0.0005	0.0005
100	0.0000	-0.0001	0.0001	-0.0002	0.0001	-0.0003	0.0001	-0.0002
	RMSE				RMSE			
(N_1N_2, T)	50	100	200	400	50	100	200	400
25	0.0295	0.0201	0.0145	0.0104	0.0368	0.0250	0.0181	0.0130
49	0.0208	0.0147	0.0102	0.0072	0.0238	0.0169	0.0120	0.0083
100	0.0148	0.0103	0.0069	0.0049	0.0161	0.0114	0.0076	0.0054
	Size				Size			
(N_1N_2, T)	50	100	200	400	50	100	200	400
25	0.0510	0.0470	0.0630	0.0570	0.1290	0.1240	0.1330	0.1260
49	0.0480	0.0540	0.0460	0.0500	0.0910	0.1010	0.1070	0.0850
100	0.0620	0.0510	0.0450	0.0490	0.0930	0.0800	0.0660	0.0700

Notes: We report the simulation results for three estimators for DGP2, (10.67) and (10.68). CTFE refers to the 3D Within estimator given by (10.5), 2D PCCE is the PCCE estimator given by (10.11), and 3D PCCE is the PCCE estimator given by (10.18).

et al. (2012) develop the spatial econometric specification to control for multilateral cross-sectional correlations across trade flows. Mastromarco et al. (2016b) compare the factor and the spatial-based gravity models to investigate the Euro impact on intra-EU trade flows over 1960–2008 for 190 country-pairs of 14 EU and 6 non-EU OECD countries. They document evidence that the CD test confirms that the factor-based model is more appropriate. Furthermore, Gunnella et al. (2015) propose the panel gravity models which accommodate both strong and weak CSD, simultane-

ously, through the use of unobserved factors and endogenously selected spatial clusters estimated by the nonlinear threshold techniques advanced by Kapetanios et al. (2014).

When we analyse 3D panel gravity models, we should control for the potential origin of biases presented by unobserved time-varying MTRs, if they are correlated with covariates. Baltagi et al. (2003) propose the 3D panel data model with CTFE specification. This approach has been popularly applied to measure the impacts of (unobserved) MTRs of the exporters and the importers in the structural gravity studies (see, e.g., Chap. 11). As discussed in Sect. 10.3, however, the 3D panel data model, typically estimated by CTFE or CTRE estimators, fails to accommodate (strong and heterogeneous) CSD. The presence of cross-sectional correlations across (ij) pairs suggests that appropriate econometric techniques need to be used, in order to avoid biased and misleading estimation results.

In this regard, we apply our proposed approach to the dataset covering the period 1960–2008 (49 years) for 182 country-pairs amongst 14 EU member countries (Austria, Belgium-Luxemburg, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Netherlands, Portugal, Spain, Sweden, and United Kingdom).[12] Our sample period consists of several important economic integrations, such as the Customs Union in 1958, the European Monetary System in 1979 and the Single Market in 1993, all of which can be regarded as promoting intra-EU trade.[13]

We consider the following generalised panel gravity specification:

$$\begin{aligned}\ln \mathrm{EXP}_{ijt} = {} & \beta_0 + \beta_1 \mathrm{CEE}_{ijt} + \beta_2 \mathrm{EMU}_{ijt} + \beta_3 \mathrm{SIM}_{ijt} + \beta_4 \mathrm{RLF}_{ijt} + \beta_5 \ln \mathrm{GDP}_{it} \\ & + \beta_6 \ln \mathrm{GDP}_{jt} + \beta_7 \mathrm{RER}_t + \gamma_1 \mathrm{DIS}_{ij} + \gamma_2 \mathrm{BOR}_{ij} + \gamma_3 \mathrm{LAN}_{ij} + u_{ijt}\,, \end{aligned} \tag{10.69}$$

where the dependent variable, EXP_{ijt} is the export flow from country i to country j at time t, CEE and EMU are dummies for European Community membership and for European Monetary Union, SIM and RLF measure similarity in size and difference in relative factor endowments, RER represents the logarithm of common real exchange rates, GDP_{it} and GDP_{jt} are logged GDPs of exporter and importer, and finally the logarithm of geographical distance (DIS) and the dummies for common language (LAN) and for common border (BOR) represent time-invariant bilateral barriers.

We report the estimation results of (10.69), employing the four estimators, namely the two-way Within estimator with $u_{ijt} = \mu_{ij} + \lambda_t + \varepsilon_{ijt}$, the CTFE estimator with $u_{ijt} = \mu_{ij} + \upsilon_{it} + \zeta_{jt} + \varepsilon_{ijt}$, the 2D PCCE estimator with $u_{ijt} = \mu_{ij} + \pi_{ij}\lambda_t + \varepsilon_{ijt}$, and the 3D PCCE estimator with $u_{ijt} = \mu_{ij} + \upsilon_{it} + \zeta_{jt} + \pi_{ij}\lambda_t + \varepsilon_{ijt}$. We also report the CD test results applied to the residuals from each of the four estimation methods

[12] It is the extended dataset analysed by Serlenga and Shin (2007), who provide detailed data description and sources in the Data Appendix. Belgium and Luxemburg are treated as a single country. Denmark, Sweden and The UK although non euro member countries, as part of the EU, experienced a similar history and faced similar legislation and regulation to euro area countries.

[13] To mitigate the potentially negative impact of the global financial crisis on our analysis, we exclude the data after 2008.

and the estimates of the CSD exponent (α). Following BKP, we compute the CD test and estimate α sequentially.

Following the structural gravity literature, our main focus is on investigating the impacts of t_{ij}, which contain both barriers and incentives to trade between i and j.[14] Here we focus on the two dummy variables, CEE (equal to one when both countries belong to the European Community) and EMU (equal to one when both trading partners adopt the same currency). Both are expected to exert a positive impact on bilateral export flows. The main motivation behind the EMU project is that a single currency will reduce the transaction costs within member countries. However, the empirical evidence on the common currency effect on trade flows is rather mixed. Rose (2001), Frankel and Rose (2002), Glick and Rose (2002), and Frankel (2008) document a huge positive effect whilst a number of studies report negative or insignificant effects (see, e.g., Persson, 2001; Pakko and Wall, 2002; de Nardis and Vicarelli, 2003). More recent studies by Serlenga and Shin (2007), Mastromarco et al. (2016b), and Gunnella et al. (2015) highlight the importance of controlling for strong CSD, and find a small but significant effect (7 to 10%) of the euro on intra-EU trade flows.[15]

In addition to the standard mass covariates, GDP_{it} and GDP_{jt}, we also consider the impact of (the logarithm of) bilateral real exchange rates (RER), which is defined as the price of the foreign currency per the home currency unit and is meant to capture the relative price effects. Further, following the New Trade Theory advanced by Krugman (1997) and Helpman (1987), we add RLF and SIM. RLF is the logarithm of the absolute value of the difference between per capita GDPs of trading countries, and measures the difference in terms of relative factor endowments. The higher RLF results in a higher volume of inter-industry trade and a lower share of intra-industry trade. SIM is the logarithm of an index that captures the relative size of two countries in terms of GDP, and is bounded between zero (absolute divergence in size) and 0.5 (equal country size).

Table 10.3 reports the estimation results for the 3D panel gravity specification in (10.69). The two-way FE estimation results are all statistically significant except RER. The impacts of home and foreign country GDPs on exports are positive, but surprisingly, the former is twice as large as the latter. The impact of similarity in size (SIM) is negative and significant, inconsistent with *a priori* expectations. Importantly, we find that trade and currency union memberships (CEE and EMU) significantly boost export flows, though their magnitudes seem to be too high. However, the CD test applied to the residuals convincingly rejects the null of no or weak CSD. The estimate of α is 0.99 with the confidence band containing unity, suggest-

[14] In the current study, we cannot consistently estimate the coefficients associated with time-invariant regressors, DIS_{ij}, BOR_{ij}, and LAN_{ij} and/or because the Within transformation wipes them out. Similar identification issues are also applied to the coefficients on GDP_{it} and GDP_{jt}. Following Serlenga and Shin (2007) and Chap. 3, we will investigate this issue in a future study.

[15] After Brexit, the issue on potential benefits of joining the currency union will be more hotly debated. In retrospect, the UK Treasury made a bold prediction in 2003 that the pro-trade effect of the Euro on the UK would be over 40%.

ing that the residuals are strongly correlated. Thus, the FE estimation results are likely to be biased and unreliable.

Next, we turn to the CTFE estimation results, which have been popularly applied in the structural gravity literature. This approach aims to control for MTRs through bilateral pair-fixed effects and origin and destination country-time fixed effects. The CD test results indicate that the CTFE residuals do not suffer from any strong CSD, suggesting that the 3D Within transformation may be able to remove strong CSD. This rather surprising result is not supported by the estimate of α, which is 0.91. Though the confidence band does not include unity, this estimate is still quite high and close to 1. Further, we find that all the coefficients become insignificant except for CEE. Focusing on the impacts on CEE and EMU, the former is still substantial (0.29), while the latter turns out to be negligible (-0.011). Combining these results together, we may conclude that the CTFE results are rather unreliable.

Moving to the 2D PCCE estimation results, we find that all the coefficients are significant with the expected signs, except for EMU. The impact of foreign country GDP on exports is substantially larger than home GDP. The RER coefficient is positive, confirming that a depreciation of the home currency increases exports. The impact of CEE is smaller (0.187), but the EMU effect is insignificant and negligible (0.018). But, the PCCE estimator still suffers from strong CSD residuals together with the estimate of α being 0.87. This may explain the conflicting findings relative to the existing studies reporting a significant effect of the euro on trade in 2D panels.

Finally, the 3D PCCE estimation results show that all the coefficients – except RFL – are significant with expected signs. The CD test fails to strongly reject the null, suggesting that this approach is able to successfully deal with strong and/or weak CSD in the 3D panels. This is also supported by the smaller estimate of α (0.77), which is close to a moderate range of weak CSD.[16] Focusing on the CEE and EMU impacts on exports, we find that the former turns out to be still substantial (0.335), while the latter becomes modest at 0.081, close to the consensus magnitudes reported in the recent 2D panel studies (see, e.g., Baldwin, 2006; Gunnella et al., 2015). Combining these results, we may conclude that the 3D PCCE estimation results are mostly reliable, providing a general support for the thesis that the potential trade-creating effects of the Euro should be viewed in the long-run historical and multilateral perspectives rather than simply focusing on the formation of a monetary union as an isolated event.

The CTFE estimator is proposed to capture the (unobserved) multilateral resistance terms and trade costs, which are likely to exhibit history and time dependence (see, e.g., Herwartz and Weber, 2015). However, it fails to accommodate strong cross-sectional correlations among MTRs, which are present in our sample of the EU countries (confirmed by CD tests and CSD exponent estimates). To capture these complex interlinkages among trading partners, we should model the time-varying interdependency of bilateral export flows in a more flexible manner than simply introducing deterministic country-time specific dummies. Baldwin (2006) stresses the importance of taking into account the fact that many omitted pair-specific variables

[16] BHP show that the values of $\alpha \in (1/2, 3/4)$ represent a moderate degree of CSD.

Table 10.3 3D panel gravity model estimation results for bilateral export flows

	FE			CTFE		
	Coeff	se	t-ratio	Coeff	se	t-ratio
gdph	2.185	0.041	52.97			
gdpf	1.196	0.041	28.98			
sim	-0.263	0.052	-5.069	-0.055	0.074	-0.754
rlf	0.031	0.006	5.011	0.006	0.005	1.294
rer	0.005	0.007	0.791	0.031	0.072	0.436
cee	0.302	0.014	22.05	0.290	0.017	16.99
emu	0.204	0.019	10.71	-0.011	0.036	-0.315
CD stat	620.1			-2.676		
	$\alpha_{0.05}$	α	$\alpha_{0.95}$	$\alpha_{0.05}$	α	$\alpha_{0.95}$
CSD exponent	0.925	0.992	1.059	0.865	0.914	0.963
	2D PCCE			3D PCCE		
	Coeff	se	t-ratio	Coeff	se	t-ratio
gdph	0.289	0.095	3.033			
gdpf	1.491	0.095	15.69			
sim	0.042	0.105	0.401	1.032	0.111	9.290
rlf	0.007	0.005	1.420	-0.004	0.005	-0.748
rer	0.144	0.019	7.427	0.168	0.114	1.471
cee	0.187	0.014	13.20	0.335	0.022	15.10
emu	0.018	0.015	1.160	0.081	0.045	1.793
CD stat	76.11			-4.19		
	$\alpha_{0.05}$	α	$\alpha_{0.95}$	$\alpha_{0.05}$	α	$\alpha_{0.95}$
CSD exponent	0.837	0.867	0.897	0.724	0.775	0.826

Notes: Using the annual dataset over 1960-2008 for 182 country-pairs amongst 14 EU member countries, we estimate the 3D panel gravity specification, (10.69). FE stands for the two-way fixed effects estimator with country-pair and time effects. CTFE refers to the 3D Within estimator given by (10.5). 2D PCCE estimator is given by (10.11) with factors $\mathbf{f}_t = \left\{\overline{gdp}_{.t}, \overline{sim}_{..t}, \overline{rlf}_{..t}, \overline{cee}_{..t}, \overline{rer}_t, t\right\}$. 3D PCCE estimator is given by (10.18) with factors $\mathbf{f}_t = \left\{\overline{sim}_{..t}, \overline{rlf}_{..t}, \overline{rer}_t, t\right\}$. CD test refers to testing the null hypothesis of residual cross-sectional error independence or weak dependence and is defined in (10.34). α denotes the estimate of CSD exponent jointly with the 90% confidence bands.

reflect time-varying factors, such as multilateral trade costs or union membership. MTRs arise from the bilateral country-pair specific reactions to global shocks or the local spillover effects across a small number of countries or both. In order to avoid biased and misleading results, we propose a novel econometric technique, called the 3D-PCCE estimator, which is the first step to developing the multi-dimensional models with the hierarchical multi-factor error structure, whereby an external shock can alter the trade costs for an individual country relative to all other countries in a heterogeneous and time-dependent way (sse, e.g., Kapetanios and Shin, 2017).

10.7 Conclusion

Given the growing availability of big datasets containing information on multiple dimensions, the recent literature on panel data has focused on extending the two-way error components models to the multi-dimensional setting. We propose in this chapter novel estimation techniques to accommodate cross-sectional error dependence within the 3D panel data models. Despite the massive development of modelling residual CSD through unobserved factors in 2D panel data models (see, e.g., Pesaran, 2006; Bai, 2009), our approach is the first attempt to introduce strong CSD into multi-dimensional error components models, and well suited to the analysis of sophisticated error CSD across triple or higher dimensions.

We develop a two-step consistent estimation procedure, called the 3D-PCCE estimator. We discuss the extent of cross-sectional dependence and develop a diagnostic test for the null hypothesis of (pairwise) residual cross-sectional independence or weak dependence in the 3D panels. The empirical usefulness and superiority of the proposed 3D-PCCE estimator are demonstrated via some Monte Carlo studies and an empirical application to the 3D panel gravity model of intra-EU trade.

At this stage, it seems appropriate to mention the number of obvious and challenging extensions and generalisations. First, as discussed in Sect. 10.4, we address the number of extensions to the analysis of incomplete panel datasets and 4D or higher dimensional models. Second, as an ongoing research, we develop general multi-dimensional heterogenous panel data models with hierarchical multi-factor error structure (see, e.g., Kapetanios and Shin, 2017). Third, our proposed approach can also easily be extended to dynamic models. Finally and more importantly, we aim to develop the most challenging models by combining both the spatial-based and factor-based techniques within the 3D or higher dimensional models. Bailey et al. (2016a) develop a multi-step estimation procedure that can distinguish the relationship between spatial units that are purely spatial from those which are due to the effect of common factors. Furthermore, Mastromarco et al. (2016a) propose a technique for allowing weak and strong CSD in modelling technical efficiency of stochastic frontier panels by combining the exogenously driven factor-based approach and an endogenous threshold regime selection by Kapetanios et al. (2014). Bai and Li (2015) and Shi and Lee (2017) have developed the framework for jointly modelling spatial effects and interactive effects (see also Gunnella et al. (2015) and Kuersteiner and Prucha (2015)). This is the most recent research trend, and thus the successful development of a general combined approach within multi-dimensional panels may broaden its appeal further.

References

Abowd, J. M., Kramarz, Francis, and Margolis, D. N. (1999). High wage workers and high wage firms. *Econometrica*, 67(2):251–334.

Anderson, J. E. and van Wincoop, E. (2003). Gravity with gravitas: A solution to the border puzzle. *American Economic Review*, 93(1):170–192.

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica*, 77(4):1229–1279.

Bai, J. and Li, K. (2015). Dynamic spatial panel data models with common shocks. *Columbia University*.

Bai, J., Liao, Y., and Yang, J. (2015). Unbalanced panel data models with interactive effects. In Baltagi, B. H., editor, *The Oxford Handbook of Panel Data*, pages 149–170. Oxford University Press: New York.

Baier, S. L. and Bergstrand, J. H. (2007). Do free trade agreements actually increase members' international trade? *Journal of International Economics*, 71(1):72–95.

Bailey, N., Holly, S., and Pesaran, M. H. (2016a). A two-stage approach to spatio-temporal analysis with strong and weak cross-sectional dependence. *Journal of Applied Econometrics*, 31:249–280.

Bailey, N., Kapetanios, G., and Pesaran, M. H. (2016b). Exponent of cross-sectional dependence: Estimation and inference. *Journal of Applied Econometrics*, 31(6):929–960.

Baldwin, R. (2006). In or out: Does it matter? an evidence-based analysis of the euro's trade effects. *Centre for Economic Policy Research*.

Baldwin, R. and Taglioni, D. (2006). Gravity for dummies and dummies for gravity equations. NBER Working Papers 12516, National Bureau of Economic Research, Inc.

Baltagi, B., Egger, P., and Pfaffermayr, M. (2003). A generalized design for bilateral trade flow models. *Economics Letters*, 80(3):391–397.

Baltagi, B., P. Egger, P., and Pfaffermayr, M. (2015). Panel data gravity models of international trade. In Baltagi, B. H., editor, *The Oxford Handbook of Panel Data*, pages 608–641. Oxford University Press: New York.

Behrens, K., Ertur, C., and Kock, W. (2012). Dual gravity: Using spatial econometrics to control for multilateral resistance. *Journal of Applied Econometrics*, 27:773–794.

Bertoli, S. and Moraga, J. F.-H. (2013). Multilateral resistance to migration. *Journal of Development Economics*, 102:79–100.

Chudik, A., Pesaran, M. H., and Tosetti, E. (2011). Weak and strong cross-section dependence and estimation of large panels. *Econometrics Journal*, 14:45–90.

Davis, P. (2002). Estimating multi-way error components models with unbalanced data structures. *Journal of Econometrics*, 106(1):67–95.

de Nardis, S. and Vicarelli, C. (2003). Currency unions and trade: The special case of EMU. *Review of World Economics*, 139(4):625–649.

Feenstra, R. C. (2004). *Advanced International Trade*. Princeton University Press, Princeton, NJ.

Frankel, J. (2008). The estimated effects of the euro on trade: Why are they below historical effects of monetary unions among smaller countries? NBER Working Papers 14542, National Bureau of Economic Research, Inc.

Frankel, J. and Rose, A. (2002). An estimate of the effect of common currencies on trade and income. *The Quarterly Journal of Economics*, 117(2):437–466.

Glick, R. and Rose, A. (2002). Does a currency union affect trade? The time series evidence. *European Economic Review*, 46:1125–1151.

Gunnella, V., Mastromarco, C., Serlenga, L., and Shin, Y. (2015). The Euro effects on intra-EU trade flows and balances: Evidence from the cross sectionally correlated panel gravity models. Technical report, University of York.

Hausman, J. A. and Taylor, W. E. (1981). Panel data and unobservable individual effects. *Econometrica*, 49(6):1377–1398.

Helpman, E. (1987). Imperfect competition and international trade: Evidence from fourteen industrialized countries. *Journal of the Japanese and International Economies*, 1:62–81.

Herwartz, H. and Weber, H. (2015). The Euro's trade effect under cross-sectional heterogeneity and stochastic resistance. Technical Report 1631, Kiel Institute for the World Economy, Germany.

Holm, H. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2):65–70.

Kapetanios, G., Mitchell, J., and Shin, Y. (2014). A nonlinear panel data model of cross-sectional dependence. *Journal of Econometrics*, 179:134–157.

Kapetanios, G. and Pesaran, M. H. (2005). Alternative approaches to estimation and inference in large multifactor panels: Small sample results with an application to modelling of asset returns. Technical Report 1416, CESifo Group Munich, CESifo Working Paper Series.

Kapetanios, G. and Shin, Y. (2017). Multi-dimensional heterogeneous panel data with hierarchical multi-factor error structure. Technical report, University of York.

Kramarz, F., Machin, S., and Ouazad, A. (2008). What makes a test score? The respective contributions of pupils, schools, and peers in achievement in english primary education. Technical report, INSEAD Working Paper.

Krugman, P. (1997). Increasing returns, monopolistic competition and international trade. *Journal of International Economics*, 9:469–479.

Kuersteiner, G. and Prucha, I. (2015). Dynamic spatial panel models: Networks, common shocks, and sequential exogeneity. Technical report, University of Maryland.

Mastromarco, C., Serlenga, L., and Shin, Y. (2016a). Modelling technical efficiency in cross sectionally dependent stochastic frontier panels. *Journal of Applied Econometrics*, 31(1):281–297.

Mastromarco, C., Serlenga, L., and Shin, Y. (2016b). Multilateral resistance and euro effects on trade flows. In Arbia, G. and Patuelli, R., editors, *Spatial Econometric Interaction Modelling*, pages 253–278. Springer: Berlin.

Pakko, M. and Wall, H. (2002). Reconsidering the trade-creating effects of a currency union. *Federal Reserve Bank of St Louis Review*, 83:37–46.

Persson, T. (2001). Currency unions and trade: How large is the treatment effect? *Economic Policy*, 16(33):433–462.

Pesaran, M. H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica*, 74(4):967–1012.

Pesaran, M. H. (2015). Testing weak cross-sectional dependence in large panels. *Econometric Reviews*, 34(6-10):1089–1117.

Rose, A. K. (2001). Currency unions and trade: The effect is large. *Economic Policy*, 16(33):449–461.

Serlenga, L. and Shin, Y. (2007). Gravity models of intra-EU trade: Application of the CCEP-HT estimation in heterogeneous panels with unobserved common time-specific factors. *Journal of Applied Econometrics*, 22(2):361–381.

Shi, W. and Lee, L. (2017). Spatial dynamic panel data models with interactive fixed effects. *Journal of Econometrics*.

Wansbeek, T. and Kapteyn, A. (1989). Estimation of error-components model with incomplete panels. *Journal of Econometrics*, 41:341–361.

Wooldridge, J. (2010). Correlated random effects models with unbalanced panels. Technical report, Michigan State University.

Chapter 11
The Estimation of Gravity Models in International Trade

Badi H. Baltagi, Peter H. Egger and Katharina Erhardt

Abstract Over the last few decades, multi-indexed data on trade, multinational activity, and even migration have become available. By far the most prominent application of multi-dimensional data in the context of international economics is the estimation of the famous gravity equation of international trade, where bilateral export or import volume (or foreign direct investment stock or migration stock) is the dependent variable of interest. This chapter provides a survey of empirical issues in gravity-model estimation from a panel econometric perspective. It sets off with a generic illustration of the theoretical foundations of gravity equations and proceeds with the modelling of the multi-dimensional stochastic structure, focusing on fixed-effects estimation.

11.1 Introduction

By far the most widely used application of multi-dimensional data in the context of international economics is the estimation of the famous gravity equation of international trade. Even though gravity equations are estimated with bilateral foreign direct investment, bilateral migration, or bilateral services trade as a dependent variable, most of the empirical work is available for bilateral goods-export or import volume as the dependent variable of interest. The explicit econometric treatment of the multi-dimensionality of the data is relatively young, and, for structural theoretical reasons, the dominant approach is the use of multiple fixed effects in estimation.

Badi H. Baltagi
Syracuse University, New York, USA

Peter H. Egger
ETH Zurich, CEPR, and CESifo, Zurich, Switzerland

Katharina Erhardt
ETH Zurich, Zurich, Switzerland

L. Matyas (ed.), *The Econometrics of Multi-dimensional Panels*, Advanced Studies in Theoretical and Applied Econometrics 50, DOI 10.1007/978-3-319-60783-2_11

This chapter provides a broad survey of the empirical literature in international economics from a panel econometric perspective. It starts off with a generic illustration of the theoretical foundations of gravity equations and the structural background of the parameters of interest. It proceeds with the modelling of the multi-dimensional stochastic structure, mainly focusing on fixed effects frameworks. Subsequently, it outlines specific problems that frequently emerge in empirical studies. Most notably, these problems relate to the estimation of log-linear versus exponential-family models, the treatment of an excessive mass of zeroes in bilateral trade data, the dependence of data in the time (dynamics) or cross-sectional dimension (spatial or contagious processes), the endogeneity of regressors, and the estimation of ratio-transformed models.

11.2 Generic Theoretical Background

As outlined above, the customary gravity equation involves bilateral commodity flows between exporting and importing countries. Therefore, it is a classical example of panel data. Even in the simplest case of cross-sectional data, the dependent variable is double-indexed, Y_{ij}, where the indices pertain to the exporter country $i \in \{1,...,N_1\}$ and the importer country $j \in \{1,...,N_2\}$. Observing such data over time (years or months) and/or products increases the dimensionality of the data, leading to multi-dimensional problems.

Not only is the gravity equation arguably the most successful empirical specification in international trade, but it is also probably one of the most important applications of panel-data – if not of multi-dimensional panel-data – techniques in general. Its success story reaches back to Tinbergen (1962), who modeled bilateral trade flows as being proportional to the product of the economic size of the trading partners reflected in their GDPs and inversely proportional to the geographic distance between these countries. It is this resemblance to Newton's law of universal gravitation that the gravity equation owes its name.

Early empirical specifications of the gravity equation in international economics lacked a deep theoretical foundation but became successful mainly thanks to their strong explanatory power. Starting with Anderson (1979), theoretical trade models were developed that actually provided a theoretical foundation for gravity equations (see also Bergstrand, 1989, 1990; Deardorff, 1998). However, regarding theory-consistent estimation, it was the seminal work of Anderson and van Wincoop (2003) that took gravity research a big step further. Anderson and van Wincoop (2003) derived the first so-called iterative-structural gravity equation emerging from a general equilibrium model of trade based on Armington preferences.

It will be useful to consider a generic version of this simplistic model for illustration. In this model, countries are endowed with a fixed amount of goods, so that the main economic problem boils down to distributing these goods across customers. Goods are differentiated by their country of origin, and agents have Dixit-Stiglitz preferences over these goods. The quantity of the country-*i*-specific goods shipped

to country j, Q_{ij}, is determined by maximizing the utility of country j's representative consumer subject to their budget constraint

$$\max \quad U_j = \left[\sum_{i=1}^{N_1} Q_{ij}^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}} \quad \text{s.t.} \quad \sum_{i=1}^{N_1} P_{ij} Q_{ij} = E_j, \tag{11.1}$$

where $\sigma > 1$ denotes the elasticity of substitution, P_{ij} is the price paid by consumers in country j for goods from country i, and E_j is the total expenditure volume in country j. While sellers offer goods at the same price to everybody, P_i, consumer prices vary across destinations, P_{ij}. Assuming iceberg-type transport costs, $C_{ij} \geq 1$, consumer prices relate to producer prices and transport costs by $P_{ij} = P_i C_{ij}$. From (11.1), it follows that bilateral trade flows (sales) from i to j are given by

$$Y_{ij} = Q_{ij} P_{ij} = \frac{(P_i C_{ij})^{1-\sigma}}{\sum_{k=1}^{N_1} (P_k C_{kj})^{1-\sigma}} E_j. \tag{11.2}$$

Put simply, how appealing a good from i is for potential buyers in j depends on its price relative to the price of all other options the same buyer has. To obtain equilibrium prices, market-clearing has to prevail: the value of total sales of good i has to equal total consumption expenditures for that good: $Y_i = \sum_{j=1}^{N_2} Y_{ij}$. Combining the market-clearing condition with (11.2) yields the so-called structural gravity that explains bilateral trade flows between country i and j taking into account the endogenous nature of prices

$$Y_{ij} = \frac{Y_i}{\sum_{k=1}^{N_2} \frac{C_{ik}^{1-\sigma} E_k}{\sum_{m=1}^{N_1} (P_m C_{mk})^{1-\sigma}}} \frac{E_j}{\sum_{k=1}^{N_1} (P_k C_{kj})^{1-\sigma}} C_{ij}^{1-\sigma} \equiv \frac{Y_i}{\mathscr{P}_i} \frac{E_j}{\mathscr{P}_j^*} C_{ij}^{1-\sigma}. \tag{11.3}$$

The terms in the denominators, $\mathscr{P}_i$ and $\mathscr{P}_j^*$, are commonly called *multilateral resistance terms* because they account for third-country effects in determining bilateral trade flows (see Anderson and van Wincoop, 2003).[1] Considering the first term, if products from i face more restrictive trade barriers in other countries than j, $\mathscr{P}_i$ falls and country i will direct a larger share of its production and sales to j. On the other hand, if products from countries other than i become cheaper, $\mathscr{P}_j^*$ rises and country j will purchase less goods from i.

Trade economists are especially interested in estimating the partial effect of bilateral trade costs on trade flows as reflected in the so-called *trade elasticity*, κ. In the context of the Armington model, trade elasticity relates to the demand parameter, $\kappa = 1 - \sigma$. Trade elasticity is an important parameter to evaluate the effect of changes in trade costs (e.g., a reduction of tariffs) on trade flows, but it is also an important statistic for evaluating welfare effects.

[1] Note that $\mathscr{P}_j^* = \left(\sum_{i=1}^{N_1} (P_i C_{ij})^{1-\sigma} \right)^{\frac{1}{1-\sigma}}$ denotes the classical constant-elasticity-of-substitution price index.

To see the latter, define welfare as the value of production in a country scaled by the price index prevailing in that country, i.e., real consumption volume, $W_i = Y_i/\mathscr{P}_i^*$. Using the convention to denote variables in logs by lower-case letters, relative changes in welfare, $dW_i/W_i = d\ln W_i = dw_i$, are hence given by

$$dw_i = dy_i - d\ln(\mathscr{P}_i^*). \tag{11.4}$$

Note that in an endowment economy, the quantity of production does not change. Taking the good produced in country i as the numeraire, its price will not change either, and $dy_i = 0$. It can then be shown that[2]

$$dw_i = -\sum_{m=1}^{N_1} \frac{Y_{mi}}{E_i}(dp_m + dc_{mi}). \tag{11.5}$$

From (11.2), it follows that

$$\frac{\frac{Y_{mi}}{E_i}}{\frac{Y_{ii}}{E_i}} = \frac{(P_m C_{mi})^{1-\sigma}}{(P_i C_{ii})^{1-\sigma}}. \tag{11.6}$$

Noting that we defined the good in i as the numeraire and setting $C_{ii} = 1$, we can rewrite (11.6)

$$d\ln\frac{Y_{mi}}{E_i} - d\ln\frac{Y_{ii}}{E_i} = (1-\sigma)(dp_m + dc_{mi}). \tag{11.7}$$

Plugging this back into (11.5), welfare changes can be expressed in terms of the share of domestic absorption and trade elasticity:

$$dw_i = \frac{d(y_{ii} - e_i)}{(1-\sigma)}. \tag{11.8}$$

Hence, to recover the welfare change from moving to autarky, where all production is consumed domestically, the researcher only needs to calculate the change from the current observed domestic absorption share to a counter-factual situation with full domestic absorption and scale it by the trade elasticity.[3] An estimate of the trade elasticity can be recovered from equation (11.3), if ad-valorem trade costs (such as bilateral applied tariff factors) are used in the parametrization of bilateral trade costs, C_{ij}. Since the multilateral resistance terms are inherently unobservable to the researcher, one can consistently estimate (11.3) using fixed or random effects. However, in finite samples, random effects estimation will not produce unbiased results, since the multilateral resistance terms are correlated with both the exporter trade potential and bilateral trade costs (see Eaton and Kortum, 2002; Anderson

[2] To see, this plug in (11.2) for Y_{mi} and rearrange terms to get $\sum_{m=1}^{N_1} \frac{Y_{mi}}{Y_i}(dp_m + dc_{mi}) = d\ln\mathscr{P}_i^*$.

[3] This is what is often referred to as the *gains from trade* (compare Arkolakis et al., 2012).

and van Wincoop, 2003).[4] Rewriting (11.3) in logs, using appropriate fixed effects $\{\bar{\alpha}_i, \bar{\alpha}_j^*\}$ and setting $1-\sigma=\kappa$ yields

$$y_{ij} = \bar{\alpha}_i + \bar{\alpha}_j^* + \kappa \cdot c_{ij} + \varepsilon_{ij}. \tag{11.9}$$

Adding a time index to the structure in (11.9), the empirical models by Matyas (1997), Matyas (1998), Hummels (1999), Egger (2000), Rose and van Wincoop (2001), Glick and Rose (2002), Eaton and Kortum (2002), Egger and Pfaffermayr (2003), Cheng and Wall (2005), Baldwin and Taglioni (2006), Baier and Bergstrand (2007), Shin and Serlenga (2007) or Armenter and Koren (2014) can be viewed as more or less restricted variants of a more generally indexed empirical model of the form

$$y_{ijt} = \lambda_t + \bar{\alpha}_i + \bar{\alpha}_j^* + \alpha_{it} + \alpha_{jt}^* + \gamma_{ij} + \varepsilon_{ijt}, \tag{11.10}$$

where either a cross section is considered (the time index is missing) so that the above model collapses to one with effects $\bar{\alpha}_i$ and $\bar{\alpha}_j^*$ as in (11.9), or the interactive effects $\{\alpha_{it}, \alpha_{jt}^*, \gamma_{ij}\}$ are assumed to be zero, or at least some of the interactive effects are restricted to zero – namely either γ_{ij} or both α_{it} and α_{jt}^*. Clearly, the model in (11.10) could not be estimated without restrictions, since the interactive effects $\{\alpha_{it}, \alpha_{jt}^*, \gamma_{ij}\}$ are nested in the main effects $\{\bar{\alpha}_i, \bar{\alpha}_j^*, \lambda_t\}$, so that the effects $\{\bar{\alpha}_i, \bar{\alpha}_j^*, \lambda_t\}$ could not be identified if $\{\alpha_{it}, \alpha_{jt}^*, \gamma_{ij}\}$ were included without restrictions. The general version of the model in (11.10) when subsuming the main effects $\{\bar{\alpha}_i, \bar{\alpha}_j^*, \lambda_t\}$ in $\{\alpha_{it}, \alpha_{jt}^*, \gamma_{ij}\}$ has been proposed and estimated by Baltagi et al. (2003).[5]

Structural gravity equations such as (11.10) follow not only from the rather parsimonious Armington model, but from a very wide class of trade models. Arkolakis et al. (2012) show that all models that feature (i) Dixit-Stiglitz preferences; (ii) one factor of production; (iii) linear cost functions; and (iv) perfect or monopolistic competition; (v) balanced trade; (vi) aggregate profits that are a constant share of aggregate revenues, as well as (vii) a CES import demand system, are nested by the general model (11.10) and can be consistently estimated by multi-dimensional fixed effects.

While these assumptions seem rather restrictive at first glance, several well-known trade models – including most quantitative trade models – are covered by the structure in (11.10) and an associated general equilibrium analysis. Specifically, let us briefly discuss two of the most influential trade models of recent years that are both nested by the structural gravity expression (11.10) but are subject to very different economic frameworks.

---

[4] Pöyhönen (1963) used fixed exporter and importer-country effects to estimate a cross-sectional gravity equation, but the fixed effects there did not have a footing in general equilibrium, unlike the work of Eaton and Kortum (2002); Anderson and van Wincoop (2003).

[5] Egger and Nigai (2015) illustrate that the customary *extraction* of trade costs from trade flows in structural calibrated models corresponds to "estimating" the model in (11.10) by estimating as many parameters as there are observations, which is a saturated model, where the variance-covariance matrix of the parameters is degenerate.

The seminal model of Eaton and Kortum (2002) is based on Ricardian comparative advantage, where productivity in each industry is drawn from a Frechet distribution. Since goods within an industry are perfect substitutes, the lowest-price seller to a specific destination will serve *all* demand for that particular industry. The probability of selling to a specific market is, hence, a function of the exporting country's location parameter of the productivity distribution, the prevailing costs of inputs in the exporting country, and the iceberg-type trade costs between exporter and importer weighted by the price distribution across all other potential exporters. Within this framework, the trade elasticity, κ, that governs the impact of a change in trade costs on trade flows depends on the shape of the productivity distribution – hence, a supply-side parameter.

In contrast to this setting, Melitz (2003) relies on a Dixit-Stiglitz preference structure. Each firm produces a variety of a good, and varieties are imperfect substitutes with the elasticity of substitution being governed by σ. Firms draw their productivity from a country-specific distribution and compete in a monopolistic competition setting with constant mark-ups over marginal costs and fixed costs of exporting.[6] Firms are heterogeneous and charge a price depending on their productivity level and the country-specific costs of inputs. Consumers choose the consumption quantity of a variety conditional on the price of this particular variety relative to the price of all other varieties and the substitutability between varieties. The more expensive a variety is due to low productivity, high input costs, or high trade costs of that variety, the lower its potential sales are in an export market. Firms with a higher productivity will ceteris paribus make higher profits. As exporting involves fixed costs, the least productive producers in a market will only serve domestic customers. Importantly, the consumption bundle of each consumer contains *every available* variety with the consumed quantity being determined by its price and the substitutability across varieties. Clearly, in this setting, the impact of trade costs on trade flows κ depends on the elasticity of substitution between varieties – hence, a demand-side parameter.

Since the trade elasticity κ – despite its different structural interpretation across theoretical models – is a sufficient statistic to recover welfare gains in all the models characterized in Arkolakis et al. (2012) and Head and Mayer (2014), proper estimation of the empirical model is absolutely crucial.

11.3 Specific Problems with Estimating Gravity Models

In this section, we focus on a selected set of issues in the estimation of gravity models that appears particularly important in view of the number of applications associated with them. Due to space constraints, we do not explicitly address issues related to the estimation of fixed versus random effects models and of the retrieval of parameters on variables which are collinear with fixed effects in a particular dimension. We refer the reader to Baltagi et al. (2015) for a treatment of these problems.

[6] Due to its simple tractability, the dominant assumption of the form of the productivity distribution is a single-parameter Pareto.

11.3.1 Heteroskedasticity

One problem with the estimation of gravity models is that the theoretically founded specification of bilateral trade flows in modern international economics is one where data on bilateral trade flows in levels are assumed to be generated by a multiplicative function of determinants as in (11.2) and (11.3). A log-transformation of an originally exponential form of the data-generating process may lead to biased response parameters of interest on observable determinants, if the disturbances are heteroskedastic and are a function of the conditional mean (see Nelder and Wedderburn, 1972; McCullagh and Nelder, 1989). Rather than log-transforming such models, it is preferable to estimate them in their original, exponential form upon specification of the density function of the disturbances and the link function between the disturbances and the conditional mean of the model (i.e., the nature of heteroskedasticity).

Santos Silva and Tenreyro (2006) emphasized this general issue in the context of gravity models, illustrating for a cross section of bilateral trade data that, depending on the degree of heteroskedasticity of the disturbances, ordinary least squares (with or without fixed exporter and importer effects) obtains biased estimates of the parameters of interest, e.g., on trade-cost variables. They assumed a Poisson distribution for the disturbances and a log-linear link function, where the variance of the disturbances is proportional to the conditional mean of the data.

Following Egger and Staub (2016), an exponential-family, generalized-linear version of the gravity model in (11.10) is given by

$$Y_{ijt} = \exp(\lambda_t + \bar{\alpha}_i + \bar{\alpha}^*_j + \alpha_{it} + \alpha^*_{jt} + \gamma_{ij})\varepsilon_{ijt}, \text{ or} \tag{11.11}$$

$$Y_{ijt} = \exp(\lambda_t + \bar{\alpha}_i + \bar{\alpha}^*_j + \alpha_{it} + \alpha^*_{jt} + \gamma_{ij}) + \nu_{ijt}, \tag{11.12}$$

where $\varepsilon_{ijt} = 1 + \nu_{ijt}/\exp(\lambda_t + \bar{\alpha}_i + \bar{\alpha}^*_j + \alpha_{it} + \alpha^*_{jt} + \gamma_{ij})$. A log-linear transformation of this model will only obtain consistent ordinary-least-squares estimates of the parameters of interest, such as $\{\lambda_t, \bar{\alpha}_i, \bar{\alpha}^*_j, \alpha_{it}, \alpha^*_{jt}\}$ and a parametrization of γ_{ij}, if $\mathrm{E}(\ln \varepsilon_{ijt})$ conditional on $\{\lambda_t, \bar{\alpha}_i, \bar{\alpha}^*_j, \alpha_{it}, \alpha^*_{jt}\}$ and the parametrization of γ_{ij} is zero. However, ε_{ijt} is a function of the conditional mean of the model, $\mathrm{E}(Y_{ijt}) = \exp(\lambda_t + \bar{\alpha}_i + \bar{\alpha}^*_j + \alpha_{it} + \alpha^*_{jt} + \gamma_{ij})$ and depends on the ingredients of the conditional mean as well as on ν_{ijt}. Hence, except for very special cases, OLS will be inconsistent if the conditional expectation function of bilateral trade flows is exponential.

In conventional gravity models, the interior of the exponential term on the right-hand side of (11.11) is additive. In general, one speaks of this function as the link function, and since it is log-linear in gravity models, they are said to have a log-link. The first-order conditions of the model are inverse-proportional to the conditional variance of the dependent variable, Y_{ijt}. Hence, for estimation the functional form of this variance may be crucial. Egger and Staub (2016) emphasized that statistical tests should be consulted in choosing the functional form within the exponential family in applications. However, they performed a set of simulation exercises, constructing data from a calibrated structural multi-country gravity model and found

that in such simulations there was little difference between Poisson, the negative binomial, the Gaussian, the inverse Gaussian, and the gamma model in the simulated data. Nevertheless, here were relatively sizable differences in these models with their empirical application. Moreover and quite interestingly, in the simulated data, an old-style empirical gravity without fixed country effects (and without their structural theoretical equivalents of them) – i.e., a model which simply included log exporter GDP, log-importer GDP, and log trade costs in the linear index as the determinants of bilateral exports akin to Tinbergen (1962) – obtained virtually unbiased estimates of the parameter on trade costs (the trade elasticity, κ). This was not the case with their empirical application, using observed rather than generated trade data. Given that the simulated data had been generated in accordance with empirical data on (factor) endowments and trade costs, the results in Egger and Staub (2016) suggest that the sizable differences between different exponential-family models, on the one hand, and the old-fashioned gravity model, on the other, with real-world data originate in problems beyond the standard structural gravity model, in particular, the endogeneity of trade costs.

In the data, the sources of heteroskedasticity may be manifold. One reason for heteroskedasticity could be a mass point at certain levels of trade flows. E.g., such a mass point could emerge with truncated or censored data, and an empirically important mass point is the one at zero bilateral trade flows. However, while zero bilateral trade flows will not be dropped or lost with an exponential model formulation as in (11.11), it should be emphasized that exponential-family models are incapable of treating any mass point unless it is specifically modelled (see Egger et al., 2011). Hence, the exponential-family-model estimation per se may only be a suitable remedy of the bias addressed above in the absence of mass points in the data. Therefore, the next subsection is explicitly devoted to a treatment of excessive zeros in the data.

11.3.2 Modelling the Mass Point at Zero Bilateral Trade Flows

Consider once again the equation explaining bilateral trade flows in the Armington model (11.3). Clearly, all countries with a positive value of production and expenditure will trade with each other even if the trade value might be small for countries that are far apart or have small domestic production. However, the data teach us otherwise, and there are many zero bilateral trade flows. Helpman et al. (2008) show that around 50% of potential trade relationships are subject to zero trade flows in a cross section of data, and the evidence in Egger and Pfaffermayr (2011) suggests that, while the percentage of zeros is falling, it is still non-negligible even in recent years. Clearly, the evidence in Helpman et al. (2008) and Egger and Pfaffermayr (2011) is for aggregate trade data, and the numbers for more disaggregated, product-level data are even worse.

Many standard gravity approaches simply neglect the presence of zeros and estimate the empirical model on positive trade flows only. This practice, however, might lead to severely biased parameter estimates of interest. Helpman et al. (2008)

make two contributions to the literature. First, they develop a theoretical model that leads to zero trade flows between countries. Second, they take this selection explicitly into account when estimating trade flows. They build on a Melitz (2003) type trade model with fixed costs of exporting between pairs of countries and a truncated Pareto distribution for firms' productivity. In this model, the maximum productivity drawn by firms in a country may be too low to profitably export to a particular other country at a given level of fixed trade costs. As before, the cross section bilateral export flows between country i and country j can be explained by importer and exporter fixed effects, $\bar{\alpha}_i$ and $\bar{\alpha}_j^*$, as well as a bilateral component which in this particular case consists of standard iceberg-type trade costs c_{ij} and a bilateral variable that depends on the share of exporters from i to j and can potentially be zero, w_{ij}:

$$y_{ij} = \lambda + \bar{\alpha}_i + \bar{\alpha}_j^* + \kappa \cdot c_{ij} + w_{ij} + \varepsilon_{ij}. \tag{11.13}$$

By including w_{ij}, Helpman et al. (2008) control for the share of exporters in i that serve customers in j (and their average productivity and the price they charge). However, by taking logs of aggregate trade flows, any zero trade flow observations are dropped. Since zeros do not emerge at random, but for those country pairs where fixed trade costs are high and/or delivery prices are high, Helpman et al. (2008) and Egger and Pfaffermayr (2011) account for the self-selection of firms and country pairs into positive exports. Helpman et al. (2008) do so with a Heckman-type model in a cross section of data, and Egger and Pfaffermayr (2011) employ a dynamic panel-data selection model. In any case, w_{ij} involves an inverse Mills' ratio to account for selection. The share of firms in i exporting to j (at a given point in time) as well as the inverse Mills' ratio are modelled as functions of a set of variables z_{ij} informed by the zero profit condition of the theoretical model. Beyond the standard gravity variables, it is customary to include identifying instruments in z_{ij}, which are suggested by the zero-profit market-entry condition for firms in i about market j. Such variables are supposed to reflect fixed market-entry costs. Helpman et al. (2008) rely on data on the regulation cost of firm entry and alternatively a variable that indicates a common religion between two countries as instruments. In a first stage, Helpman et al. (2008) and Egger and Pfaffermayr (2011) use a Probit model to estimate the probability of positive exports from i to j (in year t), $\rho_{ij(t)}$. In a second stage, they estimate gravity equations of the form

$$y_{ij(t)} = \bar{\alpha}_{i(t)} + \bar{\alpha}_{j(t)}^* + \kappa \cdot c_{ij(t)} + \hat{w}_{ij(t)}(\hat{\rho}_{ij(t)}, \delta) + \gamma \widehat{Mills}_{ij(t)}(\hat{\rho}_{ij(t)}) + \varepsilon_{ij(t)}, \tag{11.14}$$

where $\hat{w}_{ij(t)}(\cdot)$ is a non-linear function of its argument reflecting the share of exporters from i to j with the parameter δ to be estimated,[7] and $\widehat{Mills}_{ij(t)}(\hat{\rho}_{ij(t)})$ is the Mills' ratio that accounts for self-selection into positive exports (sample selection). Santos Silva and Tenreyro (2015) show that consistent estimation of a standard Helpman et al. (2008) approach relies on strong homoskedasticity assumptions that

[7] Note that $\hat{w}_{ij(t)}(\hat{\rho}_{ij(t)})$ is present only in Helpman et al. (2008) but not in Egger and Pfaffermayr (2011), who use a model with homogeneous firms.

are typically violated in trade data,[8] and that the instruments proposed in Helpman et al. (2008) are weak.

Egger et al. (2011) specify a two-part gravity model to account for zero trade flows (see Cragg, 1971; Duan et al., 1984) in a model with homogeneous firms. In contrast to a log-linearized model as in Helpman et al. (2008) or Egger and Pfaffermayr (2011), they use an exponential specification of the gravity equation

$$Y_{ij} = \mathscr{I}_{ij}(z_{ij}) \exp(\bar{\alpha}_i + \bar{\alpha}_j^* + \kappa \cdot c_{ij}) + \varepsilon_{ij}, \tag{11.15}$$

where z_{ij} denotes the determinants of participation in exporting that could potentially include other determinants beyond the covariates in (11.15). Since Egger et al. (2011) examine the effects of a binary variable that indicates membership in preferential trade agreements (PTA), they model the probabilities of positive exports and PTA membership by a bivariate Probit that allows for the endogenous nature of the explanatory variable PTA membership.[9]

In contrast to the two-step Heckman approach in Helpman et al. (2008), two-part models do not explicitly model the correlation structure between the errors in the participation equation and the outcome equation but allow for stochastic interdependence between the errors under some general class of joint distributions (see Egger et al., 2011; Duan et al., 1984). Therefore, two-part models are more efficient under the interdependence of the disturbances, while strongly correlated disturbances might call for an explicit treatment of this relationship between the errors.[10]

In contrast to the approaches outlined in this subsection, many authors follow Santos Silva and Tenreyro (2006) in allowing for zero trade flows in an exponential specification of the gravity equation without explicitly modeling the mass point at zero. In these models, the extensive country margin and the intensive country margin of trade cannot be disentangled, and, with many zeros, the regression parameters of interest may be biased (see Sect. 11.3.1).[11]

11.3.3 Dynamics

Dynamic gravity models with a lag order of S – when log-linearized – would generally be of the form

$$y_{ijt} = \left(\sum_{s=1}^{S} \lambda_s y_{ijt} \right) + \alpha_{it} + \alpha_{jt}^* + \kappa \cdot c_{ijt} + \varepsilon_{ijt}. \;\; t = 1, 2, .., T. \tag{11.16}$$

[8] Compare section 11.3.1.

[9] See Sect. 11.3.5 for a detailed treatment of endogenous regressors in gravity equations.

[10] In order to tackle a potential correlation between the disturbances, Egger et al. (2011) employ an inverse Mills' ratio similar to Helpman et al. (2008) in an alternative specification.

[11] Some authors have accounted for zero export flows by adding constants to bilateral exports and then using a Tobit approach (compare Felbermayr and Kohler, 2006). Santos Silva and Tenreyro (2006) illustrate that this practice is likely to lead to parameter bias and should not be used.

Depending on whether ε_{ijt} contains error components – in particular, ij-indexed ones – or not, the usual problems with the levels-estimation of dynamic models emerge (see Hsiao, 2015; Arellano and Bond, 1991; Baltagi, 2013). Clearly, either instrumental-variables estimation of the model in levels or the model in differences is the estimation strategy to use in this case. Such models have been estimated by Egger (2001), Bun and Klaassen (2002), Olivero and Yotov (2012), Nuroglu and Kunst (2014), to mention a few.

However, whether gravity models in differences require any instrumentation of the lagged dependent variable or not depends on the underlying assumptions. While the bilateral trade data used in most applications are characterized by a cross section of country pairs that is much larger than the time series, it is possible to track trade flows between most country pairs for several decades. Hence, for large T the Nickell (1981) bias is of a small order, and the chance to induce more bias by using an ill-suited set of instruments than when ignoring the endogeneity of the differenced lagged dependent variable is not negligible. But whether pursuing this strategy is possible or not depends on the assumptions about parameter stability. We know that fundamental parameters, such as the relative importance of geography, culture and other variables, change over time and are not constant for decades (see Disdier and Head, 2008; Egger and Lassmann, 2012). Then, pooling the data for very long periods is not possible. When estimating this model for shorter sub-sets of time periods, the usual estimation methods to control for lagged-dependent-variable bias can be used.

11.3.4 Spatial Data: Interdependence of Bilateral Trade Flows Conditional on Exogenous Determinants

Modern quantitative gravity models of international trade all postulate that the bilateral demand and supply of goods (and services) add up to aggregate income and aggregate expenditures (at least up to some constraint as in Dekle et al., 2007). Accordingly, aggregate income and expenditures are an *endogenous common factor* for bilateral trade flows of any country, as is recognized in, e.g., (11.10). Moreover, any bilateral demand depends on the prices charged by firms in every source country, since consumers in these models compare the prices at which they buy from a given country or firm to the prices charged by all other countries or firms. This establishes a generic interdependence of all bilateral trade flows, and it means that a shock – to demand, supply, or trade costs – in any country or country-pair in the world induces effects on all other country-pairs, whether the partners in a given pair were directly exposed to the shock or not. In most empirical models (structural or reduced form), this kind of generic and ubiquitous interdependence of the world trading network is specified exclusively in terms of observables. In applications, however, it is doubtful whether all relevant explanatory variables are included in the model, and hence, whether all elements establishing interdependence are included in the empirical specification. Parting with the latter assumptions requires specifying an inter-

dependence structure at least of the stochastic process (in an error-interdependence model) if not the dependent variable (in a dependent-lag model). To the extent that interdependence in the error term of the dependent variable is modelled as to relate to geography and distance between countries, we may then speak of spatial error or spatial lag models of bilateral trade.

Rewriting the deterministic part of the right-hand side of (11.10) as H_{ijt} in order to focus on the spatial part this model, we may reformulate the estimating equation for (11.10) as

$$y_{ijt} = H_{ijt} + u_{ijt}, \tag{11.17}$$

where u_{ijt} is a generic disturbance term (any fixed effects would be included in H_{ijt}). E.g., the disturbance term might contain four components, namely $u_{ijt} = u_{it} + u_{jt} + u_{ij} + \varepsilon_{ijt}$. Any more restrictive version of the stochastic process can easily be obtained.

Towards formulating a spatial model, let us introduce so-called weights matrices. Let us use w to refer to elements of weights matrices W pertaining directly to the dependent variable in logs, y_{ijt}, and use m to refer to elements of weights matrices M pertaining to the stochastic process, u_{ijt}. All examples of these generic weights matrices have the property that (at least) their diagonal elements are zero, and that the blocks pertaining to different time periods are zero, $W = \text{diag}(W_t)$ and $M = \text{diag}(M_t)$. Hence, these matrices all capture the notion of contemporaneous interdependence in the cross section. Moreover, let us assume that the elements of the weights matrices are normalized such that either the sum of elements in any row is unity or that the minimum of the maximum row and columns sums is unity (see Kelejian and Prucha, 2010). In principle the weights matrices and their elements are allowed to vary with time. Then, let us consider three forms of geography: one that pertains to the dependence among exporting countries in year t (captured by elements w_{ijt}^{it} and m_{ijt}^{it}); one that pertains to the dependence among importing countries in year t (captured by elements w_{ijt}^{jt} and m_{ijt}^{jt}); and one that pertains to the dependence among country pairs beyond exporter and importer dependence in year t (captured by elements w_{ijt}^{ijt} and m_{ijt}^{ijt}). Then, we may formulate the general spatial process pertaining directly to the dependent variable as

$$\begin{aligned} y_{ijt} &= \left(\sum_{i=1}^{N_1} \lambda^{it} w_{ijt}^{it} y_{ijt} + \sum_{j=1}^{N_2} \lambda^{jt} w_{ijt}^{jt} y_{ijt} + \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \lambda^{ijt} w_{ijt}^{ijt} y_{ijt} \right) + H_{ijt} + u_{ijt}, \\ &\equiv \lambda^{it} \bar{y}_{ijt}^{it} + \lambda^{jt} \bar{y}_{ijt}^{jt} + \lambda^{ijt} \bar{y}_{ijt}^{ijt} + H_{ijt} + u_{ijt}, \end{aligned} \tag{11.18}$$

and we may specify the stochastic process – assuming first-order spatial autocorrelation – as

$$u_{ijt} = \bar{u}_{it}^{it} + \bar{u}_{jt}^{jt} \bar{u}_{ij}^{ij} + \bar{\varepsilon}_{ijt}^{ijt}, \tag{11.19}$$

where, in vector form for time period t, the right-hand-side terms are defined as elements of

$$\overline{u}^{it} = (I_{N^2} - \rho^{it} u^{it})^{-1} + \widetilde{u}^{it}, \tag{11.20}$$

$$\overline{u}^{jt} = (I_{N^2} - \rho^{jt} u^{jt})^{-1} + \widetilde{u}^{jt}, \tag{11.21}$$

$$\overline{u}^{ij} = (I_{N^2} - \rho^{ij} u^{ij})^{-1} + \widetilde{u}^{ij}, \tag{11.22}$$

$$\overline{\varepsilon}^{ijt} = (I_{N^2} - \rho^{ijt} \varepsilon^{ijt})^{-1} + \widetilde{\varepsilon}^{ijt}, \tag{11.23}$$

and all terms with a tilde are independently distributed error components. One could alternatively formulate the process in (11.19) by way of moving averages.

Clearly, the presence of a spatial lag of the dependent variable on the right-hand side of the model as in (11.18) requires the estimation of the (nonlinear) reduced form of the model by way of maximum-likelihood estimation (see Lee, 2004) or of the structural form by way of generalized-method-of-moments estimation (see Kelejian and Prucha, 1999; Kapoor et al., 2007).

For gravity models of foreign direct investment, versions of (11.18) and (11.20)–(11.23) have been estimated by, e.g., Coughlin and Segev (2000), Baltagi et al. (2007), and Blonigen et al. (2007). However, none of these papers consider the general forms in (11.18) and (11.20)–(11.23), but they postulate models with a single spatial term on the right-hand side – typically either the dependent variable or the disturbances. Badinger and Egger (2015) consider a form with three spatial lags (interdependence among exporters, importers, and pairs) at the same time.

There is much less work involving applications of spatial econometric methods on gravity models of bilateral trade or migration in comparison to foreign direct investment. The reason is that gravity models of trade and migration tend to be structural, and many researchers analyzing such models pay little attention to the stochastic process of the model. In any case, a few papers do consider trade and migration flow models with spatial dependence. Examples are LeSage and Pace (2008), LeSage and Fischer (2010), Beenstock and Felsenstein (2012), Beenstock and Felsenstein (2015), Behrens et al. (2012), Egger and Staub (2016), LeSage and Thomas-Agnan (2015), and Egger and Pfaffermayr (2016). As with foreign direct investment as the outcome, empirical spatial trade and migration gravity models tend to be much more restrictive than the process in (11.18) and (11.20)–(11.23). For instance, the cross-sectional model in Behrens et al. (2012), includes a spatial lag of the dependent variable on the right-hand side of the model but not a specific spatial error structure. However, they present a relatively complicated spatial-moving-average model with country-specific coefficients and a spatial lag, as well as a spatial moving-average error process. Egger and Staub (2016) prove, though, that the process suggested for estimation is not consistent with the assumptions made by the authors. It turns out that the actual stochastic process flowing from their assumptions is one that can be handled without any spatial econometric methods. The reason is that the spatial weights matrices they assume do not have zero diagonal elements and contain identical weights per country. As outlined in Egger and Staub (2016), respecting this structure leads to a much simpler process than the one proposed in Behrens et al. (2012).

The empirical models proposed by LeSage and Pace (2008), LeSage and Fischer (2010), and LeSage and Thomas-Agnan (2015) explicitly focus on several spatial

lags of the dependent variable (bilateral migration) as in (11.18) but not the disturbances, and they are not derived from general equilibrium theory. On the contrary, Egger and Pfaffermayr (2016) focus on various forms of spatial dependence in the disturbances in bilateral trade-flow equations as in (11.20)–(11.23) in a structural gravity model of bilateral trade.

11.3.5 Endogenous Regressors

Multi-dimensional panel techniques can solve several endogeneity problems arising in the context of estimating bilateral trade flows. The most obvious endogeneity problem arises in naive specifications of the gravity model, where bilateral trade flows are regressed on importer and exporter-GDP, as well as on some bilateral trade cost variable c_{ijt}

$$y_{ijt} = \alpha \cdot y_{it} + \beta \cdot y_{jt} + \kappa \cdot c_{ijt} + \varepsilon_{ijt}. \tag{11.24}$$

This naive specification of the gravity equation does not include the multilateral resistance terms that relate the bilateral variables to the situation relative to the rest of the world as in (11.3). If the customary economic theory underlying (11.3) is correct and the number of countries is not too small (see Egger and Staub, 2016), estimating (11.24) will lead to biased estimates due to an omitted variables bias. Consider again the structural gravity (11.3). Assuming balanced trade, $Y_{it} = E_{it}$, equation (11.3) can be rewritten in log-terms

$$y_{ijt} = y_{it} - \log(\mathscr{P}_{it}) + y_{jt} - \log(\mathscr{P}^*_{jt}) + \kappa \cdot c_{ijt} + \varepsilon_{ijt}. \tag{11.25}$$

As already alluded to, including fixed-effects α_{it} and α^*_{jt} can easily resolve the omitted variable bias in this particular case. However, endogeneity remains to be solved if c_{ijt} includes an endogenous variable whose variation is not fully captured by fixed effects α_{it} and α^*_{jt}.

Two of the main research lines of empirical studies in international trade based on variants of the gravity equation are evaluating the role and economic effects of trade policy instruments and estimating trade elasticities which are a sufficient statistic for the welfare effects of trade frictions as targeted in many theoretical trade applications and counterfactual analyses. In both lines of research, the coefficient of interest is estimated on trade policy variables such as tariffs, monetary unions or preferential trade area (PTA) indicator variables which form part of c_{ijt}. These variables are not captured by the fixed effects α_{it} and α^*_{jt}.

Let us reconsider the standard structural gravity equation (11.3) (assuming balanced trade) and extend it in order to examine the effect of a binary policy variable such as a PTA[12]

[12] Clearly, the binary variable could be any other binary policy treatment, such as a monetary union, and specific type of PTA, such as a free trade area or a customs union. Note that we consider

$$y_{ijt} = y_{it} - \log(\mathscr{P}_{it}) + y_{jt} - \log(\mathscr{P}^*_{jt}) + \beta PTA_{ijt} + \kappa \cdot \tilde{c}_{ijt} + \varepsilon_{ijt}, \tag{11.26}$$

with $\tilde{c}_{ijt}$ denoting the *other* observable bilateral components that are observable to the econometrician, such as distance, common language dummies or a common colonial history. Estimation of β is consistent iff

$$\mathrm{E}(\varepsilon_{ijt}|PTA_{ijt}, y_{it}, y_{jt}, \mathscr{P}_{it}, \mathscr{P}^*_{jt}, \tilde{c}_{ijt}) = 0, \tag{11.27}$$

hence, if there are no unobservable determinants of PTAs that are at the same time determining trade flows.

Baier and Bergstrand (2004) examine the determinants of forming a PTA and find mostly the same driving factors behind PTA membership that are also direct drivers of trade: two countries tend to select into a PTA if they have larger and more similar GDPs, are geographically closer to each other but remote to the rest of the world, and if they are different in their respective relative factor endowments. Apart from these observable factors, one could think of several unobservable factors that might influence trade policy and trade flows jointly. E.g., any domestic trade impediment, such as internal shipping regulations that might be alleviated especially by so called "deep" trade agreements, is likely to bias the estimates on PTAs when omitted. The general result that countries have "chosen well" their PTA partners shows the very strong connection between basically all determinants of trade flows and trade policy and makes it hard to believe that the remaining unobservables are uncorrelated.[13] Given these severe endogeneity issues, it is not surprising that the trade policy effects found in a myriad of studies vary substantially across specifications and years, are highly unstable, and often yield implausible results (see Frankel, 1997; Baier and Bergstrand, 2002, 2004, 2009; Magee, 2003; Ghosh and Yamarik, 2004).

The attempts to solve the endogeneity problem for binary policy variables can be broadly classified into two subfields: one strand of the literature considers cross-sectional data and applies matching techniques, instrumental variables or control function approaches in a two-dimensional panel spanned by the bilateral nature of the trade variables; another strand of the literature relies on multi-dimensional panel-data techniques to solve the endogeneity problem.

For convenience, let us drop the time index t when considering cross-sectional techniques. Then, in the case of a binary policy variable, the estimation can be restated as a treatment evaluation problem (see Wooldridge, 2010 or Angrist and Pischke, 2009 for an introductory overview). For every (PTA) treated bilateral trade flow there are two potential outcomes, one with PTA assignment and a counterfactual (unobserved) one without PTA assignment, $y_{ij}(PTA_{ij} = 1)$ and $y_{ij}(PTA_{ij} = 0)$. Denoting all observable covariates by the vector x_{ij}, the so called average treatment effect (ATE) – the average effect of being a member of a PTA on bilateral exports –

---

a binary variable here since binary policy variables have been the main focus of the empirical literature. The endogeneity is rarely tackled in the case of continuous policy variables. For an exception see Egger and Erhardt (2016).

[13] Compare Baier and Bergstrand (2007) for an extensive discussion of potential endogeneity issues of PTAs.

can be stated as

$$ATE(x_{ij}) = \mathrm{E}\left[y_{ij}(PTA_{ij}=1|x_{ij}) - y_{ij}(PTA_{ij}=0)|x_{ij}, PTA_{ij}\right] . \tag{11.28}$$

For every bilateral trade flow, only one of the two potential outcomes is observable. Therefore, consistent estimation of (11.28) requires random assignment of a PTA such that it would actually not matter that different bilateral trade flows are compared. Since the assignment to a certain trade policy regime is, however, nonrandom, we have to make use of either instrumental variables estimation (see Egger et al. (2011)) or comparison group refinement (through weighting regression, matching, etc.; see Persson, 2001; Egger et al., 2008; Baier and Bergstrand, 2009).

The main requirement for a consistent estimation by instrumental variables is a set of suitable identifying instruments, z_{ij}, which is exogenous to the level of bilateral trade flows but is correlated with the probability of engaging in a PTA. Then, in a first step, the probability of engaging in a PTA can be estimated using standard binary choice models such as Probit or Logit

$$P(PTA_{ij}=1) = F(x_{ij}, z_{ij}, \eta, \delta), \tag{11.29}$$

where x_{ij} denotes the set of variables that determines the probability of forming a PTA jointly with the level of bilateral trade flows, z_{ij} are identifying instruments, and η and δ are the respective parameters on these variables. In a second step, one can use the fitted probability, $F(z_{ij}, x_{ij}, \hat{\delta}, \hat{\eta})$ in estimating the role of PTAs on trade flows consistently[14]

$$y_{ij} = \alpha_i + \alpha_j^* + \kappa \cdot \tilde{c}_{ij} + \beta(F(z_{ij}, x_{ij}, \hat{\delta}, \hat{\eta})) + \varepsilon_{ij}. \tag{11.30}$$

Clearly, instrumental variables methods hinge on the selection of suitable identifying instruments that do increase the likelihood of a PTA membership but do not at the same time impact trade flows. The set of identifying instruments typically used in the literature is a combination of economic variables such as relative factor endowment differences between the bilateral trade partners, the relative factor endowment differences between the bilateral trade partners and the rest of the world and measures of intra-industry trade, geographic variables such as a measure of remoteness of continental PTA partners and political variables such as democracy indices (see Baier and Bergstrand, 2002, 2007; Magee, 2003). However, at least some of these measures have previously been shown to be also relevant in determining trade flows eroding confidence in their suitability as instruments (Baier and Bergstrand, 2007; Egger et al., 2011). Egger et al. (2011) conduct a conventional test for overidentifying restrictions in order to determine suitable instruments and find historical variables such as past colonial relationship, a common colonizer indicator and a

[14] Baier and Bergstrand (2007) suggest an intermediate step in which the estimated probability of forming a PTA is regressed on the PTA dummy together with z_{ij} and x_{ij}. The predicted values of this intermediate stage are then used in the last stage. Note that the standard errors have to be corrected in the usual 2SLS manner (compare Wooldridge, 2010 or Greene, 2003).

dummy for having been part of the same country, as relevant in determining PTAs but being uncorrelated to trade flows.

Further attempts at solving the endogeneity problem using cross-sectional data sets have been conducted using matching techniques (see Persson, 2001; Egger et al., 2008; Baier and Bergstrand, 2009). The basic idea behind the matching approach is selecting a control group that is very similar to the treatment group in all observable variables, such that the control group's trade flows can be seen as the unobserved counterfactual outcome of the treatment group. A comparison of the mean effects between control and treatment group allows one to identify the treatment effect of the trade policy - assuming an appropriate choice of treatment and control group.[15] Control and treatment groups are "selected on observables", where theoretical models leading to gravity equations are of invaluable help in choosing those variables x_{ij}. The matching technique hinges on three core assumptions: first, *unconfoundedness*, which requires that conditional on the covariates x_{ij} the assignment of treatment ($PTA_{ij} = 1$) is independent of the (log) level of bilateral trade flows y_{ij}; second, the *stable unit treatment value assumption (SUTVA)*, whereby the PTA treatment assignment does not impact bilateral trade flows other than the treated country pair; third, *covariate balancing*, whereby the propensity of PTA membership is a meaningful compact comparison metric capturing similarity between treated and control observations in all columns of x_{ij}. Under these assumptions, a matching procedure combines those bilateral pairs that are most similar in their covariates but are subject to a different treatment and constructs artificial counterfactual outcomes for the – in the data – unobserved counterfactuals, $y^*_{ij}(PTA_{ij} = 1)$ or $y^*_{ij}(PTA_{ij} = 0)$, from the observed outcomes of the matched trade flows.[16] The effect of a PTA on bilateral trade flows can then be consistently estimated by

$$ATE(x_{ij}) = \mathrm{E}\left[y^*_{ij}(PTA_{ij} = 1) - y^*_{ij}(PTA_{ij} = 0)\right]. \tag{11.31}$$

Depending on the specific data situations, the three requirements for matching – unconfoundedness, SUTVA, and covariate balancing – might be difficult to achieve. Regarding unconfoundedness, there might be important covariates unobservable to the econometrician whose omission might violate unconfoundedness. Specifically, the fact that we cannot use fixed effects in order to account for multilateral resistance terms in the matching process, it becomes crucial to find proxies for those terms. Baier and Bergstrand (2009) use a Taylor-series expansion of the general equilibrium trade flows to generate reduced-form proxies for these multilateral resistance terms from exogenous determinants, such as distance, adjacency and common

[15] Compare Rosenbaum and Rubin (1983) and Abadie and Imbens (2006) for a detailed econometric treatment of the topic.

[16] There are many different ways of selecting the matched trade flows. Common metrics to compare covariates are propensity scores (Rosenbaum and Rubin, 1983) and Mahalanobis distance metrics (Rosenbaum and Rubin, 1985; Rosenbaum, 2002). Based on the similarity metric, trade flows are matched using e.g., *k*-nearest neighbors or employing an interval within the metric-space. Abadie and Imbens (2006) derive the large sample properties of a matching estimator that employs a Euclidian vector norm as a metric of similarity of covariates and matches *k*-nearest neighbors with replacement.

language dummies. Egger et al. (2008), in contrast, use a difference-in-difference approach on matched pairs in order to account for time-invariant confounders in a difference-in-differences matching framework. Regarding SUTVA, general equilibrium models of trade tell us that any change in bilateral trade costs will also impact all other trade flows through the multilateral resistance terms – in which case SUTVA would be violated. Baier and Bergstrand (2009) argue that, when looking at the Central American Common Market, which is comprised mainly of small and remote economies, expected general equilibrium effects should be small enough to be ignorable. Regarding covariate balancing, Egger and Tarlea (2016) demonstrate that the similarity of covariates between treated and untreated observations is rejected for most joint determinants of trade flows and PTA membership in most years. They propose a structural gravity approach with entropy balancing (which enforces covariate balancing). Their results suggest that the bias of the partial causal effect of PTAs on trade flows due to the lack of covariate balancing with matching is of the same order of magnitude as the bias when ignoring the endogeneity of PTA membership altogether.

In recent years, it has become customary to exploit the time-dimension of data on bilateral trade flows to solve the endogeneity problem of endogenous policy variables (see Baier and Bergstrand, 2007; Egger et al., 2008; Eicher et al., 2012; Bergstrand et al., 2015; Anderson and Yotov, 2016). Assuming that all unobservable joint determinants of PTAs and trade flows are time-invariant, a standard gravity equation extended to allow for bilateral (time-invariant) fixed effects, γ_{ij}, as in Baltagi et al. (2003), will yield consistent results (Baier and Bergstrand, 2007)

$$\log(y_{ijt}) = \alpha_{it} + \alpha^*_{jt} + \gamma_{ij} + \beta PTA_{ijt} + \varepsilon_{ijt}. \tag{11.32}$$

Bergstrand et al. (2015) argue that the role of time-varying bilateral components might still be relevant. Especially in the context of modern trade models featuring heterogeneous firms, the theory-consistent gravity equation includes fixed costs of entering a foreign market that might well be varying over time and at the same time be correlated with the decision to form a PTA, hence introducing a potential endogeneity bias that is not solved by (11.32).

Since only international trade flows are subject to these fixed costs of exporting, Bergstrand et al. (2015) include domestic trade flows, (y_{iit}), in their estimation and create a time-varying dummy $INTER_{ijt}$ that indicates non-domestic trade flows interacted with time dummies. By this methodology, they aim at controlling for any time-varying changes in fixed costs of exporting that are potentially correlated with the choice of forming a PTA. However, this procedure only provides a remedy for the problem if the self-selection bias is invariant across all country pairs (which it should not be according to economic theory).

11.3.6 Ratio Estimators

Most approaches in the gravity literature rely on the estimation of a substantial number of (multi-dimensional) fixed effects. While the fixed effects are important to account for endogeneity and potential omitted variables bias, two problems arise. First, the large number of parameters to be estimated (explicitly or implicitly) might considerably reduce the efficiency of the estimated parameters of interest. Second, one might be interested in isolating effects of covariates which are fully collinear with the fixed effects.

Instead of estimating a standard gravity model of bilateral trade flows such as (11.9), some authors have estimated normalized versions of the latter in order to avoid having to estimate numerous (multi-dimensional) fixed effects. Eaton and Kortum (2002),[17] for example, normalized bilateral trade flows by the importer's domestic sales[18]

$$\frac{Y_{ij}}{Y_{jj}} = \frac{Y_i}{\mathscr{P}_i}\frac{\mathscr{P}_j}{Y_j}(C_{ij})^{\kappa}. \tag{11.33}$$

Rewriting (11.33) when using fixed effects yields

$$y_{ij} - y_{jj} = \alpha_i - \alpha_j + \kappa \cdot c_{ij} + \varepsilon_{ij}. \tag{11.34}$$

In contrast to the standard gravity equation, the number of fixed effects is reduced by half, since the fixed effect α_i is the same if country i is importing or exporting. Note that (11.34) allows for including additional fixed effects in c_{ij}. For instance, Eaton and Kortum (2002) allow for an importer-fixed effect in the trade costs which might capture non-tariff trade barriers. This fixed effect can be isolated separately since it appears only for imports but not for consumption from domestic firms. Waugh (2010) and Simonovska and Waugh (2014) use a similar specification but isolate an exporter-specific component of trade costs instead.[19]

A way to obtain a country-(time-)fixed-effects-free formulation of the gravity equation is to take ratios of ratios (Anderson and Marcouiller, 2002; Romalis, 2007; Hallak, 2006; Martin et al., 2008; Head et al., 2010).[20] Consider the exports from i to j, as well as the exports of m to k and rewrite them in multiplicative form

$$Y_{ij} \times Y_{mk} = \left(\frac{Y_i}{\mathscr{P}_i}\frac{E_j}{\mathscr{P}_j^*}(C_{ij})^{\kappa}\right) \times \left(\frac{Y_m}{\mathscr{P}_m}\frac{E_k}{\mathscr{P}_k^*}(C_{mk})^{\kappa}\right). \tag{11.35}$$

[17] Eaton and Kortum (2002) substitute for prices to obtain a measure of "competitiveness". The ratio approach, however, remains the same as described here.

[18] Note that we – as is common in the literature – assume that domestic trade costs are zero so that $C_{jj} = 1$.

[19] Clearly, such an approach allows the identification of only one fixed effect contained in trade costs (either importer- or exporter-specific), and it generally permits modelling an asymmetry in trade costs for ij- versus ji-trade.

[20] The methodology is also known as the "tetrads" method (Head et al., 2010).

Repeating that exercise for the exports from i to k as well as the exports of m to j yields

$$Y_{ik} \times Y_{mj} = \left(\frac{Y_i}{\mathscr{P}_i} \frac{E_k}{\mathscr{P}_k^*} (C_{ik})^{\kappa} \right) \times \left(\frac{Y_m}{\mathscr{P}_m} \frac{E_j}{\mathscr{P}_j^*} (C_{mj})^{\kappa} \right) . \tag{11.36}$$

Dividing (11.36) by (11.35) and rearranging terms yields

$$\frac{Y_{ij} \times Y_{mk}}{Y_{ik} \times Y_{mj}} = \frac{(C_{ij})^{\kappa} \times (C_{mk})^{\kappa}}{(C_{ik})^{\kappa} \times (C_{mj})^{\kappa}}. \tag{11.37}$$

Taking logs, equation (11.37) can be easily estimated without relying on any country-(time-)-specific fixed effects. However, allowing for the full set of combinations dramatically increases the set of observations. Most authors rely on reference countries to reduce the set of observations. Clearly, estimating equation (11.37) may lead to severely biased parameters in the presence of self-selection into positive trade flows.[21] To accommodate zeros, Charbonneau (2012) and Egger and Staub (2016) apply a consistent GMM estimator to (11.37). Rewriting (11.37) and taking conditional expectations yields

$$\mathrm{E}\left[Y_{ij} \times Y_{mk} - Y_{ik} \times Y_{mj} * e^{\kappa * (c_{ij} + c_{mk} - c_{ik} - c_{mj})} | c_{ij}, c_{mk}, c_{ik}, c_{mj}, \alpha_i, \alpha_m, \alpha_j^*, \alpha_k^* \right] = 0. \tag{11.38}$$

The respective unconditional moment conditions read

$$\mathrm{E}\left[\left(Y_{ij} \times Y_{mk} - Y_{ik} \times Y_{mj} * e^{\kappa * (c_{ij} + c_{mk} - c_{ik} - c_{mj})} \right) (c_{ij} + c_{mk} - c_{ik} - c_{mj}) \right] = 0. \tag{11.39}$$

Instead of using four different countries to construct ratios of ratios as in (11.37), Caliendo and Parro (2015) use three countries to construct the product of trade flows from i to j, from j to k, and from k to i, and divide it by the product of the trade flows in the opposite direction, respectively

$$\frac{Y_{ij} \times Y_{jk} \times Y_{ki}}{Y_{ji} \times Y_{kj} \times Y_{ik}} = \frac{(C_{ij})^{\kappa} \times (C_{jk})^{\kappa} \times (C_{ki})^{\kappa}}{(C_{ji})^{\kappa} \times (C_{kj})^{\kappa} \times (C_{ik})^{\kappa}}. \tag{11.40}$$

As before, importer-(time-) and exporter (time) specific components of trade flows (and trade costs) cancel out with this approach. Additionally, using (11.40), all symmetric bilateral trade costs and symmetric trade-cost variables such as distance or PTA membership cancel out.

However, there are two general drawbacks with ratio estimators. First, the degrees of freedom have to be adjusted properly in order to avoid deflating standard errors on the parameter estimates of interest. Second, the mechanical dependence structure which emerges due to the repeated occurrence of specific trade flows in the various ratios must be accounted for, e.g., by using appropriate cluster-robust standard errors in order to avoid biased standard errors and test statistics.

[21] Compare Sect. 11.3.2 for an outline of this issue.

11.4 Conclusion

The econometric analysis of data on bilateral trade flows is exciting, since these data are easily available for many country pairs and years, and the economic theory establishing sound foundations for the model specification is well established and still thriving. An interesting aspect to the econometrician is the dimensionality of the data, where trade flows are at least double or triple, if not quadruple-indexed (by exporter and importer, and eventually also by time and product).

This high dimensionality of the data permits a rich treatment of fixed effects, of dynamics, of cross-sectional interdependence, and other issues, and eventually calls for methods which are not readily developed. The present chapter provides an overview of the state of the art of methods dealing with these issues. Future work on the econometrics of gravity models may fruitfully address combinations of these problems. For instance, the exponential-family applications of gravity models tend to ignore dynamics, as well as cross-sectional dependence. The dynamic models tend to ignore cross-sectional dependence. Ratio estimators tend to ignore interdependence and the complex non-independent structure of the transformed disturbances due to repeated observation of units.

References

Abadie, A. and Imbens, G. W. (2006). Large sample properties of matching estimators for average treatment effects. *Econometrica*, 74(1):235–267.

Anderson, J. E. (1979). A theoretical foundation for the gravity equation. *American Economic Review*, 69(1):106–16.

Anderson, J. E. and Marcouiller, D. (2002). Insecurity and the pattern of trade: An empirical investigation. *The Review of Economics and Statistics*, 84(2):342–352.

Anderson, J. E. and van Wincoop, E. (2003). Gravity with gravitas: A solution to the border puzzle. *American Economic Review*, 93(1):170–192.

Anderson, J. E. and Yotov, Y. V. (2016). Terms of trade and global efficiency effects of free trade agreements, 1990-2002. *Journal of International Economics*, 99(C):279–298.

Angrist, J. D. and Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Number 8769 in Economics Books. Princeton University Press.

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(2):277–297.

Arkolakis, C., Costinot, A., and Rodriguez-Clare, A. (2012). New trade models, same old gains? *American Economic Review*, 102(1):94–130.

Armenter, R. and Koren, M. (2014). A balls-and-bins model of trade. *American Economic Review*, 104(7):2127–51.

Badinger, H. and Egger, P. (2015). Fixed effects and random effects estimation of higher-order spatial autoregressive models with spatial autoregressive and heteroscedastic disturbances. *Spatial Economic Analysis*, 10(1):11–35.

Baier, S. and Bergstrand, J. (2002). On the endogeneity of international trade flows and free trade agreements. Unpublished Manuscript. University of Notre Dame.

Baier, S. L. and Bergstrand, J. H. (2004). Economic determinants of free trade agreements. *Journal of International Economics*, 64(1):29–63.

Baier, S. L. and Bergstrand, J. H. (2007). Do free trade agreements actually increase members' international trade? *Journal of International Economics*, 71(1):72–95.

Baier, S. L. and Bergstrand, J. H. (2009). Estimating the effects of free trade agreements on international trade flows using matching econometrics. *Journal of International Economics*, 77(1):63–76.

Baldwin, R. and Taglioni, D. (2006). Gravity for dummies and dummies for gravity equations. NBER Working Papers 12516, National Bureau of Economic Research, Inc.

Baltagi, B. (2013). *Econometric Analysis of Panel Data*. John Wiley & Sons, Ltd., 5th edition edition.

Baltagi, B. H., Egger, P., and Pfaffermayr, M. (2003). A generalized design for bilateral trade flow models. *Economics Letters*, 80(3):391–397.

Baltagi, B. H., Egger, P., and Pfaffermayr, M. (2007). Estimating models of complex FDI: Are there third-country effects? *Journal of Econometrics*, 140(1):260–281.

Baltagi, B. H., Egger, P., and Pfaffermayr, M. (2015). Panel data gravity models of international trade. In *The Oxford Handbook of Panel Data*. Oxford University Press.

Beenstock, M. and Felsenstein, D. (2012). Spatial gravity. Unpublished Manuscript. Hebrew University of Jerusalem.

Beenstock, M. and Felsenstein, D. (2015). Spatial dependence in the econometrics of gravity modeling. In Patuelli, R. and Arbia, G., editors, *Spatial Econometric Interaction Modelling*. Springer, Berlin.

Behrens, K., Ertur, C., and Koch, W. (2012). Dual gravity: Using spatial econometrics to control for multilateral resistance. *Journal of Applied Econometrics*, 27(5):773–794.

Bergstrand, J. H. (1989). The generalized gravity equation, monopolistic competition, and the factor-proportions theory in international trade. *The Review of Economics and Statistics*, 71(1):143–53.

Bergstrand, J. H. (1990). The heckscher-ohlin-samuelson model, the linder hypothesis and the determinants of bilateral intra-industry trade. *Economic Journal*, 100(403):1216–29.

Bergstrand, J. H., Larch, M., and Yotov, Y. V. (2015). Economic integration agreements, border effects, and distance elasticities in the gravity equation. *European Economic Review*, 78(C):307–327.

Blonigen, B. A., Davies, R. B., Waddell, G. R., and Naughton, H. T. (2007). FDI in space: Spatial autoregressive relationships in foreign direct investment. *European Economic Review*, 51(5):1303–1325.

Bun, M. J. G. and Klaassen, F. J. G. M. (2002). The importance of dynamics in panel gravity models of trade. Unpublished Manuscript. University of Amsterdam.

Caliendo, L. and Parro, F. (2015). Estimates of the trade and welfare effects of NAFTA. *Review of Economic Studies*, 82(1):1–44.

Charbonneau, K. B. (2012). Multiple fixed effects in nonlinear panel data models – theory and evidence. Unpublished Manuscript. Princeton University.

Cheng, I.-H. and Wall, H. J. (2005). Controlling for heterogeneity in gravity models of trade and integration. *Review*, 87(1):49–63.

Coughlin, C. C. and Segev, E. (2000). Foreign direct investment in China: A spatial econometric study. *The World Economy*, 23(1):1–23.

Cragg, J. G. (1971). Some statistical models for limited dependent variables with application to the demand for durable goods. *Econometrica*, pages 829–844.

Deardorff, A. (1998). Determinants of bilateral trade: Does gravity work in a neoclassical world? In *The Regionalization of the World Economy*, NBER Chapters, pages 7–32. National Bureau of Economic Research, Inc.

Dekle, R., Eaton, J., and Kortum, S. (2007). Unbalanced trade. *American Economic Review*, 97(2):351–355.

Disdier, A.-C. and Head, K. (2008). The puzzling persistence of the distance effect on bilateral trade. *The Review of Economics and Statistics*, 90(1):37–48.

Duan, N., Manning, W. G., Morris, C. N., and Newhouse, J. P. (1984). Choosing between the sample-selection model and the multi-part model. *Journal of Business & Economic Statistics*, 2(3):283–289.

Eaton, J. and Kortum, S. (2002). Technology, geography, and trade. *Econometrica*, 70(5):1741–1779.

Egger, H., Egger, P., and Greenaway, D. (2008). The trade structure effects of endogenous regional trade agreements. *Journal of International Economics*, 74(2):278–298.

Egger, P. (2000). A note on the proper econometric specification of the gravity equation. *Economics Letters*, 66(1):25–31.

Egger, P. (2001). European exports and outward foreign direct investment: A dynamic panel data approach. *Review of World Economics (Weltwirtschaftliches Archiv)*, 137(3):427–449.

Egger, P. and Erhardt, K. (2016). Heterogeneous effects of tariff and nontariff policy barriers in quantitative general equilibrium. Unpublished Manuscript. ETH Zurich.

Egger, P., Larch, M., Staub, K. E., and Winkelmann, R. (2011). The trade effects of endogenous preferential trade agreements. *American Economic Journal: Economic Policy*, 3(3):113–43.

Egger, P. and Pfaffermayr, M. (2003). The proper panel econometric specification of the gravity equation: A three-way model with bilateral interaction effects. *Empirical Economics*, 28(3):571–580.

Egger, P. and Pfaffermayr, M. (2011). Structural estimation of gravity models with path-dependent market entry. CEPR Discussion Papers.

Egger, P. and Pfaffermayr, M. (2016). A generalized spatial error components model for gravity equations. *Empirical Economics*, 50(1):177–195.

Egger, P. and Staub, K. (2016). GLM estimation of trade gravity models with fixed effects. *Empirical Economics*, 50(1):137–175.

Egger, P. and Tarlea, F. (2016). Comparing apples to apples: Estimating consistent partial effects of preferential economic integration agreements. Unpublished Manuscript. ETH Zurich.

Egger, P. H. and Lassmann, A. (2012). The language effect in international trade: A meta-analysis. *Economics Letters*, 116(2):221–224.

Egger, P. H. and Nigai, S. (2015). Structural gravity with dummies only: Constrained ANOVA-type estimation of gravity models. *Journal of International Economics*, 97(1):86–99.

Eicher, T. S., Henn, C., and Papageorgiou, C. (2012). Trade creation and diversion revisited: Accounting for model uncertainty and natural trading partner effects. *Journal of Applied Econometrics*, 27(2):296–321.

Felbermayr, G. J. and Kohler, W. (2006). Exploring the intensive and extensive margins of world trade. *Review of World Economics (Weltwirtschaftliches Archiv)*, 142(4):642–674.

Frankel, J. A. (1997). *Regional Trading Blocs in the World Economic System*. Number 72 in Peterson Institute Press: All Books. Peterson Institute for International Economics.

Ghosh, S. and Yamarik, S. (2004). Are regional trading arrangements trade creating? an application of extreme bounds anslysis. *Journal of International Economics*, 63(2):369–395.

Glick, R. and Rose, A. K. (2002). Does a currency union affect trade? the time-series evidence. *European Economic Review*, 46(6):1125–1151.

Greene, W. H. (2003). *Econometric Analysis*. Prentice Hall.

Hallak, J. C. (2006). Product quality and the direction of trade. *Journal of International Economics*, 68(1):238–265.

Head, K. and Mayer, T. (2014). *Gravity Equations: Workhorse, Toolkit, and Cookbook*, volume 4 of *Handbook of International Economics*, pages 131–195. Elsevier.

Head, K., Mayer, T., and Ries, J. (2010). The erosion of colonial trade linkages after independence. *Journal of International Economics*, 81(1):1–14.

Helpman, E., Melitz, M., and Rubinstein, Y. (2008). Estimating trade flows: Trading partners and trading volumes. *The Quarterly Journal of Economics*, 123(2):441–487.

Hsiao, C. (2015). *Analysis of Panel Data*. Cambridge Books. Cambridge University Press.

Hummels, D. (1999). Toward a geography of trade costs. GTAP Working Papers 1162, Center for Global Trade Analysis, Department of Agricultural Economics, Purdue University.

Kapoor, M., Kelejian, H. H., and Prucha, I. R. (2007). Panel data models with spatially correlated error components. *Journal of Econometrics*, 140(1):97–130.

Kelejian, H. H. and Prucha, I. R. (1999). A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review*, 40(2):509–533.

Kelejian, H. H. and Prucha, I. R. (2010). Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics*, 157(1):53–67.

Lee, L.-F. (2004). Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica*, 72(6):1899–1925.

LeSage, J. P. and Fischer, M. M. (2010). Spatial econometric methods for modeling origin-destination flows. In Fischer, M. M. and Getis, A., editors, *Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications*, pages 409–433. Springer Berlin Heidelberg.

LeSage, J. P. and Pace, R. K. (2008). Spatial econometric modeling of origin-destination flows. *Journal of Regional Science*, 48(5):941–967.

LeSage, J. P. and Thomas-Agnan, C. (2015). Interpreting spatial econometric origin-destination flow models. *Journal of Regional Science*, 55(2):188–208.

Magee, C. S. (2003). Endogenous preferential trade agreements: An empirical analysis. *Contributions in Economic Analysis & Policy*, 2(1).

Martin, P., Mayer, T., and Thoenig, M. (2008). Make trade not war? *Review of Economic Studies*, 75(3):865–900.

Matyas, L. (1997). Proper econometric specification of the gravity model. *The World Economy*, 20(3):363–368.

Matyas, L. (1998). The gravity model: Some econometric considerations. *The World Economy*, 21(3):397–401.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*, volume 37. CRC press.

Melitz, M. J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica*, 71(6):1695–1725.

Nelder, J. A. and Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society. Series A (General)*, 135(3):370–384.

Nickell, S. J. (1981). Biases in dynamic models with fixed effects. *Econometrica*, 49(6):1417–26.

Nuroglu, E. and Kunst, R. (2014). Competing specifications of the gravity equation: A three-way model, bilateral interaction effects, or a dynamic gravity model with time-varying country effects? *Empirical Economics*, 46(2):733–741.

Olivero, M. P. and Yotov, Y. V. (2012). Dynamic gravity: Endogenous country size and asset accumulation. *Canadian Journal of Economics*, 45(1):64–92.

Persson, T. (2001). Currency unions and trade: how large is the treatment effect? *Economic Policy*, 16(33):433–462.

Pöyhönen, P. (1963). A tentative model for the volume of trade between countries. *Weltwirtschaftliches Archiv*, pages 93–100.

Romalis, J. (2007). NAFTA's and CUSFTA's impact on international trade. *The Review of Economics and Statistics*, 89(3):416–435.

Rose, A. K. and van Wincoop, E. (2001). National money as a barrier to international trade: The real case for currency union. *American Economic Review*, 91(2):386–390.

Rosenbaum, P. R. (2002). *Observational Studies*, pages 1–17. Springer New York, New York, NY.

Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55.

Rosenbaum, P. R. and Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, 39(1):33–38.

Santos Silva, J. M. C. and Tenreyro, S. (2006). The log of gravity. *The Review of Economics and Statistics*, 88(4):641–658.

Santos Silva, J. M. C. and Tenreyro, S. (2015). Trading partners and trading volumes: Implementing the helpmanmelitzrubinstein model empirically. *Oxford Bulletin of Economics and Statistics*, 77(1):93–105.

Shin, Y. and Serlenga, L. (2007). Gravity models of intra-EU trade: Application of the CCEP-HT estimation in heterogeneous panels with unobserved common time-specific factors. *Journal of Applied Econometrics*, 22(2):361–381.

Simonovska, I. and Waugh, M. E. (2014). The elasticity of trade: Estimates and evidence. *Journal of International Economics*, 92(1):34–50.

Tinbergen, J. (1962). *Shaping the World Economy*. New York: The Twentiest Century Fund.

Waugh, M. E. (2010). International trade and income differences. *American Economic Review*, 100(5):2093–2124.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*, volume 1 of *MIT Press Books*. The MIT Press.

Chapter 12
Modelling Housing Using Multi-dimensional Panel Data

Badi H. Baltagi and Georges Bresson

Abstract This chapter surveys housing models using multi-dimensional panels. While there is a vast literature on housing models using two-dimensional panel data, there are few papers using multi-dimensional panels. This chapter focuses on housing models, residential mobility and location choice models derived from discrete choice theory, utilizing multi-dimensional panels. Examples include nested or hierarchical error components models, where a house is located in a street, within a block, within a city, within a county, etc. This chapter introduces some basic concepts of utility functions and discrete choice models used for hedonic functions, and residential mobility and location choices. Then it surveys some significant papers on multi-dimensional models of residential mobility and location choice. The paper concludes by surveying a few papers on dynamic housing models. It shows that both spatial and temporal dimensions in dynamic systems should be included for hedonic housing models and discrete models of residential location in a multi-dimensional framework. However, the inclusion of these multiple dimensions greatly complicates the specification and modeling of such systems.

12.1 Introduction

This chapter surveys housing models using multi-dimensional panels. For more than a decade, a huge literature within the New Economic Geography has emerged to study the causes of temporal and spatial variations in house prices, residential mobility and location choice. These are major household decisions connected with many activities and travel aspects of households' lives. These concepts have been widely researched in various fields including economics, sociology, geography, urban plan-

Badi H. Baltagi
Syracuse University, New York, USA

Georges Bresson
Université Paris II, Paris, France

L. Matyas (ed.), *The Econometrics of Multi-dimensional Panels*, Advanced Studies in Theoretical and Applied Econometrics 50, DOI 10.1007/978-3-319-60783-2_12

ning, transportation, etc. Location choices and housing investments are inherently dynamic decisions. Moreover, the choice for a household to locate in a given area is a complex decision that is influenced by, among other things, the structural elements of a dwelling, as well as the property's spatial relationship to certain amenities. One source of spatial heterogeneity comes from the natural hierarchical and nested structure of the locations of houses: whether they located in a street, within a block, within a city, within a county, within a region, etc. There is a vast literature on such topics mainly using time series and longitudinal (two-dimensional (2D)) data, but only a few papers using a multi-dimensional (three-dimensional (3D) and more) framework. In this chapter, we will focus on housing models, residential mobility and location choice models derived from discrete choice theory, focusing on examples that use multi-dimensional panels.

Baltagi et al. (2014), for example, focus on the estimation of UK house prices in which spatio-temporal variations in house prices are driven by supply and demand conditions, with spatial effects coming from two distinct sources. One is the direct dependence of house prices in a given locality on house prices in nearby localities. The second source of spatial heterogeneity comes from the presence of hierarchical error components which represent the impact of local (district) effects embedded within wider (county) effects. The panel data includes 353 local authority districts in England over the period 2000–2007. This is done using instrumental variable estimation. Another example is Baltagi et al. (2015), who estimate a hedonic housing model based on flats sold in the city of Paris over the period 1990–2003. This is done using maximum likelihood estimation, taking into account the nested structure of the data. Paris is historically divided into 20 *arrondissements*, each divided into four *quartiers* (quarters), which in turn contain between 15 and 169 blocks (*îlot*, in French) per *quartier*.

In Sect. 12.2, we introduce some basic concepts of utility functions and discrete choice models used for hedonic functions, residential mobility and location choices. Section 12.3 deals with multi-dimensional models of housing hedonic price functions, their estimation methods and some results. Section 12.4 analyses some multi-dimensional models of residential mobility and location choice. Section 12.5 focuses on multi-dimensional dynamic models of housing models and Sect. 12.6 concludes.

12.2 Discrete Choice Models and Hedonic Price Functions: A Quick Overview

The pioneering work by Daniel McFadden on location choice is an obvious starting point for a discussion on housing models. One generally considers a household i who chooses to locate in neighborhood j and buy house type k. A standard random utility model (see, e.g., Holmes and Sieg, 2014) assumes that the indirect utility of household i for location j and house k is given by

$$u_{ijk} = X_j'\beta + Z_k'\gamma + (y_i - p_{jk})\alpha + \varepsilon_{ijk} = f_{ijk}(.) + \varepsilon_{ijk}\,, \tag{12.1}$$

where X_j is a vector of observed characteristics of location j, Z_k is a vector of observed characteristics for house k, y_i is the household income and p_{jk} is the price of housing type k in location j. Each household chooses the neighborhood-housing pair that maximizes utility. Under the assumption that the error terms ε_{ijk} are independent and identically distributed (*i.i.d.*) across i, j and k and follow a type I extreme value distribution, McFadden (1973) (see also McFadden, 1974, 1978), derived the well-known conditional logit choice probabilities

$$Pr\left[d_{ijk} = 1\right] = \frac{\exp\left(f_{ijk}(.)\right)}{\sum_{j=1}^{J}\sum_{k=1}^{K}\exp\left(f_{ijk}(.)\right)}\,, \tag{12.2}$$

where $d_{ijk} = 1$ if household i has chosen neighborhood j and house type k and zero otherwise. However, the independence of irrelevant alternatives (IIA) property of this model is unattractive. McFadden (1978) proposed the use of a generalized extreme value distribution for the error terms, which gives rise to the nested logit model and allows one to relax the assumption that idiosyncratic tastes are independent across locations and houses. However, we need to choose the nesting structure before estimation, mainly if the nested structure is not natural and if we do not have knowledge about the neighborhood structure. One solution is to use random coefficients β_i, γ_i and α_i instead of fixed coefficients β, γ and α. Estimation with random coefficients is challenging and needs the use of simulation-based estimators (SBE) (see Newey and McFadden, 1974 or Judd, 1998).

Moreover, Bayesian estimators are also well suited for the estimation of discrete choice models with random coefficients. One application of such a model with SBE has been done by Hastings et al. (2006), who study the effects of open enrollment policies under a particular parent choice mechanism, sorting households among schools within the Mecklenburg Charlotte school district, North Carolina. Bajari and Kahn (2005) used Bayesian methods to study housing demand explaining racial segregation in cities.

Demand estimation has also focused on the role of unobserved neighborhood characteristics or housing quality ζ_j. In this case, the indirect utility function is written as

$$u_{ijk} = X_j'\beta + Z_k'\gamma + (y_i - p_{jk})\alpha + \zeta_j + \varepsilon_{ijk}\,. \tag{12.3}$$

Unobserved neighborhood characteristics can be recovered by matching the observed market shares of community j. Then, the remaining parameters can be estimated by a generalized method of moments (GMM) estimator using instrumental variables (IV) to deal with the correlation between housing price p_{jk} and unobserved neighborhood characteristics or housing quality ζ_j. Bayer et al. (2007), using two-dimensional (2D) panel data, estimate household preferences for school and neighborhood attributes in the presence of sorting. The model embeds a boundary discontinuity design in a heterogeneous residential choice model, addressing the endogeneity of the school and neighborhood characteristics. Their application concerns a restricted-access version of the 1990 U.S. Census, that links detailed characteristics

for nearly a quarter of a million households and their houses in the San Francisco Bay Area with their precise residential locations. Bayer et al. (2016), using three-dimensional panel data (3D), develop a dynamic model of neighborhood choice (see Sect. 12.5). They capture observed and unobserved preference heterogeneity across households and locations of housing transactions in the San Francisco Bay Area from 1994 to 2004.

We now turn to hedonic measures with a strong theoretical grounding (see, among others, Griliches, 1971; Rosen, 1974; Nelson, 1977; Blomquist and Worley, 1981, 1982 among others). In addition, we show the use of regression techniques to control for compositional and quality change (see, e.g., Witte et al., 1979; Brown and Rosen, 1982; Meese and Wallace, 1997, to mention a few). The hedonic pricing method is based on the fact that prices of goods (in our case, houses) in a market are affected by their characteristics. This method estimates the value of a commodity based on people's willingness to pay for the commodity as and when its characteristics change. In real estate economics, hedonic pricing is used to adjust for the problems associated with looking for a dwelling that is as heterogeneous as buildings. The hedonic pricing function, which explains the price of a house, will be affected by, among other things, the structural characteristics of the house, and neighborhood and environmental characteristics.

Since the seminal work of Rosen (1974), we have generally used a two-stage procedure for estimating the hedonic price function of the dwelling and for the recovery of marginal willingness to pay functions of heterogeneous individuals for the characteristics of differentiated products. Basically, hedonic models of housing price relate the price (or the logarithm of the price per square meter) to, among other things, the characteristics of the dwellings $p_{jk} = f(Z'_k, ...)$. The price gradient associated with this hedonic price function $\partial p_{jk}/\partial Z_{kl}$ denotes the implicit price of the amenity Z_{kl} (number of rooms, quality of air, etc.). The second stage of Rosen's procedure seeks to recover the coefficients of demand (or marginal willingness to pay) and supply (or marginal willingness to accept) functions for the attribute Z_{kl} from the first-order conditions of the equilibrium relationships: $\partial p_{jk}/\partial Z_{kl} = f_d(Z_k, B_k)$ for demand and $\partial p_{jk}/\partial Z_{kl} = f_s(Z_k, S_k)$ for supply, where B_k and S_k represent attributes of the buyer and seller of house k. Bartik (1987) and Epple (1987) have described a source of endogeneity in the second stage of Rosen's procedure that is difficult to overcome without exclusion restriction arguments or the use of IV methods. This has led researchers to avoid altogether the estimation of marginal willingness to pay functions, relying instead on the first-stage hedonic price function and limiting the analysis to the evaluation of marginal changes in amenities (see Gayer et al., 2000; Bishop and Timmins, 2011 to mention a few).

In some studies, dwellings were assumed to be stratified into blocks or communities j, where prices are homogeneous and price trends are roughly parallel. Ideally, a model could be estimated in each neighborhood and the elementary geographic zones could be very small sub-markets. In this case, each model is estimated in a particular block, all variables are *de facto* interacted with the block. Thus, spatial location is not without consequences and hedonic housing price models should incorporate spatial effects. In the econometric literature, spatial effects may result

from spatial dependence or from spatial heterogeneity. Spatial dependence means that observations at location j depend on other observations at locations $l \neq j$. Spatial heterogeneity refers to variation in relationships over space and, more precisely, over every point in space. The distinction comes from the structure of the dependence, which can be related to location and distance, both in a geographic space, as well as in a more general economic or social network space (see Anselin, 2001; Anselin et al., 2008).

For spatial effects in real estate, many housing models have been estimated in a 2D framework on panel data with two indexes j and t generally for location and time associated with spatial weight matrices (see for instance Baltagi and Bresson, 2011; Bresson and Hsiao, 2011; Fingleton, 2008; Glaeser, 2008; Holly et al., 2010, to mention a few). However, very few models have been developed in a three-dimensional, or higher dimensional panel data setting. In the next section, we present some of these models and their associated results for these multi-dimensional frameworks.

12.3 Multi-dimensional Models of Housing Hedonic Price Functions: Some Examples

Baltagi et al. (2015) estimate a hedonic housing model based on flats sold in the city of Paris over the period 1990–2003. This is done using maximum likelihood estimation, taking into account the nested structure of the data. Paris is historically divided into 20 *arrondissements*, each divided into four *quartiers* (quarters), which in turn contain between 15 and 169 blocks (*îlot*, in French) per *quartier*. The data set used is an unbalanced pseudo-panel data containing 156,896 transactions. The real estate literature emphasizes the importance of neighborhoods in determining the value of a house or a flat. While one can try and include as many as possible of the neighborhood characteristics in the regression to capture these effects, most attempts may fall short because many neighborhood characteristics are not observed, as in our case. One simple method of capturing the effect of neighbors' prices used by Baltagi et al. (2015) is to estimate a spatial lag regression equation with time-varying coefficients:

$$p_{taqif} = \lambda_t \tilde{p}_{taqif} + Z_{taqif}\beta + \varepsilon_{taqif}\ ,\ |\ \lambda_t\ | < 1\,, \tag{12.4}$$

where $t = 1,...,T$ for years, $a = 1,...,N$ for *arrondissements*, $q = 1,...,Q_{ta}$ for *quartiers*, $i = 1,...,M_{taq}$ for *îlots* and $f = 1,...,F_{taqi}$ for flats. p is the transaction price (in logs) for flat f, in *îlot* i nested in *quartier* q, which in turn is nested in *arrondissement* a at time t. Z_{taqif} denotes the vector of K explanatory variables describing the characteristics for this flat (surface in m^2, count data as number of rooms, bedrooms, bathrooms, garage plots, and dummy variables such as balcony, whether it is located in a street, boulevard, avenue, or place, period of construction (<1850, 1850–1913, ...,1981–2003), etc). This unbalanced panel is made up of $N = 20$ top-level *arrondissements*, each containing Q_{ta} second-level *quartiers*. The second-level *quartiers* in turn contain M_{taq} third-level *îlots*, which contain the

innermost F_{taqi} observations on flats. The number of observations in the higher level groups are $F_{taq} = \sum_{i=1}^{M_{taq}} F_{taqi}$ and $F_{ta} = \sum_{q=1}^{Q_{ta}} F_{taq}$. The total number of observations is $H = \sum_{t=1}^{T} \sum_{a=1}^{N} F_{ta}$. The number of top-level groups is NT, the number of second-level groups is $L = \sum_{t=1}^{T} \sum_{a=1}^{N} Q_{ta}$, and the number of bottom-level groups is $G = \sum_{t=1}^{T} \sum_{a=1}^{N} \sum_{q=1}^{Q_{ta}} M_{taq}$. Thus, we have a five-dimensional pseudo-panel data structure. The spatial lag coefficient λ_t may be time varying or constant over time and the spatial lag variable $\tilde{p}_{taqif}$ is defined as

$$\tilde{p}_{taqif} = \sum_{a=1}^{N} \sum_{q=1}^{Q_{ta}} \sum_{i=1}^{M_{taq}} \sum_{p=1}^{F_{taqi}} w_{taqip} p_{taqip} \,, \tag{12.5}$$

where w_{taqip} denotes the elements of the spatial weights matrices W_t, which vary with t. Elements on the diagonal of W_t are set to zero, while the off-diagonal elements define the connexion (contiguity or distances) between dwellings. There are at least two reasons why a positive spatial correlation may exist. First, dwellings in a neighborhood tend to have similar structural characteristics and second, dwellings in a neighborhood share the same location amenities (see Basu and Thibodeau, 1988). However, many of the price determining factors shared by neighborhoods are difficult to explicitly explain, but these "omitted" factors are contained in the neighborhood prices. For each year, Baltagi et al. (2015), using the "Delaunay triangle algorithm", define first-order contiguity matrices W_t for the nearest neighbors (i.e., from 10 to 140 nearest sold flats). According to the nested structure, the disturbance term is given by

$$\varepsilon_{taqif} = \delta_{ta} + \mu_{taq} + \nu_{taqi} + u_{taqif} \,, \tag{12.6}$$

where δ_{ta} is the *arrondissement* effect, μ_{taq} is the *quartier* effect naturally nested in the respective *arrondissement* and ν_{taqi} is the *îlot* effect naturally nested in the respective *quartier*. These could be fixed or random. The remainder disturbance term for the particular flat is random $u_{taqif} \sim iiN(0, \sigma_u^2)$. For the random specification, we assume that $\delta_{ta} \sim iiN(0, \sigma_\delta^2)$, $\mu_{taq} \sim iiN(0, \sigma_\mu^2)$ and $\nu_{taqi} \sim iiN(0, \sigma_\nu^2)$.

Following Antweiler (2001), Baltagi et al. (2015) use block-diagonal matrices of size $(H \times H)$ corresponding in structure to the groups or subgroups they represent. They can be constructed explicitly by using "group membership" matrices consisting of ones and zeros that uniquely assign each of the H observations to one of the G (or L or NT) groups. Let R_ν be such an $(H \times G)$ matrix corresponding to the innermost group level. Then the block-diagonal $(H \times H)$ matrix J_ν can be expressed as the outer product of its membership matrices: $J_\nu = R_\nu R_\nu'$. The inner product $R_\nu' R_\nu$ produces a diagonal matrix $\widetilde{L}_\nu$ of size $(G \times G)$, which contains the number of observations of each group. Similarly, let R_μ be such an $(H \times L)$ matrix corresponding to the second-level groups. Then the block-diagonal $(H \times H)$ matrix J_μ can be expressed as the outer product of its membership matrices: $J_\mu = R_\mu R_\mu'$. Last, let R_δ be such an $(H \times NT)$ matrix corresponding to the top-level groups. Then the block-diagonal $(H \times H)$ matrix J_δ can be expressed as the outer product of its membership matrices: $J_\delta = R_\delta R_\delta'$.

If we pool the observations, the log-likelihood is given by

$$\ln l = -\frac{1}{2} H \ln(2\pi) - \frac{1}{2} \ln|\Omega| + \ln|A| - \frac{1}{2} \varepsilon' \Omega^{-1} \varepsilon \,, \tag{12.7}$$

where

$$\varepsilon = Ay - X\beta \;, A = I_H - \lambda W \,, \tag{12.8}$$

with $W = diag(W_t)$ and $\lambda = diag(\lambda_t)$, where W is the block-diagonal spatial weight matrix of size $(H \times H)$. W_t is the spatial weight matrix[1] of size $(F_{ta} \times F_{ta})$ changing at each time period t. λ is the spatial lag matrix of size $(T \times T)$ whose elements λ_t change at each time period t. I_H is an identity matrix of size $(H \times H)$.

The variance-covariance matrix of the disturbance is defined as follows

$$\Omega = \mathrm{E}\left[\varepsilon\varepsilon'\right] = \sigma_u^2 \left[I_H + \rho_\nu J_\nu + \rho_\mu J_\mu + \rho_\delta J_\delta\right] \,, \tag{12.9}$$

with

$$\rho_\delta = \frac{\sigma_\delta^2}{\sigma_u^2} \,, \rho_\mu = \frac{\sigma_\mu^2}{\sigma_u^2} \,, \rho_\nu = \frac{\sigma_\nu^2}{\sigma_u^2} \,. \tag{12.10}$$

Extending the derivations of Antweiler (2001) to the case of the spatial lag model (12.4), Baltagi et al. (2015) get

$$\ln l = -\frac{1}{2}\left[H \ln\left(2\pi\sigma_u^2\right) + \sum_{t=1}^{T}\left\{\ln|I_t - \lambda_t W_t| + \sum_{a=1}^{N}\left\{\ln\theta_{ta} + C_{ta} - \frac{\rho_\delta}{\theta_{ta}}\frac{U_{ta}^2}{\sigma_u^2}\right\}\right\}\right] , \tag{12.11}$$

$$\text{with } C_{ta} = \sum_{q=1}^{Q_{ta}}\left\{\ln\theta_{taq} + C_{taq} - \frac{\rho_\mu}{\theta_{taq}}\frac{U_{taq}^2}{\sigma_u^2}\right\} , \tag{12.12}$$

$$\text{and } C_{taq} = \sum_{i=1}^{M_{taq}}\left\{\ln\theta_{taqi} + \frac{V_{taqi}}{\sigma_u^2} - \frac{\rho_\nu}{\theta_{taqi}}\frac{U_{taqi}^2}{\sigma_u^2}\right\} , \tag{12.13}$$

where I_t is an identity matrix of size $(F_{ta} \times F_{ta})$ and

[1] Baltagi et al. (2015) use a block-diagonal weight matrix W of $(156,896 \times 156,896)$ whose smallest sub-block is a weight matrix W_t of $(6,643 \times 6,643)$ for the year 1992 and whose largest sub-block is a weight matrix W_t of $(17,098 \times 17,098)$ for 1999.

$$\begin{cases} \theta_{taqi} = 1+\rho_v F_{taqi} & V_{taqi} = \sum_{f=1}^{F_{taqi}} \varepsilon_{taqif}^2 , \\ \theta_{taq} = 1+\rho_\mu \phi_{taq} \quad \text{with } \phi_{taq} = \left(\sum_{i=1}^{M_{taq}} \frac{F_{taqi}}{\theta_{taqi}} \right) & U_{taqi} = \sum_{f=1}^{F_{taqi}} \varepsilon_{taqif} , \\ \theta_{ta} = 1+\rho_\delta \phi_{ta} \quad \text{with } \phi_{ta} = \left(\sum_{q=1}^{Q_{ta}} \frac{\phi_{taq}}{\theta_{taq}} \right) & U_{taq} = \sum_{i=1}^{M_{taq}} \frac{U_{taqi}}{\theta_{taqi}} , \\ & U_{ta} = \sum_{q=1}^{Q_{ta}} \frac{U_{taq}}{\theta_{taq}} , \end{cases} \tag{12.14}$$

with $\varepsilon_{taqif} = y_{taqif} - \lambda_t \sum_{a=1}^{N} \sum_{q=1}^{Q_{ta}} \sum_{i=1}^{M_{taq}} \sum_{p=1}^{F_{taqi}} w_{taqip} y_{taqip} - X_{taqif}\beta$.
A gradient of this log-likelihood function (12.11) is obtained analytically, but it can also be obtained through numeric approximation. In carrying out this maximization, it is necessary to constrain the optimization such that $|\lambda_t| < 1$, the variance σ_u^2 remains positive, and that the variance ratios ρ_δ, ρ_μ and ρ_v remain non-negative.

Baltagi et al. (2015) report several ML estimation results. One for the random effects (RE) model ignoring the nested effects, one for the nested RE model ignoring the spatial lag effects, and one for the spatial nested RE model.[2] Baltagi et al. (2015) found significant spatial lag effects as well as significant nested random error effects. They emphasize the importance of nested effects in the Paris housing data as well as the spatial lag effects. In fact, they show that the impact of the adjacent neighborhoods becomes relatively small when one takes care of the nested random effects. In addition, due to the unbalanced pseudo-panel aspect of these transactions, they show that one should allow the spatial weight matrix as well as the spatial lag coefficients to vary over time, and that the likelihood ratio tests confirm that they fit the Paris housing data better.

Following LeSage and Pace (2009), Baltagi et al. (2015) compute the marginal effects – which are decomposed into direct, indirect and total marginal effects – and show that the marginal spillover effects due to the neighbors are negligible relative to the direct effects. Moreover, empirical results show that the marginal effect for a specific housing characteristic is lower on average once the nested effects are taken into account.

Baltagi et al. (2014) estimate a nested random effects spatial autoregressive panel data model to explain annual house price variation across 353 local authority districts in England over the period 2000–2007. The nested error components represent the impact of local (district) effects embedded within wider (county) effects. Baltagi et al. (2014) propose new estimators based on the instrumental variable approaches of Kelejian and Prucha (1998) and Lee (2003) for the cross-sectional spatial autoregressive model. The estimation methods allow for the endogeneity of the spatial lag variable producing the simultaneous spatial spillover of prices across districts

[2] For the estimation of a nested error components model with unbalanced panel data using simple analysis of variance (ANOVA), maximum likelihood (MLE) and minimum norm quadratic unbiased estimators (MINQUE)-type estimators of the variance components, see Baltagi et al. (2001). For Lagrange multiplier testing of a nested error components model with unbalanced panel data, see Baltagi et al. (2002).

together with the nested random effects in a panel data setting. Monte Carlo results show that these estimators perform well relative to alternative approaches and produce estimates based on real data that are consistent with the theoretical house price model underpinning the reduced form. The empirical results show that there is a significant spatial lag term indicating a positive correlation between prices locally and prices in "nearby" districts and that income within commuting distance has a positive effect, while the stock of housing has a negative effect on housing price. They also show that the nested error components attributable to district and county effects, like the spatial lag term, are necessary elements in modeling UK house prices.[3]

From hedonic price functions, we can derive temporal and/or spatial price indexes. This has been done, for instance, by Syed et al. (2008) for the Sydney region. Their data concern 15 regions in Sydney on a quarterly basis from 2001 to 2006 from a data set consisting of 418,877 house sales. As 60% of sales observations are missing for one or more of the core characteristics, they first use multiple-imputation techniques to fill in the gaps in the data set, prior to estimating the hedonic model. In a second stage, they specify and estimate a non-nested three-dimensional hedonic price function. They pool across all the regions and periods in the sample and estimate the region-time specific fixed effects and shadow prices of housing characteristics. This method was first proposed by Aizcorbe and Aten (2004), who refer to it as the "time-interaction-country product dummy" method.

$$\begin{aligned} p_{jth} = \alpha + \sum_{\tau=2}^{T} \beta_\tau q_{\tau h} + \sum_{\kappa=2}^{J} \gamma_\kappa r_{\kappa h} + \sum_{\tau=2}^{T}\sum_{\kappa=2}^{J} \delta_{\tau\kappa} b_{\tau\kappa h} \\ + \sum_{m=2}^{M_\kappa} \eta_{\kappa m} d_{\kappa m h} + Z_{jth}\theta + \varepsilon_{jth}\,, \\ \text{for } j = 1,...,J,\, t = 1,...,T \text{ and } h = 1,...,H_{jt}\,, \end{aligned} \tag{12.15}$$

where p is the log of the price of a dwelling h belonging to region-period jt, $q_{\tau h}$ (resp. $r_{\kappa h}$) are dummy variables such that $q_{\tau h} = 1$ (resp. $r_{\kappa h} = 1$) if the observation h is from period t (resp. from region j) and zero otherwise. The dummy variables $b_{\tau\kappa h}$ denote interactions between periods and regions taking the value of 1 if the observation h is from region-period jt and zero otherwise. The postcode dummies are denoted by $d_{\kappa m h}$, where $d_{\kappa m h} = 1$ for observation h's postcode and zero otherwise. Z is a set of quality characteristics including the dwelling type, the number of bedrooms, bathrooms, lot size, etc. Spatial correlation between observations is

[3] Baltagi and Pirotte (2014) derive the Best Linear Unbiased Predictor (BLUP) for a spatial nested error components panel data model. This predictor is useful for panel data applications that exhibit spatial dependence and a nested hierarchical structure. The predictor allows for unbalancedness in the number of observations in the nested groups. This could be interesting for forecasting average housing prices located in a county nested in a state. When deriving the BLUP, this paper takes into account the spatial correlation across counties, as well as the unbalancedness due to observing different numbers of counties nested in each state. Ignoring the nested spatial structure leads to inefficiency and inferior forecasts. Monte Carlo simulations show that the resulting feasible predictor is better in root mean square error performance than the usual fixed and random effects panel predictors which ignore the spatial nested structure of the data.

defined by a spatial autoregressive process on the error term: $\varepsilon_{jth} = \lambda W \varepsilon_{jth} + u_{jth}$ where $u_{jth} \sim N(0, \omega_{jth}\sigma^2)$. The spatial weight matrix W is a contiguity matrix and the variance of u_{jth} is subscripted with jt allowing for heteroskedasticity. The coefficients δ_{jt} measure the region-period specific fixed effects for the logarithms of the price level after controlling for the effects of the attributes of the dwellings. The model is estimated using the maximum likelihood method. The advantage of this region-time-dummy model is that the temporal and regional price indexes are derived directly from the estimated coefficients $\hat{\beta}_t$, $\hat{\gamma}_j$, $\hat{\delta}_{jt}$, $\hat{\eta}_{jm}$ and $\hat{\theta}$. Let $P_{j,t,s}$ the price index for region j in year t and quarter s. Then, the relative prices are given by

$$\begin{aligned} \frac{P_{j,t,s}}{P_{j,t,1}} &= \exp\left(\hat{\beta}_{t,s} + \hat{\delta}_{j,t,s}\right) \text{ for } s = 2,3,4, \\ \text{and} \quad \frac{P_{j,t+1,1}}{P_{j,t,1}} &= \exp\left(\hat{\beta}_{t+1,1} + \hat{\delta}_{j,t+1,1}\right). \end{aligned} \tag{12.16}$$

Therefore, it is possible to construct a temporal price index for each region j over the entire time period of the dataset. Results are normalized such that the price index for the initial region (Inner Sydney) is equal to 1 for the first quarter of 2001. One can also construct a spatial price index for each quarter s of a specific year t for the entire set of regions. For a given quarter (t,s), spatial price indexes can be constructed from the estimated coefficients $\hat{\gamma}_j$, $\hat{\delta}_{jt}$, $\hat{\eta}_{jm}$ and $\hat{\theta}$. The starting point is a comparison between a postcode m in region l and a postcode n in region j for a particular dwelling h with amenities vector Z_{ch}. This spatial price index is defined as

$$\begin{aligned} P_{lmts,jnts}(Z_{ch}) = &\exp\left[(\hat{\gamma}_j - \hat{\gamma}_l) + \left(\hat{\delta}_{jt} - \hat{\delta}_{lt}\right) + (\hat{\eta}_{jn} - \hat{\eta}_{lm})\right] \\ &\times \left[\prod_{c=1}^{C} \exp\left[Z_{ch}\left(\hat{\theta}_{jc} - \hat{\theta}_{lc}\right)\right]\right], \end{aligned} \tag{12.17}$$

and the spatial index can be generalized to take into account all dwellings sold in postcodes lm

$$\begin{aligned} P_{lmts,jnts} = &\exp\left[(\hat{\gamma}_j - \hat{\gamma}_l) + \left(\hat{\delta}_{jt} - \hat{\delta}_{lt}\right) + (\hat{\eta}_{jn} - \hat{\eta}_{lm})\right] \\ &\times \left[\prod_{h=1}^{H_{lmts}} \prod_{c=1}^{C} \exp\left[Z_{ch}\left(\hat{\theta}_{jc} - \hat{\theta}_{lc}\right)\right]\right]^{1/H_{lmts}}. \end{aligned} \tag{12.18}$$

This is close to a Laspeyres price index.

Combining the temporal and spatial indexes allows a price comparison of dwellings between different location-year-quarter triplets. Syed et al. (2008) found that their hedonic house price indexes rose significantly from 2001 to 2003, after which they fell slightly. This finding is consistent with the Australian Bureau of Statistics (ABS) index. Their indexes, however, are less volatile than their ABS counterpart, rising noticeably less in the boom and falling less thereafter. In the spatial dimen-

sion, they found large and systematic differences in the price of housing across regions of Sydney. The regional dispersion narrowed during the boom period but appears to have increased again since then.

Several authors have shown that values of complex assets are difficult to accurately quantify and information asymmetry affects asset prices through various channels (see, e.g., Agarwal and Hauswald, 2010; Baker and Wurgler, 2007; Carlin et al., 2013; Kelly and Ljungqvist, 2012). The subprime crisis (poor household mortgage decisions and subsequent foreclosure), and the housing market collapse in the US, followed by the financial crisis have revealed that uninformed buyers overpay. The house buying mechanism is a field in which households' ability (or inability) to use market information may have strong effects on housing decisions. This could be through the choice of mortgage product and through the purchase transaction (see Carlin et al., 2013; Turnbull and van der Vlist, 2015). House purchases may involve residential mortgages and associated complex financial instruments, which have been identified as a major cause of waves of foreclosures during and after the 2007–2008 financial crisis. Turnbull and van der Vlist (2015) show that buyers who are uninformed of the housing market pay more for houses than buyers who are informed. They use pseudo-panels of repeated sales based on neighborhood census block-level. This data is for 426,021 parcels located in Orange County, Florida, over the period 2000–2012. The authors split fair market value and uninformed buyer effects by first identifying for each of the market sales in the period 2000–2006 which of the units foreclosed in 2007–2012. The future foreclosure dummy FF equals 1 if a market transaction completed in 2000–2006 is followed by a foreclosure in 2007–2012 and equals zero otherwise. Turnbull and van der Vlist (2015) estimate a hedonic price function of the log of market price in first differences on the neighborhood block-level j

$$p_{itj} - p_{lsj} = \left(Z_{itj} - Z_{lsj}\right)\beta_Z + \left(FF_{ij} - FF_{lj}\right)\beta_{FF} + \varepsilon_{itj} - \varepsilon_{lsj}, \quad \text{for } t,s = 1,...,T,\ i,l = 1,...,N,\ j = 1,...,J \text{ for all } i \neq j \text{ and } t \neq s, \tag{12.19}$$

where p_{itj} is the log of the price of property i sold at time t located in area j. Z is the vector of relevant house characteristics, and amenities and FF is the penalty associated with being foreclosed *ex post* (over 2007–2012). The model of first differences at the neighborhood block-level basically treats sales within the neighborhood block as repeat sales while accounting for observed structural differences. This is a model on pseudo-panels of repeated observations "*à la* Deaton (1985)". This model also allows for clustered errors at the neighborhood block-level j. Results show that buyers who are later foreclosed paid a 2.7% (resp. a 4.6%) premium for properties bought between 2000 and 2006 (resp. between 2005 and 2006). Estimation on different sub-periods also reveal a strong correlation between home buyers' house prices and future foreclosures. To check whether effects vary across housing market segments, Turnbull and van der Vlist (2015) estimate quantile regression models. Results show that the effect for the penalty associated with being foreclosed is larger for the lower end of the housing market. Buyers in 2005–2006 who ended up foreclosed paid up

to 3.5% above the fair market value in the lower end of the housing market, while foreclosed owners paid a little over 1% percent more in the higher end of the housing market.

12.4 Multi-dimensional Models of Residential Mobility and Location Choice: Some Examples

Residential mobility and location choice are significant household decisions and have been widely researched in various fields including economics, sociology, geography, regional science, urban planning, housing policy, transportation, etc. Decisions of residential mobility and location choice are closely related to the household housing process with a large range of factors that contribute to each choice. Due to the vastness of the literature on such topics, we will focus on a few examples of residential mobility and location choice. Readers could profitably read the survey by Dieleman (2001) on residential mobility. Since the seminal works of Rossi (1955) and Alonso (1964), a huge amount of research on residential location choice has been published. *"Reasons for moving are divided into those which pertain to the decision to move out of the former home - 'pushes' - and those reasons pertaining to the choice among places to move to - 'pulls' "* (Rossi, 1955, p. 8). For instance, push factors may include negative externalities like noise, pollution or crime, changes in housing affordability, dissatisfaction with the current dwelling, changes in household structure, etc. Pull factors often include better access to good quality public services (schools and health care facilities), employment, leisure and recreational opportunities, etc. (see Lee and Waddell, 2010; Hoang and Wakely, 2000 for a review). Our purpose is not to review the main factors of residential mobility and relocation but to summarize a few multi-dimensional studies of residential mobility and relocation.

One interesting study has been done by Davies and Pickles (1985) in a multi-dimensional framework. They propose a model that conceptualizes residential mobility as a sequence of choices between staying and moving. Household i will move in time period t if and only if random utility derived from the most-favored alternative dwelling available u_{itb} is larger than the random utility derived from the current dwelling u_{ita}:

$$\begin{aligned} u_{ita} &= V(y_{it}, Z_{ta}) + \varepsilon_{ita} = V_{ita} + \varepsilon_{ita} \text{ with } \varepsilon_{ita} = \mu_{ia} + g(d_{it}) + \nu_{ita}, \quad (12.20) \\ u_{itb} &= V(y_{it}, Z_{tb}) + \varepsilon_{itb} = V_{itb} + \varepsilon_{itb} \text{ with } \varepsilon_{itb} = \mu_{ib} + h(t) + \nu_{itb}, \end{aligned}$$

where y_{it} is a vector of observed characteristics of household i at time t, Z_{ta} (resp. Z_{tb}) is a vector of the observed characteristics of the current dwelling (resp. the most-favored alternative dwelling available). V_{ita} and V_{itb} are the systematic utilities while ε_{ita} and ε_{itb} are the random components of utilities. These random components are likely to be correlated over time for each household. ε_{ita} is the sum of the unexplained household heterogeneity μ_{ia}, a function $g(d_{it})$ of the duration of stay

for household i at time t and a remainder term v_{ita}, independently distributed over both households and time. For the other random component ε_{itb}, the unexplained household heterogeneity μ_{ib} also applies. Moreover, a time trend $h(t)$ represents fluctuations in market conditions. Davies and Pickles (1985) used a quadratic specification for the duration of stay $g(d_{it}) = \beta_1 d_{it} + \beta_2 d_{it}^2$, and a cubic specification for the housing market function $h(t) = \beta_3 t + \beta_4 t^2 + \beta_5 t^3$.

The likelihood $L(z_{it})$ of the observed sequence of outcomes is the product of the probabilities of the observed choice for each time period:

$$L(z_{it}) = \prod_{t=1}^{T} \{Pr[u_{itb} > u_{ita}]\}^{z_{it}} \{1 - Pr[u_{itb} > u_{ita}]\}^{1-z_{it}} , \tag{12.21}$$

$$\text{with} \qquad Pr[u_{itb} > u_{ita}] = \int_{-V_{it}-\mu_i+g(d_{it})-h(t)}^{\infty} \phi(v_{itb} - v_{ita})\, d(v_{itb} - v_{ita}) ,$$

where $z_{it} = 1$ if household i moves in time period t and zero elsewhere, $V_{it} = V_{itb} - V_{ita}$, $\mu_i = \mu_{ib} - \mu_{ia}$ and $\phi(.)$ is the probability density of the difference between the two random components. Assuming that they follow Weibull distributions, leads to the following likelihood with a household-specific error term μ_i:

$$L(z_{it}) = \prod_{t=1}^{T} \frac{\exp[-V_{it} - \mu_i + g(d_{it}) - h(t)]^{z_{it}}}{1 + \exp[-V_{it} - \mu_i + g(d_{it}) - h(t)]} . \tag{12.22}$$

Three problems arise with this likelihood: the integration over the error term distribution is almost analytically intractable; the initial observation complicates the handling of endogenous variables such as duration of stay d_{it}, and numerical methods are required for parameter estimation. To overcome these problems, Davies and Pickles (1985) derived an approximation of the likelihood using the generalized Beta-logistic approach developed by Davies (1984).

The panel data is for 887 households participating in the Michigan Panel Study of Income Dynamics over the period 1968–1977. The dependent variable was a residential move within the county or an intercounty move with no change in the head-of-household's job. Among the main explanatory variables were the duration of stay, a room adequacy index (actual rooms / required rooms), an income adequacy index (actual income / needs), the age of the head of household, and the education level. First, they show that the room adequacy index has a U-shaped relationship with residential mobility. Renters have the shortest initial duration status, while owners have the longest. But, there is no evidence of a similar U-shaped relationship anticipated for the income adequacy index. Second, they show that changing financial circumstances does not seem to play any role in the life cycle variation in residential mobility in the United States. Moreover, they are unable to show any effect of income surplus on residential mobility.

These are unexpected results. Davies and Pickles (1985) argue that these results may be due to the housing market being highly segmented, not just between renting and owner-occupation, but between different types of property and their location. It could be interesting to redo this study with more recent data. It will probably give

different conclusions for the last decade which has known troubled financial periods. Davies and Pickles (1985) found a strong negative relationship between the age of the head of household and residential mobility. This strong negative relationship is present even when changing space requirements and financial pressures are accounted for. The age of the head of household is the dominant life cycle and acts as a proxy variable for changing needs and financial circumstances through the life cycle.

Explaining the factors which determine housing tenure choices is important. For instance, Fu et al. (2015) estimate multilevel multinomial logistic regressions for housing types to study home ownership in urban China. They base their estimation on a sample data of $2,585,480$ households from the 2005 National Population Sample Survey of China and available information for 205 urban areas (prefecture-level data) (see Huang and Clark, 2002 for a similar study in China but in a 2D framework). For one household i in prefecture j, the within-prefecture multinomial logistic model for the odds of housing type m are given by

$$\ln\left[\frac{Pr\left(\text{housing type}_{mij}\right)}{Pr\left(\text{private rental housing}_{ij}\right)}\right]=\beta_{mj0}+\sum_{k=1}^{K}\beta_{km}\left(Z_{h,kijm}-\overline{Z}_{h,kjm}\right)+\varepsilon_{ij}\,. \tag{12.23}$$

The $m=1,..,5$ housing types refer to owning self-built housing, owning commodity housing, owning affordable housing, owning privatized *danwei* housing and public rental housing. $Z_{h,kijm}$ is the value of household-level covariate k associated with household i in prefecture j for the m-th housing type. $\overline{Z}_{h,kjm}$ is the sample mean of covariate k within prefecture j. The household-level error term ε_{ij} is assumed to be $i.i.N(0,\sigma^2)$. The between-prefecture model for housing types is

$$\beta_{mj0}=\gamma_{00m}+\sum_{s=1}^{S}\gamma_{0sm}Z_{p,sjm}+\eta_{0jm}\,, \tag{12.24}$$

where $Z_{p,sjm}$ is the prefecture-level covariate s in prefecture j for the m-th housing type and η_{0jm} is the prefecture-level error term, which is assumed to be $i.i.N(0,\sigma_m^2)$.

Using a generalized linear mixed model with random effects estimation methods (GLMM), Fu et al. (2015) show at the household level that redistributors (e.g., cadres) and supporting clerical staff were more likely to achieve home ownership than manual workers did. Both non-agricultural status and working in state sectors confer benefits in obtaining reform-era housing with heavy subsidies or better quality. When one takes into account education and earnings, the advantage of redistributors (e.g., cadres) over manual workers in home ownership could be explained by work units. At the prefecture-level, they show that the marketization only reduced the local home ownership of self-built housing, affordable housing and privatized *danwei* housing but not that of commodity housing. In contrast, political and market connections promote all types of home ownership except self-built housing, and have a significant positive association with the odds of renting public housing.

Numerous studies focus on how neighborhoods change in terms of income level, housing values, environment amenities or different racial preferences, etc. Racial and ethnic composition may have effects on neighborhood economic change (see for instance Sykes, 2003). Some studies have examined how neighborhood minority composition is associated with change in neighborhood relative economic status. For instance, the paper by Jun (2016) in a 3D framework uses the Neighborhood Change Database (NCDB), which includes the decennial census data across the USA from 1970 to 2000 at the census tract level. The multilevel modeling fits the data structure that a neighborhood is nested in a metropolitan area and allows for answering the research question whether the effect of neighborhood racial/ethnic composition on neighborhood economic change is conditioned by metropolitan-level factors. Jun (2016) shows that both neighborhood percentages of Blacks and Hispanics are negatively related to neighborhood economic gain and are conditioned by metropolitan-level factors. Although this negative effect of neighborhood minority composition has been consistent over the four ten-years panel, – the 1970s, 1980s, 1990s, and 2000s – its impact level is lower in the latest panel compared to the earliest. The negative effect of neighborhood minority composition has also declined as a result of the interactions with metropolitan minority composition. In the later panels, metropolitan minority composition turned out to moderate the negative effect of neighborhood minority composition.

Explaining residential choices and residential mobility is not sufficient. It seems important to jointly model residential mobility and the duration of stay at a location preceding relocation. A considerable amount of research has treated the decision to move as a binary choice decision (move/no-move) and modeled this decision as a function of various factors (see above). Others have used duration models (see Deng et al., 2003) to represent the stay at a location between moves, treating the reason for a move as an exogenous variable. An interesting study done in a multidimensional framework by Eluru et al. (2009) has extended these previous studies in three ways. First, the move decision is treated as an endogenous variable in a multinomial unordered choice modeling framework. Second, the duration of stay is modeled as a grouped choice, supposing that households treat the duration of stay at a residential location in terms of time-period ranges as opposed to exact continuous durations. Third, they consider heterogeneity of exogenous variables using random coefficients in both the equation for the move as well as the equation for the duration of stay preceding a relocation. In sum, Eluru et al. (2009) estimated a joint unordered choice-grouped choice model system with random coefficients.

Let the households be represented by the index $i = 1, ..., N$, let the different move reasons (e.g., personal reasons, employment reasons, etc.) be represented by the index $m = 1, ..., M$ and let the duration categories (e.g., < 2 years, $2 - 5$ years, $5 - -10$ years, etc.) be represented by the index $j = 1, 2, ..., J$. The specification of Eluru et al. (2009) allows the possibility of multiple move records per household to be defined by the index $t = 1, 2, ..., T$ as the different moving choice occasions for households i. The system of equations jointly models the reason for move and the duration of stay as follows

$$\begin{cases} u_{imt} = X_{it}'\beta_{im} + \eta_{im} + \varepsilon_{imt}\,, \\ d_{imt} = j \text{ if } \psi_{m,j-1} < d_{imt}^{*} < \psi_{m,j}\,, \end{cases} \tag{12.25}$$

$$\text{with } d_{imt}^{*} = X_{it}'\alpha_{im} \pm \eta_{im} + \zeta_{imt}\,. \tag{12.26}$$

The first equation of the system is associated with the random utility u_{imt} for a household i corresponding to the reason to move m at choice occasion t. The $(Q \times 1)$ vector X_{it} is the vector of attributes associated with household i and its choice environment (e.g., sex, age, employment status, family type, transportation mode to work, etc.) at the t-th choice occasion. The $(Q \times 1)$ random coefficient vector $\beta_{im} = \beta_m + \gamma_{im}$ is the sum of a vector β_m of mean effects of the elements of X_{it} for move reason m and a random vector γ_{im} with its q-th element $(q = 1, ..., Q)$ representing unobserved factors specific to household i and its choice environment. η_{im} expresses unobserved individual factors that simultaneously impact the propensity of moving for a certain reason m and the duration of stay. ε_{imt} is an idiosyncratic random error term assumed to be identically and independently standard Gumbel distributed across individuals, move reasons and choice occasions.

The second equation of the system is associated with d_{imt}^{*}, being the latent (continuous) duration of stay for household i before moving for reason m on the t-th choice occasion. This latent duration is mapped to the grouped duration category d_{imt} by the ψ thresholds (with infinite bounds as in the usual ordered-response modeling framework). d_{imt} is observed only if the end of the duration of stay at a residential location is associated with alternative m. The $(Q \times 1)$ random coefficient vector $\alpha_{im} = \alpha_m + \delta_{im}$ is the sum of the vector α_m of mean effects for category m, and the random vector δ_{im} of unobserved factors specific to household i and its duration of stay. ζ_{imt} is an idiosyncratic random error term, assumed identically and independently distributed with a logistic distribution across individuals, reasons for move, and choice occasions, with variance λ^2. The elements of the random vectors γ, δ and η are normally distributed: $\gamma_{imq} \sim N(0, \sigma_{\gamma_{mq}}^2)$, $\delta_{imq} \sim N(0, \sigma_{\delta_{mq}}^2)$ and $\eta_{im} \sim N(0, \sigma_{\eta_m}^2)$ for $q = 1, ..., Q$.

Correlation in unobserved individual factors between the reason to move and the duration of stay may be positive or negative, it is indicated by the $\pm$ sign in front of η_{im} in the duration category equation. If a positive sign seems logical for the propensity of a move for a given reason m in the first equation, a negative sign in the second equation suggests that unobserved individual factors will decrease the duration of stay preceding such a potential move. In the estimation, Eluru et al. (2009) considered both the positive and negative signs on the η_{im} terms in the second equation of the system. But the negative sign for all m provided statistically superior results. Conditional on γ_{im} and η_{im} for each (and all) m, the probability of a household i choosing to move for reason m on the t-th choice occasion is given by

$$P_{imt} = \frac{\exp\left(X_{it}'\beta_{im} + \eta_{im}\right)}{\sum_{m=1}^{M} \exp\left(X_{it}'\beta_{im} + \eta_{im}\right)}\,. \tag{12.27}$$

Conditional on δ_{im} and η_{im}, the probability of a household i choosing to stay for a particular duration category j preceding a move for reason m on the t-th choice occasion is given by

$$R_{imtj} = G\left(\frac{\psi_{m,j} - \{X'_{it}\alpha_{im} \pm \eta_{im}\}}{\lambda}\right) - G\left(\frac{\psi_{m,j-1} - \{X'_{it}\alpha_{im} \pm \eta_{im}\}}{\lambda}\right), \quad (12.28)$$

where $G(.)$ is the cumulative distribution of the standard logistic distribution. Let Ω be a vector that includes all the parameters β_m, α_m, λ, $\sigma_{\gamma_{mq}}$, $\sigma_{\delta_{mq}}$ and σ_{η_m} for $m = 1,...,M$ and $q = 1,...,Q$. Let c_i be a vector stacking the coefficients γ_{im}, δ_{im} and η_{im} across all m for household i. Let Σ be another vector stacking the standard error terms $\sigma_{\gamma_{mq}}$, $\sigma_{\delta_{mq}}$ and σ_{η_m} and let $\Omega_{-\Sigma}$ represent a vector of all parameters except the standard error terms. Then, the unconditional likelihood function for all the households is given by

$$L(\Omega) = \prod_{i=1}^{N} L_i(\Omega) = \prod_{i=1}^{N} \int_{c_i} \{L_i(\Omega_{-\Sigma} \mid c_i)\}\, d\Phi(c_i \mid \Sigma), \quad (12.29)$$

$$\text{with } L_i(\Omega_{-\Sigma} \mid c_i) = \prod_{m=1}^{M}\prod_{t=1}^{T}\prod_{j=1}^{J} [P_{imt}R_{imtj}]^{D_{imt}E_{ijt}},$$

where $\Phi(.)$ denotes the multi-dimensional cumulative normal distribution and $L_i(\Omega_{-\Sigma} \mid c_i)$ is the likelihood function, for household i and for a given value of $\Omega_{-\Sigma}$ and c_i. D_{imt} (resp. E_{ijt}) is a dummy variable taking the value of 1 if household i chooses to move for reason m (resp. chooses to stay for duration category j) on the t-th choice occasion and 0 otherwise. Equation (12.29) needs the evaluation of a multi-dimensional integral of size equal to the number of rows in c_i. Eluru et al. (2009) apply Quasi-Monte Carlo simulation techniques based on the Halton sequence to approximate this integral in the likelihood function and maximize the logarithm of the resulting simulated likelihood function across individuals with respect to Ω (see Bhat, 2001, 2003). Eluru et al. (2009) use a longitudinal data set of households from a stratified sample of municipalities in the Zurich region of Switzerland over the period 1985–2004. The data set includes 1012 households and 2590 move records. They found that several demographic, socioeconomic, and commute related variables (e.g., age, gender, family reasons, education/employment reasons, accommodation related reasons, surrounding environment related reasons, vicinity to family and friends, etc.) have a significant influence on the reason for move and the duration of stay. In the duration of stay model, Eluru et al. (2009) found that household size creates heterogeneity across the sample of households. They show that people who own dwellings have a lower probability of moving for surrounding vicinity related reasons than those renting their units. Likewise, people who live in smaller homes have higher probabilities of short duration stays probably because they are looking for larger homes. Having a mix of job opportunities located close to residential neighborhoods increases the duration of stay in the dwelling. Reducing commute distances promotes longer durations of stay, etc. Eluru et al. (2009) found that common unobserved factors jointly affect the reason to move and the du-

ration of stay and call for a joint modeling framework that allows error correlation structures.

Endogeneity (or simultaneity) is a fundamental aspect of modelling housing that should be taken into account both for hedonic housing price functions and for choice models of residential location. This is the object of the next section.

12.5 Multi-dimensional Dynamic Models of Housing Models

In hedonic housing price functions, some explanatory variables, in addition to the dependent variable and its spatial lag, may be endogenous following the simultaneous choice of the house price and of the quantities of attributes. This is particularly true for floor space (see Fingleton and LeGallo 2008, who extended Kelejian and Prucha's 1998 feasible generalized spatial two-stage least squares estimator to account for endogenous variables due to system feedback, given an autoregressive or a moving average error process). As for hedonic price functions, endogeneity is expected to occur mainly as a result of the omission of attributes in discrete choice models of residential mobility. In the literature, several methods have been proposed to consider endogeneity. Berry et al. (1995) proposed a fixed effects procedure by product and market to solve market-level endogeneity in the automobile sector. Guevara and Ben-Akiva (2006) applied to residential location choice models the control function method, which is based on the inclusion of an additional variable that controls for the endogeneity problem (see Heckman, 1978; Blundell and Powell, 2004). They applied residential location choice models based on 630 households of renters who had moved to their present location between 1999 and 2001 in Santiago (Chile). The results show that price endogeneity is significant in choice models of residential location and that the control function method can account for it.

Endogeneity is not limited to the correlation between the dependent variables and attributes (in the equation or omitted) or to the simultaneity of demand and supply, the marginal willingness to pay and the marginal willingness to accept. Location choices and housing investments are fundamentally dynamic decisions over multiple time periods. In the 2D panel data literature, some dynamic models have been applied to real estate topics. For instance, Engle et al. (1985) used a version of a dynamic multiple-indicator multiple-cause (DYMIMIC) model for a hedonic price model of the resale housing market for a suburb of San Diego, California, during the period 1973–1980. The specification of the model features hedonic equations for each house sale and a dynamic equation for the capitalization rate, which is taken to be an unobservable time series to be estimated jointly with the unknown parameters. Engle et al. (1985) used maximum likelihood with an EM algorithm based upon Kalman filtering.

Some authors have used, in a 2D framework, the dynamic factor models (DFM) and/or large-scale Bayesian vector autoregressive (LBVAR) models to forecast housing prices. These models are interesting to study the "ripple effect", i.e., the propagation of shocks to house prices across regions. For instance, Das et al. (2010)

forecast regional house price inflation for five metropolitan areas of South Africa, using principal components obtained from quarterly macroeconomic time series in the period 1980 to 2006. In the majority of cases, the dynamic factor model statistically outperforms the vector autoregressive models, using both the classical and the Bayesian treatments. They also considered spatial and non-spatial specifications. Das et al. (2010) indicate that macroeconomic fundamentals in forecasting house price inflation are important. Li and Leatham (2011) investigate moving trends of house prices in 42 metropolitan areas in the United States from the perspective of large-scale models, which are also DFM and LBVAR models. These models accommodate a large panel data comprising 183 monthly series for the U.S. economy, and an in-sample period of 1980 to 2007 are used to forecast the one to twelve-months-ahead house price growth rate over the out-of-sample horizon of 2008 to 2010. Li and Leatham (2011) show that DFM consistently outperforms its LBVAR alternative for forecasting the house price growth rate for the overall U.S. housing market. The forecasting power of DFM does not decrease as the number of forecast periods ahead increases, while LBVAR has its best performance for the two-months-ahead forecast and then its forecasting accuracy decays.

Beenstock and Felsenstein (2015) using data from 9 regions of Israel over 1987–2010, apply spatial panel cointegration methods for a dynamic model of regional housing markets in which people prefer to live where housing is cheaper and building contractors prefer to build in regions where construction is more profitable. Based on dynamic hedonic price functions, the analysis of nonstationary spatial panel data shows that although housing starts vary directly with profitability as measured by house prices relative to building costs, they vary inversely with profitability in neighboring regions. Beenstock and Felsenstein (2015) show that there is a non-negligible spatial substitution in housing construction and this substitution effect suggests that contractors have local building preferences since they regard neighboring regions as close substitutes but not more distant regions. Abate and Anselin (2016) investigate the interactions between house price fluctuations and output growth rate across 373 metropolitan statistical areas in the US over the period 2001–2013. In a panel data context, they use time varying spatial econometric hedonic price functions. They show that the spatial correlation coefficient across metropolitan areas has been increasing over time, indicating an increasing synchronization of house prices across metropolitan statistical areas during the sample period.

Spatio-temporal models of hedonic price functions have recently been proposed to jointly take into account time effects and spatial effects either through multifactor error structure or through specific weight matrices. For instance, Holly et al. (2010) considered the determination of real house prices in a panel made up of 49 US States over 29 years. An error correction model with a cointegrating relationship between real house prices and real incomes is found once they take proper account of both heterogeneity and cross-sectional dependence (see also Latif, 2015 for a study on the impact of new immigration on housing rent, using Canadian province-level panel data from 1983 to 2010). Latif (2015) uses panel cointegration regressions and panel

vector error correction models and shows that immigration flow has a significant positive impact on housing rent both in the short and in the long run.

There are also extensions of the spatial hedonic price functions which use a weight matrix that expresses spatio-temporal rather than purely spacial relations. A general $(N \times N)$ spatio-temporal weight matrix W is obtained by splitting its construction into two separate matrices of the same dimension. The first matrix, S, captures the spatial relations among the N observations and a second matrix, T, expresses the temporal direction of observations. Smith and Wu (2011) have proposed a spatio-temporal weight matrix defined as the Hadamard product between two spatial and temporal distance weight matrices $W = S \odot T = \left[s_{jl}\right] \odot \left[t_{jl}\right]$. It identifies the spatio-temporal neighbors that affect hedonic price determination. The elements s_{jl} indicate the way observation j is spatially connected to observation l. The elements on the diagonal s_{jj} are set to zero, while the off-diagonal elements are defined by an inverse distance function: $s_{jl} = d_{jl}^{-\gamma}$ if $d_{jl} < \overline{d}$ and 0 elsewhere, where d_{jl} is the geographic distance between locations j and l, $d_{jl} < \overline{d}$ is a critical cut-off and $\gamma \geq 0$. The elements t_{jl} represent the time that elapsed between the realization of observations j and l. One assumes that observations have been ordered chronologically: the first row of T corresponds to the earliest observation, while the last row corresponds to the latest observation. The elements on the diagonal t_{jj} are set to zero, while the off-diagonal elements are defined by an inverse function of the time that elapsed between two observations: $t_{jl} = \mid t_j - t_l \mid^{-\alpha}$ if $\mid t_j - t_l \mid < \bar{t}$ and 1 elsewhere. t_j (resp. t_l) is the time when dwelling j (resp. l) is sold. $\bar{t}$ is a critical cut-off value and α is a penalty parameter to be fixed.

Several authors have used spatio-temporal models of hedonic price functions with standard spatial specifications (spatial autoregressive (SAR), spatial error (SEM), spatial Durbin model, etc.) but with different spatio-temporal matrices W. They got better results in terms of estimation and/or forecasting as compared to those obtained with the usual purely spatial weight matrices. See for instance, Pace et al. (2000) for an application on the residential market of Bâton Rouge, Louisiana, during 1984–1992, Liu (2013) for an application of housing in Randstad, the Netherlands, during the years 1997–2007, Nappi-Choulet and Maury (2011) for the residential market of Paris for the years 1995–2005, or Thanos et al. (2016) for the Aberdeen, Scotland, housing market during 2004–2007, to mention a few. To our knowledge, unfortunately, nobody has used these spatio-temporal multifactor error structures or the spatio-temporal weight matrices in a three-dimensional framework. However, this could be a promising development for future research.

The developments in the dynamics of modelling housing are focused not only on hedonic price functions. Some authors have been interested in dynamic versions of discrete models of location choice. Forward-looking behavior in the housing market justifies dynamic considerations in a model of location choice. Several authors have underlined the need to use dynamic specifications for modelling housing. For instance, Case et al. (2012), using questionnaire surveys for home buyers in four U.S. cities over 2003–2012, have shown that the root causes of the speculative bubble can be seen in their long-term home price expectations, which reached abnormal levels relative to the mortgage rate at the peak of the boom and have sharply declined since.

The downward turning point around 2005 of the long boom that preceded the crisis was associated with the changing public understanding of speculative bubbles. But estimating dynamic discrete models of location choice is a rather challenging and stimulating objective and is technically difficult. Bayer et al. (2016) noted that first, the estimation of residential sorting and hedonic equilibrium models needs to match a large sample of households, their characteristics to the location and the features of their housing choices. Second, the high dimensionality of the state space (consisting of current lifetime utilities and neighborhood characteristics) – required to define the evolution of an urban system – leads to the curse of dimensionality, which puts a brake on the estimation of an acceptable sized dynamic model of residential location decisions.

Diao et al. (2015) propose a real-option based dynamic model to simulate real estate developer behavior. In a three-dimensional framework (property, type of property and time for private residential housing in Singapore during 1995–2012), they extend the standard discrete choice model approach by adding an explicit probabilistic representation of development templates available to developers to take into account both the developers' option to hold the land undeveloped and the market volatility of different development types. In their proposed simulation framework, Diao et al. (2015) suppose that a developer making investment decisions for a parcel faces a set of alternative development templates in a market with uncertainty. In each time period, the developer estimates future revenue and the construction cost of feasible development templates under planning constraints and related real option values. He chooses the template based on the principle of profit maximization, but only does so if the return of the development template is higher than a threshold level (value of the call option), which is a function of the market volatility of the built property as suggested by the real option theory, otherwise, he keeps the *status quo*. The model components in the proposed simulation framework are calibrated with private housing data in Singapore. The results show significant volatility in housing prices and construction costs, relevant differences in volatility across housing types, and good fit in the hedonic model of market prices and construction costs. This kind of research contributes to the microsimulation literature by proposing an interesting approach which takes into account the dynamic and volatile nature of the real estate market but, unfortunately, this remains a simulation study.

Bayer et al. (2016) have proposed a new approach for estimating a three-dimensional dynamic model of demand for houses and neighborhoods that is computationally tractable. Using a semi-parametric estimation approach, they control for unobserved household and neighborhood heterogeneity. Their model adapts dynamic demand models for durable goods in a housing market context. They treat houses as assets and allow households' wealth to evolve endogenously. Households anticipate selling their homes at some point in the future and then consider the expected evolution of neighborhood amenities and housing prices when deciding where and when to purchase or sell their house. They relax the index sufficiency assumption which is standard in the dynamic demand literature.

This assumption helps to deal with the computational challenges posed by the large state space typically arising in models of dynamic demand. Instead of treat-

ing the logit inclusive value as a sufficient statistic for predicting future continuation values, Bayer et al. (2016) define the continuation value from predicted future lifetime utilities, which depend on the state space in a flexible manner. Last, they use stable and uniform realtor fees to estimate the marginal utility of consumption without the need for a price instrument. They use the fact that households face a monetary trade-off both in the standard sense of deciding which product (neighborhood) to purchase, but also in terms of deciding when to move. They take advantage of the fact that realtor fees during the sample period were quite uniform (6% of the house value) in order to identify the marginal utility of consumption when estimating each resident's move-stay decision. The decision variable, d_{it}, denotes both of the choices made by household i in period t, whether to move and where to move, conditional on deciding to move. If a household decides to move, the decision is denoted $d_{it} = j$, $j = 0, 1, ..., J$, where j indexes neighborhoods, J denotes the total number of neighborhoods in the region and 0 denotes the outside option. The data concern housing transactions in the San Francisco Bay Area from 1994–2004 for more than $220,000$ households and 2398 neighborhoods. We give only some results as the paper is highly technical. However, the model and estimation procedure presented in this paper are very general and can be applied to a broad range of dynamic studies in housing markets. The model uses a two-stage estimator. In the first stage, Bayer et al. (2016) use the household location and the mobility decisions to estimate the value of lifetime expected utility for each neighborhood, time period, and household type, as well as an unobservable characteristic that captures a household's preference for sub-regions within the San Francisco Bay Area. In the second stage, they recover fully-flexible estimates of per-period utility and regress them on a set of observable attributes. They use a semi-parametric estimation approach to control for the endogeneity of price in this second stage, utilizing outside information relating to the financial cost of moving to pin down the coefficient on house prices.

The results indicate that the downward biases associated with static demand estimation are significant for three important non-marketed amenities: air quality, crime, and neighborhood race. For instance, for a 10% change in each amenity, the static model overestimates the willingness to pay for living in close proximity to neighbors of the same race for low-income households. The static estimation is $\$1,627.03$, whereas the corresponding dynamic estimation is \$612.09. For high-income households, the bias runs in the opposite direction and the static model underestimates the willingness to pay by a factor of more than two. The static model always underestimates the willingness to pay for living in close proximity to crimes. For low-income households and for a 10% increase in violent crime, the static estimation is -\$291.14, while the corresponding dynamic estimation is -\$350.18. This is also true for air pollution.

12.6 Conclusion

The development of modelling housing in multi-dimensional frameworks (3D, 4D or more) is still in its infancy, as compared to the huge literature in a 2D framework, which explains why there are relatively few multi-dimensional housing studies. The limitation comes from the availability of data and the complexity of methods relative to time series or longitudinal dimensions. The previous papers show that both spatial and temporal dimensions in dynamic systems should be included for hedonic housing models and discrete models of residential location in a three-dimensional framework. But the inclusion of these multiple dimensions substantially complicates the specification and modeling of such systems. Extending models with unobserved neighborhood characteristics to deal with the endogenous neighborhood characteristics or introducing rationing in housing markets (see Geyer and Sieg, 2013) is not trivial.

Part of the attractiveness of a neighborhood may be due to the characteristics of neighbors (for instance, higher-income households attract higher-income households, while lower-income households repel higher-income households). As Kuminoff et al. (2013) said *"households 'sort' across neighborhoods according to their wealth and their preferences for public goods, social characteristics, and commuting opportunities ... These 'equilibrium sorting' models use the properties of market equilibria, together with information on household behavior, to infer structural parameters that characterize preference heterogeneity. These results can be used to develop theoretically consistent predictions for the welfare implications of future policy changes. Analysis is not confined to marginal effects or a partial equilibrium setting. Nor is it limited to prices and quantities... These capabilities are just beginning to be understood and used in applied research"* (p. 1007).

Over three decades, econometric methods have made significant progress and considerably improved to eliminate non-credible assumptions, such as homogenous preferences and exogenous amenities. But now, in a 2D framework, the structural estimators still rely on parametric assumptions for utility functions, on specific statistical distributions (log-normal, Type I extreme value, generalized extreme value, etc.) used to capture sources of unobserved heterogeneity and some strong assumptions to eliminate potential sources of market frictions. As suggested by Kuminoff et al. (2013), one approach could be to refine the current estimators through the lens of the econometric literature on partial identification (see Manski, 2007), which views economic models as sets of assumptions, some of which are plausible and some of which are "esoteric" (according to Tamer's (2010) expression) and are needed only to complete a model. One of the key advantages of this approach is that it could characterize the potential sensitivity of outcomes to the least credible assumptions. However, the presence of numerous latent variables, omitted variables, the definition of dynamic and spatial structures within multi-dimensional frameworks (3D, 4D or more) and the econometric complexity that results will not make things any better and must move us towards the use of flexible models and methods. One of the many other promising future pathways is probably the use of variational Bayesian approximations (see, for instance, Ormerod and Wand, 2010; Lee and Wand, 2016).

These methods facilitate approximate inference for the parameters in complex statistical models and provide fast, deterministic alternatives to Monte Carlo methods to potentially overcome many problems in the applied modelling of housing.

References

Abate, G. D. and Anselin, L. (2016). House price fluctuations and the business cycle dynamics. *Aarhus University, Dept. of Economics and Business Economics, Working Paper 2016-6*.

Agarwal, S. and Hauswald, R. (2010). Distance and private information in lending. *The Review of Financial Studies*, 23:2757–2788.

Aizcorbe, A. and Aten, B. (2004). An approach to pooled time and space comparisons. *SSHRC Conference on Index Number Theory and the Measurement of Prices and Productivity, Vancouver*.

Alonso, W. (1964). *Location and Land Use*. Harvard University Press.

Anselin, L. (2001). Spatial econometrics. In Baltagi, B. H., editor, *A Companion to Theoretical Econometrics*, pages 310–330. Blackwell.

Anselin, L., LeGallo, J., and Jayet, H. (2008). Spatial panel econometrics. In Mátyás, L. and Sevestre, P., editors, *The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice*, pages 625–660. Kluwer Academic Publishers.

Antweiler, W. (2001). Nested random effects estimation in unbalanced panel data. *Journal of Econometrics*, 101:295–313.

Bajari, P. and Kahn, M. (2005). Estimating housing demand with an application to explaining racial segregation in cities. *Journal of Business and Economic Statistics*, 23:20–33.

Baker, M. and Wurgler, J. (2007). Investor sentiment in the stock market. *Journal of Economic Perspectives*, 21:129–151.

Baltagi, B. H. and Bresson, G. (2011). Maximum likelihood estimation and Lagrange multiplier tests for panel seemingly unrelated regressions with spatial lag and spatial errors: An application to hedonic housing prices in Paris. *Journal of Urban Economics*, 69:24–42.

Baltagi, B. H., Bresson, G., and Etienne, J.-M. (2015). Hedonic housing prices in Paris: An unbalanced spatial lag pseudo-panel model with nested random effects. *Journal of Applied Econometrics*, 30:509–528.

Baltagi, B. H., Fingleton, B., and Pirotte, A. (2014). Spatial lag models with nested random effects: An IV procedure with application to English house prices. *Journal of Urban Economics*, 80:76–86.

Baltagi, B. H. and Pirotte, A. (2014). Prediction in a spatial nested error components panel data model. *International Journal of Forecasting*, 30:407–414.

Baltagi, B. H., Song, S., and Jung, B. (2001). The unbalanced nested error component regression model. *Journal of Econometrics*, 101:357–381.

Baltagi, B. H., Song, S., and Jung, B. (2002). Simple LM tests for the unbalanced nested error component regression model. *Econometric Reviews*, 21:167–187.

Bartik, T. (1987). The estimation of demand parameters in hedonic price models. *Journal of Political Economy*, 95:81–88.

Basu, S. and Thibodeau, T. (1988). Analysis of spatial autocorrelation in house prices. *Journal of Real Estate Finance and Economics*, 17:61–85.

Bayer, P., Ferreira, F., and McMillan, R. (2007). A unified framework for measuring preferences for schools and neighborhoods. *Journal of Political Economy*, 115:588–638.

Bayer, P., McMillan, R., Murphy, A., and Timmins, C. (2016). A dynamic model of demand for houses and neighborhoods. *Econometrica*, 84:893–942.

Beenstock, M. and Felsenstein, D. (2015). Estimating spatial spillover in housing construction with nonstationary panel data. *Journal of Housing Economics*, 28:1–144.

Berry, S., Levinsohn, J., and Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica*, 63:841–889.

Bhat, C. (2001). Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model. *Transportation Research Part B*, 35:677–693.

Bhat, C. (2003). Simulation estimation of mixed discrete choice models using randomized and scrambled halton sequences. *Transportation Research Part B*, 37:837–855.

Bishop, K. C. and Timmins, C. (2011). Hedonic prices and implicit markets: Estimating marginal willingness to pay for differentiated products without instrumental variables. *NBER working paper 17611*.

Blomquist, G. and Worley, L. (1981). Hedonic prices, demand for urban housing amenities and benefit estimates. *Journal of Urban Economics*, 9:212–221.

Blomquist, G. and Worley, L. (1982). Specifying the demand for housing characteristics: the exogeneity issue. In Diamond, D. and Tolley, G., editors, *The Economics of Urban Amenities*, pages 89–102. Academic Press.

Blundell, R. and Powell, J. (2004). Endogeneity in semi-parametric binary response models. *Review of Economic Studies*, 71:655–679.

Bresson, G. and Hsiao, C. (2011). A functional connectivity approach for modeling cross-sectional dependence with an application to the estimation of hedonic housing prices in paris. *Advances in Statistical Analysis*, 95:501–529.

Brown, J. and Rosen, H. (1982). On the estimation of structural hedonic price models. *Econometrica*, 50:765–768.

Carlin, B., Kogan, S., and Lowery, R. (2013). Trading complex assets. *The Journal of Finance*, 58:1937–1960.

Case, K., Shiller, R., and Thompson, A. (2012). What have they been thinking? Home buyer behavior in hot and cold markets. *NBER Working Paper No. 18400*.

Das, S., Gupta, R., and Kabundi, A. (2010). Dynamic factor models and vector autoregressive models. *Journal of Forecasting*, 30:288–302.

Davies, R. B. (1984). A generalized Beta-logistic model for longitudinal data with an application to residential mobility. *Environment and Planning, A*, 16:1375–86.

Davies, R. B. and Pickles, A. (1985). A panel study of life-cycle effects in residential mobility. *Geographical Analysis*, 17:199–216.

Deaton, A. (1985). Panel data from time series of cross-sections. *Journal of Econometrics*, 30:109–126.

Deng, Y., Gabriel, S. A., and Nothaft, F. E. (2003). Duration of residence in the rental housing market. *The Journal of Real Estate Finance and Economics*, 26:267–285.

Diao, M., Ma, X., and Ferreira, J. (2015). A real-option based dynamic model to simulate real estate developer behavior. *14th International Conference on Computers in Urban Planning and Urban Management (CUPUM), Cambridge, MA USA*.

Dieleman, F. (2001). Modelling residential mobility: A review of recent trends in research. *Journal of Housing and the Built Environment*, 16:249–265.

Eluru, N., Sener, I., Bhat, C., Pendyala, R., and Axhausen, K. (2009). Understanding residential mobility: A joint model of the reason for residential relocation and stay duration. *Transportation Research Board 88th Annual Meeting, Washington, DC*.

Engle, R. F., Lilien, D. M., and Watson, M. (1985). A DYMIMIC model of housing price determination. *Journal of Econometrics*, 28:307–326.

Epple, D. (1987). Hedonic prices and implicit markets: Estimating demand and supply functions for differentiated products. *Journal of Political Economy*, 95:59–80.

Fingleton, B. (2008). Housing supply, housing demand and affordability. *Urban Studies*, 45:1545–1563.

Fingleton, B. and LeGallo, J. (2008). Estimating spatial models with endogenous variables, a spatial lag and spatially dependent disturbances: Finite sample properties. *Papers in Regional Science*, 87:319–339.

Fu, Q., Zhu, Y., and Ren, Q. (2015). The downside of marketization: A multilevel analysis of housing tenure and types in reform-era urban China. *Social Science Research*, 49:126–140.

Gayer, T., Hamilton, J., and Viscusi, W. (2000). Private values of risk tradeoffs at superfund sites: Housing market evidence on learning about risk. *Review of Economics and Statistics*, 82:439–451.

Geyer, J. and Sieg, H. (2013). Estimating a model of excess demand for public housing. *Quantitative Economics*, 4:483–513.

Glaeser, E. (2008). *Agglomeration and Spatial Equilibrium*. Oxford University Press.

Griliches, Z. (1971). *Price Indexes and Quality Change, Studies in New Methods of Measurement*. Harvard University Press.

Guevara, C. A. and Ben-Akiva, M. (2006). Endogeneity in residential location choice models. *Transportation Research Record*, 1977:60–66.

Hastings, J., Kane, T., and Staiger, D. (2006). Parental preferences and school competition: Evidence from a public school choice program. *NBER Working Papers 11805*.

Heckman, J. (1978). Dummy endogenous variables in a simultaneous equation system. *Econometrica*, 46:931–959.

Hoang, H. and Wakely, P. (2000). Status, quality and the other trade-off: Towards a new theory of urban residential location. *Urban Studies*, 37:7–35.

Holly, S., Pesaran, M. H., and Yamagata, T. (2010). A spatio-temporal model of house prices in the USA. *Journal of Econometrics*, 158:160–173.

Holmes, T. and Sieg, H. (2014). Structural estimation in urban economics. In Duranton, G., Henderson, J. V., and Stran, W. C., editors, *Handbook of Regional and Urban Economics, vol.5*, pages 69–114. Elsevier, North-Holland.

Huang, Y. and Clark, W. A. V. (2002). Housing tenure choice in transitional urban China: A multilevel analysis. *Urban Studies*, 39:7–32.

Judd, K. (1998). *Numerical Methods in Economics*. MIT Press.

Jun, H.-J. (2016). The effect of racial and ethnic composition on neighborhood economic change: A multilevel and longitudinal look. *Housing Studies*, 31:102–125.

Kelejian, H. and Prucha, I. (1998). A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *Journal of Real Estate Finance and Economics*, 17:99–121.

Kelly, B. and Ljungqvist, A. (2012). Testing asymmetric-information asset pricing models. *The Review of Financial Studies*, 25:1366–1413.

Kuminoff, N. V., Smith, V. K., and Timmins, C. (2013). The new economics of equilibrium sorting and policy evaluation using housing markets. *Journal of Economic Literature*, 54:1007–1062.

Latif, E. (2015). Immigration and housing rents in canada: A panel data analysis. *Economic Issues*, 20:91–108.

Lee, B. H. Y. and Waddell, P. (2010). Residential mobility and location choice: A nested logit model with sampling of alternatives. *Transportation*, 37:587–601.

Lee, C. Y. and Wand, M. P. (2016). Streamlined mean field variational Bayes for longitudinal and multilevel data analysis. *Biometrical Journal*, 4:868–895.

Lee, L. (2003). Best spatial two-stage least square estimators for a spatial autoregressive model with autoregressive disturbances. *Econometric Reviews*, 22:307–335.

LeSage, J. and Pace, R. (2009). *Introduction to Spatial Econometrics*. CRC Press/Taylor & Francis.

Li, Y. and Leatham, D. J. (2011). Forecasting housing prices: Dynamic factor model versus LBVAR model. *Agricultural & Applied Economics Association's 2011 AAEA and NAREA Joint Annual Meeting, Pittsburgh, Pennsylvania*.

Liu, X. (2013). Spatial and temporal dependence in house price prediction. *Journal of Real Estate Finance and Economics*, 47:341–369.

Manski, C. (2007). *Identification for Prediction and Decision*. Harvard University Press.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. In Zarembka, P., editor, *Frontiers in Econometrics*, pages 105–142. Academic Press.

McFadden, D. (1974). The measurement of urban travel demand. *Journal of Public Economics*, 3:303–328.

McFadden, D. (1978). Modelling the choice of residential location. In Karlqvist, A., Lundqvist, L., Snickars, F., and Weibull, J., editors, *Spatial Interaction Theory and Planning Models*, pages 75–96. North Holland.

Meese, R. and Wallace, N. E. (1997). The construction of residential housing price indices: A comparison of repeat-sales, hedonic-regression and hybrid approaches. *Journal of Real Estate Finance and Economics*, 14:51–73.

Nappi-Choulet, I. and Maury, T. (2011). A spatial and temporal autoregressive local estimation for the Paris housing market. *Journal of Regional Science*, 51:732–750.

Nelson, J. (1977). Residential choice, hedonic prices, and the demand for urban air quality. *Journal of Urban Economics*, 5:357–369.

Newey, W. K. and McFadden, D. (1974). Large sample estimation and hypothesis testing. In Engle, R. F. and McFadden, D., editors, *Handbook of Econometrics*, volume 4, pages 2111–2245. Elsevier, North-Holland.

Ormerod, J. and Wand, M. (2010). Explaining variational approximations. *The American Statistician*, 64:140–153.

Pace, P., Barry, R., Gilley, O., and Sirmans, C. (2000). A method for spatial-temporal forecasting with an application to real estate prices. *International Journal of Forecasting*, 16:229–246.

Rosen, S. (1974). Hedonic prices and implicit markets. *Journal of Political Economy*, 82:34–55.

Rossi, P. (1955). *Why Families Move: A Study in the Social Psychology of Urban Residential Mobility*. The Free Press, Glencoe, Illinois.

Smith, T. and Wu, P. (2011). A spatio-temporal model of housing prices based on individual sales transactions over time. *Journal of Geographical Systems*, 11:333–355.

Syed, I., Hill, R. J., and Melser, D. (2008). Flexible spatial and temporal hedonic price indexes for housing in the presence of missing data. *The University of New South Wales, School of Economics, Discussion Paper: 2008/14*.

Sykes, L. L. (2003). Income rich and asset poor: A multilevel analysis of racial and ethnic differences in housing values among baby boomers. *Population Research and Policy Review*, 22:1–20.

Tamer, E. (2010). Partial identification in econometrics. *Annual Review of Economics*, 2:167–195.

Thanos, S., Dubé, J., and Legros, D. (2016). Putting time into space: The temporal coherence of spatial applications in the housing market. *Regional Science and Urban Economics*, 58:78–88.

Turnbull, G. K. and van der Vlist, A. J. (2015). Uninformed house buyers and foreclosures. *University of Central Florida, Department of Finance, Working Paper 1503*.

Witte, A., Sumka, H., and Erekson, H. (1979). An estimate of a structural hedonic price model of the housing market: An application of Rosen's theory of implicit markets. *Econometrica*, 47:1151–1173.

Chapter 13
Modelling Migration

Raul Ramos

Abstract The use of three-dimensional models as a tool to analyse migration flows has substantially increased during the last decade. These models are very popular in analysing economic phenomena related to the movement of goods and services, capital and people. In fact, while gravity models have been used extensively in the analysis of trade determinants, only due to the recent availability of time-varying origin-destination migration data, have multi-dimensional panel data models been applied in order to improve our understanding of migration flows. Recent contributions have also provided the micro-foundations of gravity models in the context of migration analysis and, nowadays, the literature is expanding by considering the role of different factors and policies under the framework provided by Random Utility Maximization (RUM) models. The aim of this chapter is to provide a survey of recent developments in multi-dimensional panel data models in the context of migration analysis, to identify the problems that researchers face when using these models and to review potential new directions in this field of research.

13.1 Introduction and Objectives

As shown in Chap. 11, multi-dimensional data in the context of international economics has been mainly used to analyse the determinants of international trade. Only due to the recent availability of time-varying origin-destination migration data, have multi-dimensional panel data models been applied in order to improve our understanding of the drivers of international migration flows.

Earlier research related bilateral migration to the relative size of the origin and destination countries (measured by population or GDP) and the distance between them, a specification that clearly resembles Newton's 1687 law of gravity. However,

Raul Ramos
AQR-IREA, Universitat de Barcelona, and IZA, Barcelona, Spain

L. Matyas (ed.), *The Econometrics of Multi-dimensional Panels*, Advanced Studies in Theoretical and Applied Econometrics 50, DOI 10.1007/978-3-319-60783-2_13

as richer datasets have become available, simpler specifications of the gravity models have beeen enlarged with variables related to different migration pull and push factors; for instance, better economic opportunities in the destination country, safer conditions, or higher political freedom, among others (see Beine et al., 2016 for a review).

The aim of this chapter is to provide a survey of recent developments in multidimensional panel data models in the context of migration analysis, to identify some of the problems that researchers face when using these models and to review potential new directions in this field of research.

The chapter is structured in three sections. First, Sect. 13.2, briefly describes Random Utility Maximization (RUM) models that have provided the micro-foundations of gravity models in the context of migration analysis. Second, Sect. 13.3 focuses on some of the problems and challenges that researchers face when estimating multidimensional panels to analyse the determinants of migration flows. In particular, the section starts describing data limitations and considering how the transformation of the multiplicative specification of RUM based models into a log-linearised specification introduces the problem of how to deal with the potential presence of zeros. Next, potential ways of dealing with multilateral resistance to migration through the inclusion of different types of fixed effects are described. The section ends with a discussion on how the migration literature has dealt with endogeneity and spatial interactions in the context of multi-dimensional panels. Last, Sect. 13.4 provides some concluding remarks.

13.2 Micro-foundations of the Gravity Model of Migration

Following Beine et al. (2016), the Random Utility Maximization (RUM) model describes the utility that individual i located in country j at time $t-1$ derives from moving to country k that belongs to a particular choice set D at time t

$$U_{ijkt} = w_{jkt} - c_{jkt} + \varepsilon_{ijkt} \,, \tag{13.1}$$

where w_{jkt} represents a deterministic component of utility, c_{jkt} is the cost of moving from j to k between $t-1$ and t, and ε_{ijkt} is an individual specific stochastic term. When we assume that ε_{ijkt} follows an iid extreme value distribution as in Grogger and Hanson (2011),[1] we can apply the results in McFadden (1974) and write the expected probability that individual i opts for country k as

$$\mathrm{E}(p_{jkt}) = \frac{e^{w_{jkt}-c_{jkt}}}{\sum_l w_{jlt} - c_{jlt}} \,, \tag{13.2}$$

[1] As shown by Beine et al. (2016), similar results are obtained under alternative distributional assumptions.

where l represents any country belonging to the choice set D. Assuming that the deterministic component of utility does not vary with the origin j, this allows us to define the expected gross migration flow from country j to k, m_{jkt}, as the product of the stock of population in j at time t, s_{jt}, by the expected probability $\mathrm{E}(p_{jkt})$

$$\mathrm{E}(m_{jkt}) = s_{jt} \frac{e^{w_{kt}}}{e^{c_{jkt}}} \frac{1}{\sum_l w_{lt} - c_{jlt}} \,. \tag{13.3}$$

Specification (13.3) clearly looks like the traditional gravity model, as expected gross migration flows from one country to any other depend in a multiplicative way on origin characteristics, the attractiveness of the destination, and they are inversely related to the accessibility of the destination for potential migrants (the costs of moving from j to destination k). As highlighted by Poot et al. (2016), in the most commonly applied form, the gravity law of population usually considers population at origin and destination as the main factors related to attractiveness, while distance captures accessibility. However, (13.3) also includes a last term related to the attractiveness of this destination compared to the different alternatives in the choice set D including the option of not migrating. This last term is related to multilateral resistance to migration, the influence that the attractiveness of alternative destinations exerts on bilateral migration flows (Bertoli and Fernández-Huertas Moraga, 2013).

However, the simplest empirical specification of (13.3) that has been used in several applications is the following one

$$m_{jkt} = \frac{\beta_j' x_{jt} \beta_k' x_{kt}}{\gamma_{jk}' x_{jkt}} \varepsilon_{jkt} \,, \tag{13.4}$$

where x_{jt} are time-varying origin characteristics, x_{kt} time-varying destination characteristics and x_{jkt} a measure of the accessibility of destination k for residents in j. Last, ε_{jkt} is an error term with an expected value equal to 1. In (13.4) the multilateral resistance term is ignored, an omission that generates biases in the estimation of the coefficients of the determinants of migration. Different strategies have been developed in order to control for multilateral resistance to migration, an issue that will be further developed in Sect. 13.3.

In addition, it is important to highlight that one relevant assumption of the RUM model is that the attractiveness of a destination is not supposed to be affected by migration (Ramos, 2016). For instance, if one particular destination is attractive due to its low levels of unemployment when compared to a particular origin, massive inflows of immigrants could increase unemployment in the destination while at the same time decreasing it in the origin country. Empirical specifications based on RUM models cannot capture these second-round effects, which is an important point to consider in order to appropriately interpret the results. In any case, RUM models provide an appropriate theoretical justification of the intuition behind empirical models used in this field. The use of RUM models makes clear what the assumptions made by researchers are, and how these assumptions yield different empirical specifications.

13.3 Data Limitations and Estimation Issues

Some of the problems and challenges that researchers face when estimating multi-dimensional panels to identify the determinants of migration flows have already been discussed in Chap. 11 for trade analysis. The transformation of the multiplicative specification in (13.4) into a log-linearised specification not only introduces the problem of how to deal with the potential presence of zeros, but could also lead to heteroskedasticity. While the first point is discussed below for migration analysis, regarding heteroskedasticity we refer the reader to the specific Sect. in Chap. 11. The inclusion in the model of different types of fixed effects to control for multilateral resistance and how to deal with endogeneity and spatial interactions are other relevant aspects that have already been developed in Chaps. 9 and 11, but that are discussed below for the specific case of international migration models.[2]

13.3.1 Data and Measurement Issues

Until recently, the analysis of the determinants of international migration mainly focused on migrants arriving to and from OECD countries due to data availability. However, the publication of global bilateral migration matrices by the World Bank and the United Nations have opened new possibilities to researchers. In particular, as described in Özden et al. (2011), matrices of bilateral migrant stocks for 226 countries spanning 1960–2000 and disaggregated by gender are provided by the World Bank.[3] More than one thousand census and population register records were combined to construct decennial matrices of stocks corresponding to the five census rounds between 1960 and 2000 based primarily on the foreign-born definition of migrants.

However, as we have seen in Sect. 13.2, RUM models provide the basis to analyse migration flows and rather than stocks, and it is quite clear that variations in stocks are subjected to measurement errors and cannot be used as a proxy for gross flows. In fact, variations in stocks are influenced by return migration or migration to third countries, and, as a result, negative values could be obtained. For this reason, some researchers have opted to apply indirect methods to obtain estimates of flows. For instance, Abel and Sander (2014) obtained Maximum Likelihood estimates for the

[2] This chapter focuses on the analysis of international movements of people. The interested reader in the use of multi-dimensional panels to explain internal migration flows can consult Cushing and Poot (2004), Faggian et al. (2015) and Poot et al. (2016).

[3] Efforts have also been devoted to improve the information regarding usual controls in gravity models. For instance, CEPII's GeoDist database (Mayer and Zignago, 2011) provides data on different measures of bilateral distances for 225 countries. It also incorporates country-specific geographical variables, including capital cities coordinates, languages spoken in the country, a variable indicating whether the country is landlocked, and colonial links. CEPII Gravity dataset (Head et al., 2010) adds some additional time-varying variables to the GeoDist data set: GDP, population, and other institutional variables, such as regional trade agreements or currency unions.

number of movements required to meet the changes over time in migrant stock data using an iterative proportional fitting algorithm. Other researchers, such as Grogger and Hanson (2011), have opted to use migration stocks and interpret their results as a representation of long-term equilibrium, although their analysis would not be compatible with the theoretical basis provided by RUM models.

The literature also recognises that the key driving forces behind migration can be very different for economic-based movements and forced displacements. The United Nations High Commissioner for Refugees (UNHCR) has also compiled a database on the number of refugees and asylum seekers according to their countries of asylum/residence and origin. Barthel and Neumayer (2015) and Echevarria and Gardeazabal (2016) use this dataset in order to analyse forced migration flows. While Barthel and Neumayer (2015) focus on asylum migration to developed countries, Echevarria and Gardeazabal (2016) study global flows. The covariates used in both articles include additional determinants in the gravity specification than the ones usually employed. In particular, the existence of armed conflict and the lack of democracy or civil liberties seem to play central roles in the determination of forced migration, aspects that are often difficult to conceptualize and measure.

Last, a new emerging literature also focuses on the analysis of temporary migration flows. As Dustmann and Görlach (2016) emphasize, many migrations are not permanent: ten years after arrival, close to 50 percent of the original arrival cohort leaves the destination country in the case of Europe and 20 percent in the case of the United States and Canada, but also in Australia and New Zealand. However, data limitations are (again) an important barrier as it is common to register new immigrants but not even feasible to register out-migration. The solution to this problem has consisted in combining multiple data sources at destination and origin countries, but also compiling information about intentions to return through specific surveys, as in Artuç and Özden (2016), to explain transit migration to the United States. The more than probable availability of multi-dimensional panels in the near future will allow researchers to analyse the patterns of repeated and circular migration and other types of temporary movements of people like international students (Beine et al., 2014, Perkins and Neumayer, 2014 and Abbott and Silles, 2016), tourists (Eilat and Einav, 2004, Morley et al., 2014, among others) and high-mobile inventors (Fink et al., 2017).

13.3.2 Missing and Incomplete Data

Although as previously explained, the availability and quality of migration data has significantly improved, in almost all cases collected datasets are unbalanced (a common feature of the empirical applications covered in this book – see Chap. 1). However, in the specific case of migration analysis, researchers have faced the additional difficulty that censuses are only available on a decennial basis, while data related

to the main factors of interest in the analysis usually refer to relatively short time periods, but they are available on an annual basis.[4]

For this reason, several authors have used interpolation methods to overcome this difficulty.[5] In some cases, interpolation affects a small number of cases, as in Ortega and Peri (2013), where they interpolate observations to fill in missing values in intermediate years. They apply a linear interpolation method when a data point for a bilateral migration flow is missing and both the previous and following years are available. A similar approach is adopted by Docquier and Bhargava (2007), who build an annual panel data set on medical brain drain from all countries in the world from 1991 to 2004. The medical brain drain is defined as the proportion of physicians trained in their country and working abroad. Annual data on foreign-trained doctors were available from national medical associations in the most important host countries, representing 75 per cent of their sample, but they need to interpolate 10-year or 5-year data where the data source used was national censuses. In this case, they used a log-linear adjustment to obtain annual data from the two or three data points available. For domestic physicians, only a few missing observations were found and they interpolated them on the basis of regional trends or using the figures obtained in neighbouring countries with similar economic and health records.

In other cases, more sophisticated interpolation methods are applied to obtain annual estimates of migration flows or stocks for the period between Census years. For instance, in order to obtain annual estimates of the foreign-born stock by country, Clark et al. (2007) interpolate between the benchmarks established from the census using a stock adjustment equation: $S_{t+1}=M_t+dS_t$, where S_t is the stock at the beginning of year t, and M_t is the flow during that year. The parameter d reflects deaths, return migration, and illegal immigration, which subtract or add to the stock independently of the additions through gross immigration. This depreciation rate is calculated for each interval between censuses using an iterative procedure beginning with S_t, such that the value of S_{t+10} obtained by cumulating forward is reconciled with that of the next census benchmark. This procedure allows to calculate a different value of d for each country for each interval between benchmarks. Variants of this method are still applied nowadays. Fitzgerald et al. (2014) apply a similar method but they estimate the parameter d by non-linear least squares.

13.3.3 Logs and Zeros

Taking as a starting point the specification in (13.4), researchers usually estimate a log-linearised form of this model:

[4] Baltagi et al. (2015) focus on the implications of non-random missing values in multi-dimensional panels for trade. In this case, sample selection models or two-part models should be used. However, this problem does not seem to be as frequent in the case of migration analysis as it is in trade.

[5] One notable exception is Llull (2016), who prefers to work with grouped data for missing pairs rather than interpolating country-specific values.

$$\ln[m_{jkt}] = \beta_j' \ln[x_{jt}] + \beta_k' \ln[x_{kt}] - \gamma_{jk}' \ln[x_{jkt}] + \varepsilon_{jkt} \ . \tag{13.5}$$

However, as already explained in Chap. 11 in the context of trade, one problem with this approach is how to deal with the potential presence of zero values for bilateral flows (or even negative values as due to data problems some researchers work with stocks or net flows, the variation in stocks). As in the case of trade, when analysing migration, most researchers exclude these values from the sample but others prefer to correct them as in Beine et al. (2016) or Figueiredo et al. (2016).[6] However, by deleting zero flows, relevant information on the pair of countries where there are no migratory movements is not taken into account. For these reasons, and similarly to the trade literature, migration researchers are considering two alternative procedures: first, to use count data models such as Poisson, negative binomial and zero-inflated models and, second, to apply Heckman's selection model in order to correct for the probability of migration in (13.5). However, in the case of migration, Poisson Pseudo Maximum Likelihood (PPML) has become the standard procedure in the literature (Santos Silva and Tenreyro, 2006) hereby avoiding the need to take natural logarithms. The problem with this approach is that it tends to over-weight high flows, and the estimation procedure might face problems of convergence towards the optimal values of the parameters (Burger et al., 2009).

Regarding Heckman's selection model, the two-step procedure is only applied by a reduced number of authors (opposed to the case of trade).[7] The first step involves the estimation of the selection equation, a probit model for a given country pair to have a positive migration flow. The usual procedure implies the use of an instrument in the probit equation, i.e., a bilateral variable that influences the probability of observing a positive migration flow between the two countries but does not influence the size of the flow, although it is not strictly required (Puhani, 2000). However, as highlighted by Beine et al. (2011), the drawback of not using an additional instrument in the selection equation is that the Mills' ratio might become highly collinear with the explanatory variables of the flow equation, which in turn lowers the significance of the coefficients. In the second step, (13.5) is enlarged with the inverse of the Mills' ratio to account for selection

$$\ln[m_{jkt}] = \beta_j' \ln[x_{jt}] + \beta_k' \ln[x_{kt}] - \gamma_{jk}' \ln[x_{jkt}] + \lambda \widehat{mills}_{jkt} + \varepsilon_{jkt} \ . \tag{13.6}$$

Using data on migration flows disaggregated by gender and educational attainment from 195 source countries to OECD countries, Beine et al. (2011) use the existence or not of diplomatic representation among the countries considered as an instrument to explain positive migration flows. Their argument is that in the absence of any diplomatic representation of country k in country j, the cost of obtaining a visa could discourage citizens of country j from trying to migrate to country k. However, while it seems clear that this variable could affect the probability of initial

[6] Adding a positive number is also problematic as small variations in the selected number produce big variations in the results, as demonstrated by Burger et al. (2009).

[7] As highlighted by Baltagi et al. (2015), a nonparametric control function approach can also be applied, as suggested by Cameron and Trivedi (2005).

migration, it does not necessarily modify the magnitude of the flows between the two countries.

13.3.4 Multilateral Resistance to Migration

As previously mentioned in Sect. 13.2, a relevant aspect when specifying a gravity model is to consider multilateral resistance to migration. This term is related to the influence of third countries in determining migration flows between two particular countries. Following the explanation by Krugman (1995) regarding trade flows and introduced in the migration literature by Bertoli and Fernández-Huertas Moraga (2013), if two European countries were moved to Mars, migration flows between them would clearly increase, due to the current lack of alternative destinations, although their relative characteristics remain unchanged. Therefore, not considering the influence of alternative destinations could bias the results of the analysis. For instance, if there is some degree of coordination in migration policies in destination countries, controlling for multilateral resistance to migration tends to find much larger effects of these policies than the ones that would have been obtained without controlling for it.

Earlier attempts to control for multilateral resistance to migration consisted in including origin-year fixed effects, α_{jt}, as in Ortega and Peri (2013). In particular, (13.5) is modified as follows

$$\ln[m_{jkt}] = \alpha_{jt} + \beta_k' \ln[x_{kt}] - \gamma_{jk}' \ln[x_{jkt}] + \varepsilon_{jkt} \,. \tag{13.7}$$

Origin-year fixed effects capture all terms that vary across origin and year but not across destinations. This identification strategy is consistent with their underlying RUM model and, as highlighted by Bertoli and Fernández-Huertas Moraga (2013), "the underlying pattern of substitution across alternative locations is richer than in the traditional approach: an increase in the attractiveness of destination l can draw from another destination k more than it does from the origin country j, so that the bilateral migration rate m_{jkt} falls". However, it is not clear that this solution would be enough to account for all potential sources of unobserved heterogeneity.

According to Beine and Parsons (2015), the inclusion of destination-year fixed effects, α_{kt} in (13.5) is a valid approach to account for multilateral resistance in destination countries as the migration policies are defined at that level

$$\ln[m_{jkt}] = \alpha_{kt} + \beta_j' \ln[x_{jt}] - \gamma_{jk}' \ln[x_{jkt}] + \varepsilon_{jkt} \,. \tag{13.8}$$

More complex fixed effects specifications have also been adopted in the literature (Bertoli and Fernández-Huertas Moraga, 2013).[8] For instance, extending to

[8] Although it is common practice in other fields of research to include different types of fixed effects from the initial specification, migration researchers have tried to link their empirical specification with the one derived from their RUM-based model. It is worth mentioning that control-

the multiple origin and multiple destination case, the specification in Bertoli et al. (2016) would yield the following expression

$$\ln[m_{jkt}] = \alpha_{jk} + \lambda_t + \beta_j' \ln[x_{jt}] - \gamma_{jk}' \ln[x_{jkt}] + \varepsilon_{jkt} \,, \tag{13.9}$$

where α_{jk} are time-invariant fixed effects for all country pairs considered in the analysis, also called dyadic fixed effects, and λ_t are time fixed effects that are also introduced to account for common shocks in origin and destination countries. However, one inconvenience of this approach is that the parameters of time-invariant dyadic explanatory factors of migration flows such as distance, common language or common border, among others, can no longer be identified. Moreover, as suggested by Bertoli and Fernández-Huertas Moraga (2013) and Bertoli et al. (2016), the Common Correlated Effects (CCE) estimator by Pesaran (2006) can also be applied providing some advantages when compared to (13.9). In particular, it is possible to identify the effect of determinants of bilateral migration flows that are specific to each origin or destination country. It is not required to have data on multiple origins and multiple destinations to be able to control for multilateral resistance to migration, although it cannot be applied in short panels. As demonstrated by Bertoli et al. (2016), a last advantage of the CCE estimator is that "it delivers unbiased estimates even in the presence of the confounding influence exerted by the future attractiveness of the various options in the choice set, while the fixed effects specification, does not". This method consists in introducing the cross-sectional averages of the dependent and independent variables as additional regressors together with dyadic fixed-effects as in Bertoli and Fernández-Huertas Moraga (2013). In particular, the following regression can be estimated

$$\ln[m_{jkt}] = \alpha_{jk} + \lambda_t + \beta_j' \ln[x_{jt}] - \gamma_{jk}' \ln[x_{jkt}] + \theta_{jk}' \tilde{z}_t + \varepsilon_{jkt} \,, \tag{13.10}$$

where $\tilde{z}_t$ represents the vector of auxiliary regressors formed by the cross-sectional averages of migration flows and all the regressors.

Although the use of fixed effects has been predominant in the literature, other authors, such as Egger (2000), have defended the use of random effects models.[9] As highlighted by Shepherd (2016) in relation to trade analysis, this approach is only consistent under restrictive assumptions as to the pattern of unobserved heterogeneity in the data. In the context of RUM models, accounting for both origin and destination multilateral resistance requires the specification of a two-dimensional random effects model and to assume that multilateral resistance is normally distributed. For this reason, dyadic random effects are usually only estimated for comparison with fixed effects models (Baltagi et al., 2015).

---

ling for multilateral resistance to migration by the inclusion of different types of fixed effects can also capture a large part of the omitted factors making instrumentation unnecessary unless reverse causality is also present (Beine et al., 2016).

[9] Baltagi et al. (2015) discuss the estimation of fixed effects versus random effects models in the context of panel data gravity models of international trade.

13.3.5 Endogeneity

Another challenge that researchers in the field have tried to solve is related to the potential presence of endogeneity of some of the regressors. In order to solve this problem, one possibility would be to apply instrumental variables estimators, although as usual, it is quite difficult to find appropriate instruments. We need to find variables related to the policy we want to analyse but at the same time they should not be correlated to the rest of the regressors in the model. For this reason, the use of internal instruments, such as past bilateral flows, does not always solve the problem, and the identification of external instruments is always difficult unless there are some historical events that can help us to identify the instrumental variable.

Beine et al. (2014) consider the role of migration networks to explain the mobility of international students, using data for 13 OECD countries. Networks are proxied by the stock of educated migrants in destination countries at the beginning of the period considered. The idea is that students engaged in higher education benefit from the support of skilled migrants in the destination country. However, if the destination countries favour migration from some particular origin countries, it stands to reason that they will also favour the arrival of students from those countries. As a result, the observed positive impact of networks on migration flows will actually be the result of students simply following the general pattern of economic migrants or those affected by family reunification programs that come from the same origin countries as the students. In order to disentangle the effects of networks on flows, researchers use the following instrument: the existence of guest worker programs after the Second World War that attracted economic migrants to work in some specific industries, like coal mines or steel factories. These guest worker agreements led to important diasporas in the destination countries, and are good independent predictors of migrant networks. The results of using this instrument still support the positive impact of networks on international student flows. The second variable upon which they apply an instrument is enrolment fees. According to their initial results, there is a positive correlation between higher fees in destination countries and higher flows of international students. Although this unexpected result could be explained as a signal for the presence of higher quality education in the destination country, it could also be related to reverse causality. Those universities that are more attractive for international students can afford to charge them higher fees. As this policy can be easily implemented by private universities, researchers use the following instrument: the private sector's share of total expenditures in the higher education systems in destination countries, a variable that is related to the capacity of universities to charge higher fees, but not necessarily explaining international student arrivals. When using this instrument, they obtain no significant effect of fees on flows, a result that could also be related to the existence of grants for international students.

An additional example in this context relates to the analysis of visa policies and irregular immigrant flows by Czaika and Hobolth (2016). The authors' objective is to analyse if restrictive asylum and visa policies in Europe have triggered an unintended behavioural response of potential and rejected asylum seekers shifting into

irregular migrants (deflection effect). In particular, although more restrictive asylum policies might discourage potential immigrants from entering the country, it is also likely that this policy has the unintended effect of pushing asylum seekers into an irregular status. With the aim of testing if the European asylum and visa policies have triggered this unintended effect, they estimate a three-dimensional gravity model using Eurostat's data on irregular migrant flows into 29 European states from more than 200 origin countries over the period 2008 to 2011. As their policy variables can be potentially endogenous either due to reverse causality or an omitted variable bias, they employ an instrumental variable approach using as a first instrument the share of the Muslim population in origin countries. They assume this instrument to be relevant in order to define bilateral asylum and visa policies due to security and terrorism concerns from Muslim countries after the September 11 2001 attacks. As a second instrument, they employ the size of the informal sector in the origin country. The results obtained support the hypothesis that a significant number of rejected asylum seekers seem to opt for the irregular stay option: a 10 percent increase in rejected asylum seekers results in a 2 percent increase in on-territory arrests of irregular immigrants when using PPML and in a 7 percent increase according to IV-GMM estimates. In both cases, origin and destination fixed effects have been considered.

13.3.6 Spatial Models

Chap. 9 has already shown how the increasing availability of spatial data at different levels has allowed to consider new theoretical and empirical analysis in the context of multi-dimensional nested spatial panel data.

The literature on migration has also advanced in this context from two perspectives that are not explicitly considered in Chap. 9. First, researchers have been concerned about the validity of their analysis if flows are not spatially independent. A failure to account for spatial dependence may lead to biased parameter estimates and misleading inferences. One solution to this problem, originally developed by Griffith (2003), consists in the use of eigenvector spatial filtering variants of multi-dimensional Poisson or negative binomial gravity models along with pseudo ML estimators as suggested by Chun and Griffith (2011). Eigenvector spatial filtering relies on the spectral decomposition of the spatial weight matrix W into eigenvalues and eigenvectors, and then using a subset of the eigenvectors as additional regressors in the gravity model. In particular, spatial filtering uses the spectral decomposition of the spatial weight matrix W

$$W = E\Lambda E', \tag{13.11}$$

where E is a matrix of eigenvectors and Λ is a matrix containing the corresponding eigenvalues on the diagonal. For multi-dimensional models, the eigenvectors need to be concatenated T times so that the resulting eigenvectors match the total number of observations. This process is based on the assumption that the underlying spatial

process does not change over time. Eigenvector selection is then undertaken with stepwise regression procedures, and the Q selected eigenvectors are then used as additional regressors in (13.5)

$$\ln[m_{jkt}] = \beta_j' \ln[x_{jt}] + \beta_k' \ln[x_{kt}] - \gamma_{jk}' \ln[x_{jkt}] + \sum_{q=1}^{Q} E_q \theta_q + \varepsilon_{jkt} \ . \tag{13.12}$$

Term $\sum_{q=1}^{Q} E_q \theta_q$ is known as a spatial filter and it reduces the potential bias in the parameter estimates associated to origin-destination spatial dependence.

The second direction that migration researchers have followed in this context has been to apply spatial models considering "substantive" spatial dependence in a similar vein to Fingleton and Lopez-Bazo (2006) regarding spatial interactions in growth models. While most specifications described below and the estimation methods used resemble those described in Chap. 9, the main difference is the way authors justify the inclusion of spatial spillovers.

In particular, Neumayer and Plümper (2010) introduce five different options for modelling spatial dependence when analysing "directed" dyadic data. According to these authors, in directed dyadic data, two actors, j and k, have an asymmetric interaction and one can distinguish dyad jk, where in the case of migration, j would be the sending country and k the receiving one, from dyad kj, where these roles are reversed. Although these authors define different types of contagion effects in a cross-sectional setting, their proposal can be easily extended to multi-dimensional gravity models, as also developed in the case of trade in Chap. 11. In particular, Neumayer and Plümper (2010) define *directed dyad contagion* as follows

$$m_{jkt} = \rho \sum_{ml \neq jk} \omega \, m_{mlt} + \varepsilon_{jkt} \ , \tag{13.13}$$

where, following a similar notation as the one used in Chap. 9, ρ is the spatial lag parameter to be estimated, and ω is a time-invariant spatial weight matrix. (13.13) describes a situation where migration flows between j and k at time t depend on the weighted sum of all other migration flows between source countries m and target countries l. In the remaining four forms of contagion, Neumayer and Plümper (2010) assume that only a subset of the sources or target flows could be relevant for the particular dyad. When flows between the dyad jk spatially depend only on the aggregate flows of the other sources m ($\neq j$), that is on their relationship with all other targets l and not just the specific target k, they define *aggregate source contagion* as

$$m_{jkt} = \rho \sum_{m \neq j} \sum_{l} \omega \, m_{mlt} + \varepsilon_{jkt} \ . \tag{13.14}$$

Alternatively, when flows between the dyad jk spatially depend on the aggregate flows of the other targets l with all other sources m, then this leads to *aggregate target contagion*

$$m_{jkt} = \rho \sum_{m} \sum_{l \neq k} \omega \, m_{mlt} + \varepsilon_{jkt} \ . \tag{13.15}$$

The next possibility is that, instead of spatially depending on the aggregate choices of all sources and targets, migration flows between j and k depend on the choices of sources in relation to the specific dyad under consideration, a situation that Neumayer and Plümper (2010) term *specific source contagion*

$$m_{jkt} = \rho \sum_{m \neq j} \omega m_{mkt} + \varepsilon_{jkt} \ . \tag{13.16}$$

In comparison to the aggregate source contagion, (13.16) represents a situation in which other sources m affect j's spatial interaction with k only if countries m have non-zero migration flows with k.

The last form of contagion introduced by Neumayer and Plümper (2010) is the *specific target contagion*, a situation in which other targets l affect k interaction with j only if countries l have non-zero migration flows with country j

$$m_{jkt} = \rho \sum_{l \neq k} \omega m_{jlt} + \varepsilon_{jkt} \ . \tag{13.17}$$

The different possible spatial lags of directed dyadic data described in (13.13)–(13.17) could be combined into a single equation and different spatial weight matrices could also be introduced for the different type of spatial interactions (Lesage and Pace, 2008; Neumayer and Plümper, 2016). However, Neumayer and Plümper (2010) recommend that "researchers should not mine the data for potential evidence of all types of contagion but test only those types of contagion specified by their theory".

Properly identifying the spatial weight matrix is also an important point. Neumayer and Plümper (2016) argue that the specification of ω should take into account the causal mechanism through which spatial dependence works, while most applied researchers often use proxies for connectivity, such as geographical contiguity or proximity. Moreover, as shown in Chap. 9 it is common practice by spatial econometricians to row-standardise the spatial weight matrix,[10] but Plümper and Neumayer (2010) demonstrate that row standardisation changes the relative influence of other units on the spatial effect, thereby altering the estimation results. For this reason, they argue that this procedure should not be applied by default unless there are theoretical justifications to do so. Plümper and Neumayer (2010) also demonstrate that the functional form of the spatial weight matrix for matrices depending on continuous variables, like distance, also matters: small changes can lead to very different results. For instance, it is not the same to use the inverse of distance, the inverse of squared distance or a dichotomous weighting matrix in which units beyond a certain threshold distance (or bandwith) exert no influence at all. Last, according to Neumayer and Plümper (2016), it is also important not to assume that there is only a unique causal mechanism to explain spatial dependence. They argue that the

[10] In a row-standardised matrix, the sum of all cells in the same row adds up to one (each cell of the original matrix has been divided by its row sum). When using the row-standardised matrix, the spatial lag is a weighted average of the lagged dependent variable. If the spatial weight matrix is not row-standardised, then the spatial lag is a weighted sum of the lagged dependent variable.

assumption of uni-dimensionality could be correct in some research areas such as epidemiology, in which a spatial effect may depend on a unique type of contact as a causal mechanism. However, in other fields such as migration research, there could be different channels that should be considered in order to specify the spatial weight matrix.

Barthel and Neumayer (2015) justify the consideration of different types of contagions effects in their empirical analysis of asylum seekers and refugees flows by the fact that if one country introduces restrictive migration policies, flows are going to increase in other destination countries. Moreover, in addition to these negative externalities among destination countries, there could also be potential spillover effects among source countries. The existence of networks reduces the uncertainty for potential migrants and, in fact, these networks do not need to be strictly related to the migrants' country of origin but to a wider geographical area sharing the same language, religion and culture.

For this reason, they consider the possibility of spatial interdependence not only between destination countries, but also between origin countries (specific target contagion and specific source contagion in terms of Neumayer and Plümper (2010) terminology). Barthel and Neumayer (2015) also consider different spatial weight matrices for specific target contagion and for specific source contagion. In the first case, they consider two matrices obtained by the product between the change in asylum policy and the inverse of the squared distance. One matrix captures linkages among targets, where policy became relatively more restrictive while the second is related to linkages among targets where policy became relatively less restrictive. Regarding specific source contagion, they used three different spatial weight matrices: The first one is based on the inverse of the squared distance, the second on sharing a common language (spoken by at least 9 percent of the population in both countries) and, the third on common colonial experience.

Their estimation results show that spatial lag variables capturing spatial dependence among target countries have the expected negative sign, although spillover effects are quantitatively small. A 10 percent decrease in the number of applications from a source country in geographically proximate targets that have become relatively more restrictive in their asylum policies in the previous year leads to an increase of 0.23-0.26 percent in the number of asylum applications from this source in the target country under observation. In a similar way, a 10 percent increase in the number of applicants from a country of origin to geographically proximate targets that have become less restrictive in their policies in the previous year leads to a decrease of 0.61-0.69 percent in the number of asylum applications from this source in the target country considered.

By contrast, regarding spatial dependence among source countries, a 10 percent increase in asylum seekers from other geographically proximate source countries to a specific destination is estimated to increase the flow of asylum seekers from the source country considered to this same destination by 4.4 percent. Spatial source contagion based on sharing a common language is statistically significant but of a modest size, while Barthel and Neumayer (2015) find no evidence of statistically

significant spatial dependence among source countries sharing similar previous experience of having been colonised by the same target country.

13.4 Concluding Remarks

The dyadic dimension of recently available international migration data allows to analyse many new questions that could not be previously addressed. Using multidimensional panels, researchers are now improving our knowledge about the impact of traditional pull and push factors on migration flows, such as distance (Abbott and Silles, 2016) or income.

Regarding income, Llull (2016) presents evidence on the existence of heterogeneous effects of income gains in destination countries on migration depending on distance (individuals from closer countries can easily move back if they do not succeed in the new country). For example, according to his results, a 1000$ increase in US per capita income would increase the stock of Mexican immigrants in the US by a percentage that is 2.6 times larger than the percentage increase in the stock of Chinese immigrants. This differs from the standard gravity equation predicting positive linear effects of income gains on immigrant stocks. As Llull (2016) highlights, this result is clearly relevant for immigration policy design as after a positive income shock in a particular country, the migration response from the rest of the world would be very different and it could imply a significant change in the nationality composition of the immigrant population in that country. In a similar vein, Beine et al. (2016) question the traditional view that low-income in origin countries always acts as a push factor. In particular, they explain that low average income in a country may work as a push-factor, but at the same time it can act as a credit constraint, so that fewer people can afford to migrate. Only after a certain level of financial resources is reached, can people living in low-income countries and willing to migrate effectively do so. Bergh et al. (2015) find evidence supporting this hypothesis. The same authors analyse the asymmetric role of institutional quality depending on the income level of the country considered, a dimension that was not considered in the literature previously due to data limitations. According to Bergh et al. (2015), individuals may accept to stay in relatively poor countries if institutions are good, as this is one of the basic enhancers of economic growth. Accordingly, people could also be willing to migrate to a country with better institutions even if income differentials are not very high. Their empirical analysis suggests that the quality of institutions matters more as a push factor for migration in non-OECD than in OECD countries.

However, the main contribution of this new literature is related to the consideration of new factors affecting migration flows such as networks (Beine, 2016), linguistic proximity (Adserà and Pytliková, 2015), cultural barriers (Belot and Ederveen, 2012), climatic factors (Beine and Parsons, 2015), and different policy measures such as visa restrictions (Czaika and Hobolth, 2016) that could not be properly analysed in the context of time-invariant dyadic models. Regarding networks, as

noted by Beine et al. (2016), new empirical analyses have shed light on the importance of its effect: around one-third of the explained variability of migration flows is associated with networks. Beine (2016) finds that, on average, a 10 percent increase in the existing stock of migrants from a given origin country tends to increase the bilateral flow from that country by 5 percent. However, the results obtained also show that there are heterogenous effects depending on the types of destinations and over time. In particular, the network effect in receiving countries with higher levels of income is less important. The network effect has increased over time, a result that could be related to the fact that family reunification policies have recently become less restrictive. Some of the findings of the other studies considering the role of language or climatic factors have already been summarised in previous sections of the chapter.

In sum, this chapter has reviewed some recent contributions using multi-dimensional panels in the context of migration analysis and has identified some methodological challenges such as the consideration of fixed effects and the treatment of multilateral resistance that are not exclusive to this field but common to other analyses that have already been summarised in other chapters of the book. Spatial models and gravity trade equations that are described in Chaps. 9 and 11, respectively, are clearer examples. For instance, dynamic models have been extensively used in the context of trade analysis, but their use in migration modelling is still very scarce, although the role of expectations in migration decisions has been highlighted both on a theoretical and empirical basis (Beine et al., 2016; Bertoli et al., 2016). Regarding spatial dependence, as highlighted by Patuelli and Arbia (2016), "the last 10 years have shown a resurgence of the gravity/spatial interaction modelling literature, and the emergence of innovative estimation approaches making use of spatial statistics and econometrics". Last, the triple dimension, origin, destination and time, or even the quadruple dimension when gender or skill groups are analysed, have started to be considered by a few authors (see, for instance, Artuç et al., 2015 or Beine et al., 2011), but this literature will clearly expand in the future, posing new challenges for the development of new theoretical and empirical models that could be used in this context.

Naturally, migration researchers would, at the same time, benefit and contribute to advances in this field. But, apart from the previously mentioned issues, future research will certainly extend to the consideration of other topics already described in this book. For instance, panel quantile regression models described in Chap. 8 provide an adequate tool to measure potential heterogeneous effects of different migration drivers, while non-parametric models described in Chap. 7 could also be used in this context extending previous research for time-invariant dyadic data (see, for instance, Henderson and Millimet, 2008).

References

Abbott, A. and Silles, M. (2016). Determinants of international student migration. *The World Economy*, 39:621–635.

Abel, G. J. and Sander, N. (2014). Quantifying global international migration flows. *Science*, 343(6178):1520–1522.

Adserà, A. and Pytliková, M. (2015). The role of language in shaping international migration. *The Economic Journal*, 125(586):F49–F81.

Artuç, E., Docquier, F., Özden, Ç., and Parsons, C. (2015). A global assessment of human capital mobility: The role of non-OECD destinations. *World Development*, 65:6–26.

Artuç, E. and Özden, Ç. (2016). Transit migration: All roads lead to America. *World Bank Policy Research Working Paper*, 7880.

Baltagi, B. H., Egger, P., and Pfaffermayr, M. (2015). Panel data gravity models of international trade. In Baltagi, B. H., editor, *The Oxford Handbook of Panel Data*. Oxford University Press.

Barthel, F. and Neumayer, E. (2015). Spatial dependence in asylum migration. *Journal of Ethnic and Migration Studies*, 41(7):1131–1151.

Beine, M. (2016). The role of networks for migration flows: an update. *International Journal of Manpower*, 37(7):1154–1171.

Beine, M., Bertoli, S., and Fernández-Huertas Moraga, J. (2016). A practitioners' guide to gravity models of international migration. *The World Economy*, forthcoming:1–17.

Beine, M., Docquier, F., and Özden, Ç. (2011). Diasporas. *Journal of Development Economics*, 95(1):30–41.

Beine, M., Noël, R., and Ragot, L. (2014). Determinants of the international mobility of students. *Economics of Education Review*, 41:40–54.

Beine, M. and Parsons, C. (2015). Climatic factors as determinants of international migration. *Scandinavian Journal of Economics*, 117(2):723–767.

Belot, M. and Ederveen, S. (2012). Cultural barriers in migration between OECD countries. *Journal of Population Economics*, 25(3):1077–1105.

Bergh, A., Mirkina, I., and Nilsson, T. (2015). Pushed by poverty or by institutions? An analysis of migration determinants. *IFN Working Paper*, 1077.

Bertoli, S., Brücker, H., and Fernández-Huertas Moraga, J. (2016). The European crisis and migration to Germany. *Regional Science and Urban Economics*, 60:61–72.

Bertoli, S. and Fernández-Huertas Moraga, J. (2013). Multilateral resistance to migration. *Journal of Development Economics*, 102:79–100.

Burger, M., van Oort, F., and Linders, G.-J. (2009). On the specification of the gravity model of trade: Zeros, excess zeros and zero-inflated estimation. *Spatial Economic Analysis*, 4(2):167–190.

Cameron, C. A. and Trivedi, P. K. (2005). *Microeconometrics, Methods and Applications*. Cambridge University Press, Cambridge.

Chun, Y. and Griffith, D. A. (2011). Modeling network autocorrelation in spacetime migration flow data: An eigenvector spatial filtering approach. *Annals of the Association of American Geographers*, 101(3):523–536.

Clark, X., Hatton, T. J., and Williamson, J. G. (2007). Explaining U.S. immigration, 1971-1998. *The Review of Economics and Statistics*, 89(May):359–373.

Cushing, B. and Poot, J. (2004). Crossing boundaries and borders: Regional science advances in migration modelling. *Papers in Regional Science*, 83(1):317–338.

Czaika, M. and Hobolth, M. (2016). Do restrictive asylum and visa policies increase irregular migration into Europe? *European Union Politics*, 17(3):345–365.

Docquier, F. and Bhargava, A. (2007). A new panel data set on physicians' emigration rates (1991-2004). *http://perso.uclouvain.be/ frederic.docquier/ oxlight.htm*.

Dustmann, C. and Görlach, J.-S. (2016). The economics of temporary migrations. *Journal of Economic Literature*, 54(1):98–136.

Echevarria, J. and Gardeazabal, J. (2016). Refugee gravitation. *Public Choice*, 169(3-4):269–292.

Egger, P. (2000). A note on the proper econometric specification of the gravity equation. *Economics Letters*, 66(1):25–31.

Eilat, Y. and Einav, L. (2004). Determinants of international tourism: A three-dimensional panel data analysis. *Applied Economics*, 36(12):1315–1328.

Faggian, A., Corcoran, J., and Partridge, M. (2015). Interregional migration analysis. In Karlsson, C., Andersson, M., and Norman, T., editors, *Handbook of Research Methods and Applications in Economic Geography*. Edward Elgar, Cheltenham, UK.

Figueiredo, E., Lima, L. R., and Orefice, G. (2016). Migration and regional trade agreements: A new gravity estimation. *Review of International Economics*, 24(1):99–125.

Fingleton, B. and Lopez-Bazo, E. (2006). Empirical growth models with spatial effects. *Papers in Regional Science*, 85(2):177–198.

Fink, C., Miguelez, E., and Raffo, J. (2017). The determinants of the international mobility of knowledge workers. In Fink, C. and Miguelez, E., editors, *The International Mobility of Talent and Innovation: New Evidence and Policy Implications*. Cambridge University Press.

Fitzgerald, J., Leblang, D., and Teets, J. C. (2014). Defying the law of gravity: The political economy of international migration. *World Politics*, 66(03):406–445.

Griffith, D. A. (2003). *Spatial Autocorrelation and Spatial Filtering: Gaining Understanding through Theory and Scientific Visualization*. Springer, Berlin Heidelberg New York.

Grogger, J. and Hanson, G. H. (2011). Income maximization and the selection and sorting of international migrants. *Journal of Development Economics*, 95(1):42–57.

Head, K., Mayer, T., and Ries, J. (2010). The erosion of colonial trade linkages after independence. *Journal of International Economics*, 81(1):1–14.

Henderson, D. J. and Millimet, D. L. (2008). Is gravity linear? *Journal of Applied Econometrics*, 23(2):137–172.

Krugman, P. (1995). Increasing returns, imperfect competition and the positive theory of international trade. In Gopinath, G., Helpman, E., and Rogoff, K., editors, *Handbook of International Economics*, volume 3 of *Handbook of International Economics*, pages 1243–1277. North-Holland, Amsterdam.

Lesage, J. P. and Pace, R. K. (2008). Spatial econometric modeling of origin-destination flows. *Journal of Regional Science*, 48(5):941–967.

Llull, J. (2016). Understanding international migration: Evidence from a new dataset of bilateral stocks (1960–2000). *SERIEs*, 7(2):221–255.

Mayer, T. and Zignago, S. (2011). Notes on CEPII's distance measures: The GeoDist database. *CEPII Working Papers*, 25.

McFadden, D. I. (1974). Conditional logit analysis of qualitative choice analysis. In Zarembka, P., editor, *Frontiers in Econometrics*, pages 105–142. Academic Press, New York.

Morley, C., Rosselló, J., and Santana-Gallego, M. (2014). Gravity models for tourism demand: Theory and use. *Annals of Tourism Research*, 48:1–10.

Neumayer, E. and Plümper, T. (2010). Spatial effects in dyadic data. *International Organization*, 64(1):145–166.

Neumayer, E. and Plümper, T. (2016). W. *Political Science Research and Methods*, 4(1):175–193.

Ortega, F. and Peri, G. (2013). The effect of income and immigration policies on international migration. *Migration Studies*, 1(1):1–35.

Özden, C., Parsons, C. R., Schiff, M., and Walmsley, T. L. (2011). Where on earth is everybody? *World Bank Economic Review*, 25(1):12–56.

Patuelli, R. and Arbia, G. (2016). Spatial econometric interaction modelling: Where spatial econometrics and spatial interaction modelling meet. In Patuelli, R. and Arbia, G., editors, *Spatial Econometric Interaction Modelling*. Springer.

Perkins, R. and Neumayer, E. (2014). Geographies of educational mobilities: Exploring the uneven flows of international students. *Geographical Journal*, 180(3):246–259.

Pesaran, M. H. (2006). Estimation and inference in large heterogeneous panels with cross section dependence. *Econometrica*, 74(4):967–1012.

Plümper, T. and Neumayer, E. (2010). Model specification in the analysis of spatial dependence. *European Journal of Political Research*, 49(3):418–442.

Poot, J., Alimi, O., Cameron, M. P., and Maré, D. C. (2016). The gravity model of migration: The successful comeback of an ageing superstar in regional science. *Investigaciones Regionales - Journal of Regional Research*, 36:63–86.

Puhani, P. A. (2000). The Heckman correction for sample selection and its critique. *Journal of Economic Surveys*, 14(1):53–68.

Ramos, R. (2016). Gravity models: A tool for migration analysis. *IZA World of Labor*.

Santos Silva, J. M. C. and Tenreyro, S. (2006). The log of gravity. *The Review of Economics and Statistics*, 88(4):641–658.

Shepherd, B. (2016). *The Gravity Model of International Trade: A User Guide (An updated version)*. ARTNeT Gravity Modeling Initiative.

Chapter 14
Modeling Heterogeneity in Country-Industry-Year Panel Data: Two Illustrative Econometric Analyses

Jimmy Lopez and Jacques Mairesse

Abstract The macroeconomic empirical literature based on three-dimensional country-industry-year panel data, such as the widely used OECD STAN and EUKLEMS databases, has become extremely abundant, providing more observations for studies previously investigated on two-dimensional country-year panels. However, a large part of this literature does not take advantage of the development of panel data methods to deal with heterogeneity and dynamic and non-stationarity issues. We explain in this chapter how one can put them into practice and circumvent the lack of variability left in the data once one controls for simple and two-way interacted fixed effects in these industry-country-year panels. We illustrate what to do in the context of two econometric analyses, in which we first try to estimate the productivity impacts of ICT and R&D, and second the productivity impacts of product market anticompetitive regulations.

14.1 Introduction

The macroeconomic literature on economic growth and productivity has greatly benefited from country-industry panel databases, notably the widely used OECD STAN and EU KLEMS databases, allowing cross-country comparative investigations, based on easily accessible information, to study important issues that were previously only considered at a national level. Many of these studies, however, do not take advantage of the panel data methods that have been largely developed for two-dimensional (2D) panels, to deal, for example, with heterogeneity issues on non-stationary variables, and most ignore the possibilities opened with three-

Jimmy Lopez
Université de Bourgogne (LEDI) and Banque de France, Paris, France

Jacques Mairesse
CREST-ENSAE, Maastricht University (UNU-MERIT), Maastricht, the Netherlands, Banque de France and NBER, Paris, France

L. Matyas (ed.), *The Econometrics of Multi-dimensional Panels*, Advanced Studies in Theoretical and Applied Econometrics 50, DOI 10.1007/978-3-319-60783-2_14

dimensional (3D) panels, in order to overcome major difficulties due to the relatively short duration and the lack of exogenous data variability over time.[1] In this chapter, we use the EUKLEMS and OECD STAN databases and show how various heterogeneity issues can be investigated by considering the 3D structure of these data rather than treating them as individual (country×industry) and year 2D panels.

We present our simple approach on country-industry level data via two applications: (i) an estimation of the Information and Communication Technology (ICT) and Research & Development (R&D) capital effects on productivity; and (ii) an investigation of the impact of Non-Manufacturing Regulations (NMR) on productivity.[2] We believe that apart from these applications, our approach may be useful for many other macroeconomic issues. Through these applications, we mobilize 3D estimation methods, notably the 3D fixed effects and 3D Pooled Common Correlated Effect (3D-PCCE) estimators presented in this book (see Chaps. 1 and 10), paying special attention to variable non-stationarity and cointegrated relationships.

Since Paul Douglas's work beginning in the 1920s and the famous production function Douglas proposed with Charles Cobb (Cobb and Douglas, 1928), there has been an impressive amount of literature on the "laws of production". The increasing availability of panel data has shown the complexity of these "laws" resulting from heterogeneous behaviors. We begin the presentation of our approach from within this well-known framework. More specifically, we use a knowledge production function including R&D capital as well as a breakdown of physical capital between ICT and other assets. Thus, this first application fills a gap between two bodies of literature, one focusing on R&D and the other on ICT.

The second application emphasizes the impact of regulations in non-manufacturing industries, producing intermediate inputs on the productivity of industries using these inputs (called upstream and downstream industries, respectively). Regulations that protect rents in upstream industries can reduce incentives to implement efficiency improvements in downstream industries, since downstream firms will have to share the rents expected from such improvements with upstream industries. This application introduces a major advantage of investigations on country-industry panel data in tackling issues that were previously investigated using national data: the possibility of testing whether the effects of national level variables differ between industries depending on industry-specific characteristics. In this application, we test whether the impact of upstream NMR on downstream productivity grows with the more intensive use of regulated intermediate inputs. This method solves empirical issues, such as reverse causality if public authorities change the NMR depending on industry activity, but introduces difficulties in the treatment of heterogeneity, which we discuss.

[1] The EU KLEMS and OECD STAN databases provides annual unbalanced panels covering the period 19702014 for EU KLEMS and 1970 2011 for STAN. However, the lack of data on specific country-industry variables may result on shorter estimation periods. For instance, applications presented in this chapter cover the period 1987–2008, with an average length of 18 years.

[2] These applications borrow from Cette et al. (2013), but extend that analysis to take into account various heterogeneity issues.

Both applications mobilize the OECD STAN and EU KLEMS databases to measure productivity and physical production factors, whereas R&D investments are taken from the OECD ANBERD database. The second application also requires the OECD indicators of NMR. Thus, both our analyses are grounded on almost the same unbalanced panel of about four thousand observations covering 15 countries and 18 industries for the period 1987 to 2008.

Through the two applications, we show how the methodology dealing with the heterogeneity of unobserved individual effects, unobserved common factor effects, and explanatory variable coefficients takes account of our country-industry data characteristics. However, prior to this, we must discuss the dynamic issues. As we will see further, estimations of dynamic models through an Autoregressive Distributed Lag (ADL) specification on our panel suffers from: (i) autocorrelated residuals, leading to endogeneity of the lagged dependent variable; and (ii) very unlikely estimation results in many cases. Therefore, we focus on the long-term relationship and take advantage of the non-stationarity of our data through the Dynamic Ordinary Least Squares (DOLS) estimator.

The country, industry, and time effect heterogeneity may be taken into account by various possible sets of fixed effects (see Chap. 1 on fixed effects models). These different sets may lead to marked differences in estimation results, as shown by the literature and our applications. We explain our specific choices based on variance analysis, potential omission bias, and – for the second application – the requirement of our approach to prevent reverse causality. Notably, we show that considering only two dimensions (individual and time) when choosing the set of fixed effects would account for a huge part of our data variability, leading to serious difficulties in identifying the estimated coefficients, while leaving several omission bias sources.

Time fixed effects specifications assume homogeneous impacts of unobserved common factors. Under this constraint, our estimators could be biased if these common factors are correlated with explanatory variables (see Chap. 10 on cross-sectional error dependence), so we relax this assumption. In the same way, our empirical investigation must take into account the heterogeneity of explanatory variables effects, as most of our variables are non-stationary. Otherwise, the estimation residuals would be non-stationary, even if the "true" data generating process is cointegrated. Our approach is based on Pesaran's (2006) common correlated estimator and Pesaran et al.'s (1999) mean group estimator. Our panel is too short to estimate the "individual" (country$\times$industry) coefficients required by these estimators, but considering the 3D structure of our data helps take this heterogeneity into account. We show that industry and country coefficient heterogeneity can be taken into account when we consider them as two different dimensions and this heterogeneous specification appears essential to prevent spurious regression when variables are non-stationary.

Our estimation results confirm previous studies on country-industry panel data and show their sensitivity to heterogeneity issues. In the first application, the ICT and R&D elasticities are strongly robust to the various heterogeneity issues, with values ranging from 9% to 10% for ICT and from 6% to 7% for R&D. In the second application, we find that our approach is necessary to detect the impact of NMR on

productivity. More specifically, according to our estimation results, the impact of upstream NMR on downstream productivity increases with the intensity of the use of regulated intermediate inputs.

This chapter is organized in two parts around the two applications (Sects. 14.2 and 14.3), both starting with a literature review (Sects. 14.2.1 and 14.3.1) and a presentation of the model and data (Sects. 14.2.2 and 14.3.2).[3] The first application continues with the dynamic specification used (Sect. 14.2.3), before setting out how we consider the 3D structure of our data in order to take account of heterogeneity through fixed effects (Sect. 14.2.4) or common correlated and mean group estimators (Sect. 14.2.5). We conclude the first application by putting in perspective our approach and usual panel cointegration tests (Sect. 14.2.6). Next, the second application presents our estimation strategy when using national level NMR indicators (Sect. 14.3.3) and the corresponding estimation results (Sect. 14.3.4), as well as some limitations regarding heterogeneity issues (Sect. 14.3.5). A few reflections on open issues and directions for future research conclude the chapter (Sect. 14.4). More detailed information on data sources and calculations are provided in the Appendix, along with supplementary estimation tables.

14.2 ICT, R&D, and Productivity

14.2.1 Literature Review

We begin by investigating the impact of ICT and R&D on productivity. Both ICT and R&D are driving forces behind modern growth that have undergone considerable change in recent decades. The extraordinary development of ICTs is a general purpose technological shock: most economic activities have benefited from the steady and sizeable fall in prices. Therefore, the economic literature devotes a great deal of attention to ICT's contribution to growth, mainly through accounting analysis (e.g., Jorgenson et al., 2016; Timmer et al., 2011) but also econometric analysis. Stiroh's (2005) meta-analysis shows an average ICT capital elasticity of value-added of 0.05, with higher values for industry level data, as ICT elasticities estimated on industry level data take within-industry externalities into account. At the same time as ICT has developed, R&D spending has also surged. The econometric analysis of R&D capital elasticity provides very different results, from 0.01 to 0.25, with 0.08 on average, but generally concludes that there have been excess returns: estimated R&D elasticity induces a higher net return (marginal productivity less depreciation rate) than for physical capital (see Hall et al., 2010 for a survey).

ICT and R&D investment are closely linked, but few empirical analyses investigate their simultaneous impact on economic growth. Kocoglu and Mairesse (2005) provide a growth accounting analysis for France and the USA underlining the im-

[3] The literature reviews may be skipped by readers interested in only the methodology without impairing their understanding of the other sections.

portance of R&D externalities.[4] According to Venturini's (2015) main estimation results on national panel data for 15 OECD countries, ICT and R&D elasticities are 0.05 and 0.13, respectively, confirming previous results for ICT and strong R&D externalities. Cette et al. (2013) mobilize country-industry panel data for 15 OECD countries and also estimate an ICT elasticity of 0.05, but a smaller R&D elasticity of 0.08, which is closer to the empirical literature specific to R&D. This difference with Venturini's (2015) estimated elasticities may be explained by inter-industry externalities.

14.2.2 Model and Data

In order to investigate the impact of ICT and R&D on productivity, we assume a Cobb-Douglas production function including R&D capital as well as a breakdown of physical capital between ICT and other assets (country, industry, and year indices omitted to alleviate the equations):

$$Y = A \cdot L^{\lambda} \prod_{a} C_a^{\gamma_a} \,, \tag{14.1}$$

where Y is the value added, A disembodied technical change, L total employment, C_a the capital stock of asset a, λ and γ_a the output elasticities. We distinguish four capital assets: ICT equipment (C_I), non-ICT equipment (C_{NI}), non-residential construction (C_c), and R&D capital (C_k).

In order to estimate ICT and R&D elasticities parsimoniously, we define and calibrate the following partial MultiFactor Productivity (MFPp) taking account of the non-ICT physical capital contributions only (with small letters for logarithms):

$$mfp^p \equiv (y-l) - \tilde{\gamma}_{NI}(c_{NI}-l) - \tilde{\gamma}_c(c_c-l) \,. \tag{14.2}$$

We calibrate the elasticities $\tilde{\gamma}_{NI}$ and $\tilde{\gamma}_c$ at the industry level by the average shares of their user costs computed for the USA over the whole estimation period (our estimation results are robust to this choice of calibration). These shares are proxies for private returns. The C_a/L ratios are termed the capital intensities.

By introducing this partial MFP definition into our production function equation and by assuming constant returns to scale, our main estimated specification becomes

$$mfp^p_{ijt} = \alpha + \gamma_t(c_I-l)_{ijt} + \gamma_k(c_k-l)_{ijt} + \varepsilon_{ijt} \,, \tag{14.3}$$

where i, j, and t are country, industry, and time indices; ε is the error term.

This simple equation underlines our focus on long-term average elasticities. However, the various estimations presented broaden this equation further by in-

[4] According to Kocoglu and Mairesse's (2005) accounting analysis, the R&D contribution to annual economic growth for the 1980–2000 period was 0.08% in France and 0.10% in the USA without externalities but 0.24% and 0.56% with externalities, respectively.

cluding various sets of fixed effects, dynamics of the explanatory variables, and by allowing for coefficient heterogeneity.

Our analysis of relation (14.3) is based on an unbalanced panel of 3,999 observations (4,750 for the static model) from 15 OECD countries and 18 industries, covering the period from 1987 to 2008.[5] We mobilize data from three databases: (i) STAN for value added and total employment in numbers of persons engaged; (ii) EU KLEMS for physical capital; and (iii) ANBERD for R&D investment (see Appendix 1 for a more detailed presentation of the data sources and authors' calculations).

In order to test for the non-stationarity of equation (14.3) variables, we perform an Im et al. (2003) test. This test has the desirable characteristic of allowing for heterogeneous roots of the variables, but assumes cross-sectional independence. To mitigate this issue, we subtract the cross-sectional averages from the variable prior to implementing the test. The Im et al. (2003) test is particularly sensitive to the specification of the dynamic. As our panel is relatively short (18 periods on average), we only include up to three lags in the specifications; so we may underestimate the relevant number of lags and greatly reduce the test's power. According to this test, partial MFP and ICT capital intensity are not stationary, but the results for R&D capital intensity are more surprising: when – and only when – this variable is demeaned and a trend is introduced into the test, does the test point to stationarity. However, as a precaution we treat the three variables as non-stationary in what follows.

14.2.3 Discussion of the Dynamic Specification

As for stationary variables, the serial correlation of the estimated cointegration residuals leads to an inefficient OLS estimator and biased variance estimators, but also to a biased OLS estimator when using an Autoregressive Distributed Lag (ADL) model as lagged dependent variables are then endogenous.[6] Unfortunately, when using an ADL model for relation (14.3), residual serial correlation remains, even with the introduction of up to four lagged dependent and explanatory variables in the estimated specification, according to Wooldridge's (2002) test for serial correlation in the idiosyncratic errors. Moreover, the ADL model estimation results from relation (14.3) were very unlikely (with negative elasticities).[7]

---

[5] Some industries make almost no investment in R&D, in particular: wood products, energy, construction, retail distribution, hotels & restaurants, and banking services. Therefore, we estimate the R&D output elasticity on just the 12 industries investing markedly in R&D. Our estimation results are robust to this choice of industry exclusion.

[6] We assume in this section and the next that relation (14.3) variables are cointegrated, which is tested and discussed after presenting our approach to dealing with heterogeneity.

[7] See Chap. 10 for more information on estimation of ADL models on 2D and 3D panel data. The assumption of no serial correlation of residuals is a base for estimating these models.

We therefore preferred to use the Dynamic OLS (DOLS) estimator proposed by Stock and Watson (1993). This estimator benefits from the non-stationarity of variables to make sure that the estimated elasticities are not biased by short-term correlations between the variables and the idiosyncratic error, and that we can consider them as long-term parameters. The DOLS estimator eliminates these short-term correlations by including leads and lags of the first differences of the non-stationary explanatory variables in the regressions. To choose the number of leads and lags, we compare OLS estimates and DOLS estimates with up to three lags and leads. We use a Hausman test to conclude that we must use the DOLS estimator rather than the OLS estimator and that one lead and one lag are enough, but estimation results are very similar whatever the specification (see Table 14.6 in Appendix 2).

According to Wooldridge's (2002) test, the DOLS specification does not remove the serial correlation of the residual. As with stationary variables, the standard practice is to compute heteroscedasticity and serial correlation-consistent standard errors. However, on a sample with a small time dimension, the corrected standard errors may lead to larger errors in test rejection than the uncorrected standard errors, so it is good practice to check whether both lead to the same conclusions. We therefore observe the following rule in the next section: we present standard errors corrected for residual heteroscedasticity and serial correlation, using the Newey and West (1987) method in the main text, and comment on the robustness of our inference to the use of uncorrected standard errors.

14.2.4 Three-dimensional Structure and Fixed Effects

Table 14.1 summarizes the variance analysis of our three main variables. Column 1 presents the coefficient of variation of each variable. Column 2 documents the R-squares of their regression on the three one-way fixed effects separately, as a basic control for the usual sources of specification errors, such as omitted (time invariant) country and industry characteristics. The next columns document the additional variability lost when we also include two-way interaction effects in order to control for other potential sources of specification errors to be discussed in the estimation section. The R-squares presented in this second step correspond to the regressions on the set of two-way fixed effects of the first step residual, i.e., the residual of the regression of our variables on the one-way fixed effects. In columns 3 to 5, the two-way fixed effects are introduced separately. Column 6 presents the simultaneous introduction of country×year and industry×year fixed effects (we will argue further that this specification is particularly relevant to our data), and column 7 shows what happens if we introduce all the possible two-way fixed effects simultaneously.

We see that the one-way country, industry, and year fixed effects taken alone already account for large (at least 85%) shares of the variability of the three variables of our model (column 2). Among the two-way fixed effects, the country×year effects are the least expensive in terms of variability, explaining at most 9% of the first-step residual variance (column 3). Both industry×year and country×industry

Table 14.1 Variance analysis – application 1

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Variable (in logarithm)	Coef. of variation	First step R^2	Second step R^2				
Partial MFP	0.142	0.975	0.052	0.542	0.373	0.583	0.913
ICT capital intensity	0.230	0.864	0.093	0.045	0.740	0.137	0.859
R&D capital intensity[a]	0.282	0.853	0.030	0.073	0.796	0.102	0.881
Observations	4,750	4,750	4,750	4,750	4,750	4,750	4,750
Residual degree of freedom		4,697	4,391	4,303	4,445	4,018	3,797
Fixed effects:							
Country, industry, year		Y	-	-	-	-	-
Country×year		N	Y	N	N	Y	Y
Industry×year		N	N	Y	N	Y	Y
Country×industry		N	N	N	Y	N	Y

Notes: First step R^2: R-squares of the regressions of our model variables on the three one-way effects separately. Second step R^2: R-squares of the regressions of the first step residuals on two-way fixed effects.

[a] The analysis sample for R&D capital intensity is reduced to 12 industries as six industries make almost no investment in R&D, thus reducing the sample size (to 3,006 observations) and degree of freedom.

fixed effects are particularly costly, the former for the dependent variable (54% of the first-step residual variance, column 4) and the latter for the two explanatory variables (between 74% and 80%, column 5). Introducing all the two-way fixed effects leads us to take into account around 90% of the first-step residual variance (column 7), which is already a very small part of the variable variances. In other words, this full set of fixed effects takes into account almost all the data variability. The omission of the country×industry fixed effects allows us to keep some data variability, still taking into account 58% of the first-step residual for the partial MFP but only between 10% and 14% for the explanatory variables (column 6).

We must choose among the two-way fixed effects as the introduction of the full set of possible fixed effects in the estimated equation would take into account almost all the variability of the dependent and explanatory variables.[8] We assume that country×year and industry×year fixed effects, either alone or taken together, can act as good proxies for a variety of omitted variables. In particular, they can take into account differences between countries and/or industries in technical progress, in the development of labor force education and skills, in changes in international trade conditions, and so on. Moreover, variance analysis in Table 14.1 shows that the country×year fixed effects are the least costly of the two-way fixed effects in terms of variance. We do not expect introducing country×industry fixed effects to prevent specific sources of omission bias, therefore we favored specifications with country×year and/or industry×year fixed effects.

[8] A mixed fixed and random-effects model would allow taking into account parsimoniously all the specific effects, thus increasing the fitting and predictive power of our model (see Chap. 2).

Reflecting the variance analysis, Table 14.2 presents the sensitivity of the estimation results to the various sets of fixed effects. Column 5 presents the estimation results for our preferred specification including country×year and industry×year fixed effects. The ICT and R&D estimated elasticities are strongly statistically significant and more so than their average shares in total costs, suggesting large social excess returns: these elasticities are of 10% and 6%, whereas private returns, approximated by average shares in total costs, are of 3% and 4%.[9] These results are strongly robust to different sets of fixed effects for all but two: (i) if we include the full set of fixed effects (column 6), the ICT elasticity is weakly statistically significant and negative, confirming the lack of remaining variability to identify the coefficients and that we must choose between the two-way fixed effects; and (ii) if we include only the individual (country×industry) two-way fixed effects (column 4), the ICT estimated elasticity falls to 7% (but the difference with our preferred estimate is not statistically significant) and the R&D elasticity rises to 15%, suggesting omission bias and illustrating the importance of the choice of fixed effects based on economic considerations and variance analysis results.

Table 14.2 ICT and R&D estimated elasticity sensitivity to fixed effects

	(1)	(2)	(3)	(4)	(5)	(6)
ICT capital intensity	0.0934***	0.0999***	0.0925***	0.0683***	0.101***	−0.0240*
	[0.0126]	[0.0132]	[0.00996]	[0.0228]	[0.0105]	[0.0129]
R&D capital intensity[a]	0.0740***	0.0691***	0.0645***	0.146***	0.0612***	0.0637***
	[0.00994]	[0.00975]	[0.00728]	[0.0230]	[0.00727]	[0.0142]
Observations	3,991	3,991	3,991	3,991	3,991	3,991
R-squared	0.980	0.981	0.991	0.988	0.991	0.998
RMSE	0.258	0.258	0.185	0.203	0.182	0.0821
Residual degree of freedom	3,933	3,704	3,628	3,714	3,399	3,180
Fixed effects:						
Country, industry, year	Y	Y	Y	Y	Y	Y
Country×year	N	Y	N	N	Y	Y
Industry×year	N	N	Y	N	Y	Y
Country×industry	N	N	N	Y	N	Y

Notes: Estimator: DOLS(1;1), dependent variable: Partial MFP. Newey-West standard errors in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

[a] The R&D output elasticity is estimated on 12 industries as six industries make almost no investment in R&D.

We also test the sensitivity of the results to (i) uncorrected standard errors; (ii) omission of ICT or R&D capital stock; and (iii) estimating rather than calibrating the "structure" and "non-ICT equipment" capital stocks elasticities. ICT and R&D estimated elasticities are robust to these three changes, except that the ICT estimated elasticity is half as much and weakly statistically significant when "structure" and

[9] ICT and R&D elasticities are estimated on different samples, as six industries make almost no investment in R&D, so their average shares in total costs are also calculated on different samples. On the whole estimation sample, R&D estimated elasticity and average share over total cost would be 4.1% and 2.7%, respectively.

"non-ICT equipment" capital stocks elasticities and country×industry fixed effects are introduced (see Tables 14.7, 14.8, and 14.9 in Appendix 2). Moreover, the "structure" and "non-ICT equipment" estimated elasticities are statistically significant and close to their calibration values, except again when country×industry fixed effects are included in the estimated specification. These results may be explained by the lack of remaining variability when country×industry fixed effects are included and argue against their introduction.

14.2.5 Heterogeneity of Factor Effects

The effects of explanatory variables are generally heterogeneous. In our study, R&D and ICT elasticity must differ between countries and industries, which explains why some industries make almost no investment in R&D and/or little in ICT, for instance. However, most econometric methods mobilized in the applied economic literature assume homogeneous effects for observable factors as well as for unobservable common factors. For observable factors, the homogeneity constraint underlines the focus on average effects as OLS regression produces variance-weighted averages of the individual effects. Nevertheless, this constraint may be particularly harmful for our empirical investigation inferences because we mobilize non-stationary variables; it induces non-stationary residuals even if the true data generating process were cointegrated.[10] For unobservable common factors, sources of cross sectional error dependence, the homogeneity constraint on their effects leads to biased OLS estimators if these common factors are correlated to the (observed) explanatory variables, whether or not they are stationary. If some of the common factors are non-stationary, the homogeneity constraint also leads to non-stationary residuals.

It is standard practice to estimate separate individual regressions and calculate the coefficient means. Pesaran et al. (1999) call this method the mean group estimator and derive its asymptotic distribution for 2D panels. But this method requires panel data with a large time dimension. When estimating separate individual regressions on our panel with an average length of 18 years, the confidence intervals of estimated coefficients are huge and their averages yield unlikely results. This may explain why the empirical literature using country-industry panel data barely explores the coefficient heterogeneity issue. We argue that considering the 3D structure of our data provides a feasible way to take coefficient heterogeneity into account.

We are unable to take into account country and industry heterogeneity together through individual (country×industry) coefficients, but we can do so separately; we assume that the individual coefficients are composed by country and industry specific components (i.e., we assume $\gamma_{(I,ij)} = \gamma_{(I,i)} + \gamma_{(I,j)}$ and $\gamma_{(k,ij)} = \gamma_{(k,i)} + \gamma_{(k,j)}$ for ICT and R&D elasticities). Indeed, when estimating the industry (or country)

[10] This is easy to see with a simple model $y_{it} = \alpha_i + \beta_i x_{it} + u_{it}$, with i and t individual and time indices, x a non-stationary variable and u_{it} a stationary residual. Under the coefficient homogeneity constraint, the estimated equation becomes $y_{it} = \alpha_i + \beta x_{it} + v_{it}$ with the residual term $v_{it} = (\beta_i - \beta)x_{it} + u_{it}$ non-stationary.

specific coefficient, we mobilize not only a time series, but a country (or industry) panel, leading to more precise estimation results. This sensitivity analysis increases the confidence in our estimation results when they are robust to it, and otherwise avoids conclusions based on spurious regressions.[11]

In the same way, we should not assume homogeneous effects of unobservable common factors, which is done when the common factors are taken into account by time fixed effects. If these common factors are correlated to the explanatory variables, the homogeneity assumption would lead to the endogeneity of the explanatory variables. This correlation must appear in our application; for instance the quality of the institutions may be correlated to investments in ICT and R&D. A facilitating factor to take account of this heterogeneity is that we are not interested in the unobservable common factors for themselves, but rather in eliminating their influence on the estimation of the explanatory variable coefficients. For 2D panel data, Pesaran (2006) proposes filtering the explanatory variables by introducing the cross-sectional averages of dependent and explanatory variables into the estimated specifications and calls this method the common correlated estimator. Chap. 10 develops this estimator for 3D panel data. As with the fixed effects, allowing for individual (country×industry) heterogeneity of the coefficient of variables' cross-sectional averages would explain at least 99.8% of our data variability, preventing the identification of the parameters. Therefore, as with the explanatory variable but for another issue, we take into account the country and industry heterogeneity of the unobservable common factors separately.

Table 14.3 presents the sensitivity of our estimation results to allowing for the heterogeneity of factors' effects. All estimations assume country×year and industry×year unobserved common factors. Column 1 presents previous estimates under the homogeneity constraint for observable and unobservable factors (as in Table 14.2 column 5). Columns 2 and 4 present average estimation results when ICT and R&D elasticities are allowed to differ between countries and industries. Columns 3 and 4 present estimation results when the effects of unobserved common factors are allowed to differ. ICT and R&D elasticities are always positive and strongly statistically significant, pushing back the spurious regression issue. The estimated elasticity values are robust to taking into account the heterogeneity for the observed explanatory variables or unobserved common factors separately (columns 2 and 3), but the average estimated ICT elasticity declines markedly from 10% to 4% when the heterogeneity of the observed and unobserved factors are considered together (column 4). We have neither economic nor empirical explanations for this ICT elasticity reduction, except the lack of remaining data variability. We guess that this last estimation specification is over the limit of what is feasible with our country-industry panel data, whereas column 2 and 4 are at the limit.

[11] The Random Coefficient linear Model (RCM) for 3D panels developed in Chap. 5 allows taking into account parsimoniously coefficient individual heterogeneity, but this estimator is consistent only under the independence assumption of coefficients with respect to the explanatory variables. However, this approach could be mixed with our fixed coefficients in order to improve the fitting and predictive power of our model.

Table 14.3 Estimation of ICT and R&D elasticity allowing for heterogeneous effects

	(1)	(2)	(3)	(4)
ICT capital intensity	0.101***	0.0945***	0.0925***	0.0444***
	[0.00717]	[0.00855]	[0.00734]	[0.0132]
R&D capital intensity[a]	0.0612***	0.0655***	0.0752***	0.0816***
	[0.00542]	[0.00716]	[0.00556]	[0.0125]
Observations	3,991	3,991	3,991	3,991
R-squared	0.991	0.994	0.994	0.996
RMSE	0.182	0.151	0.138	0.123
Residual degree of freedom	3,399	3,175	3,796	3,572
Heterogeneous effects of:				
Observable variables	N	Y	N	Y
Unobserved common factors	N	N	Y	Y

Notes: Estimator: DOLS(1;1), dependent variable: Partial MFP. Standard errors in brackets - *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Country×year and industry×year unobserved common factors are taken into account by fixed effects (columns 1 and 2) or cross-sectional averages of the dependent and explanatory variables (columns 3 and 4). Columns 2 and 4 present unweighted averages of industry and country specific coefficients.

[a] The R&D output elasticity is estimated on 12 industries as six industries make almost no investment in R&D.

14.2.6 Discussion of Cointegration

Previous sections assume that relation (14.3) variables are cointegrated. To test this assumption, Pedroni's (1999; 2001) panel cointegration tests are often performed (see, for instance, Hurlin and Mignon, 2007 for a review of the various first and second generation panel cointegration tests). The first step of Pedroni's cointegration tests is to estimate the long-term relationship, allowing for the individual heterogeneity of coefficients. Based on the first step residuals, seven cointegration tests are calculated, taking account of different nuisance parameters. Unfortunately, on samples with small time dimensions, we often get contradictory results. In our application, four tests conclude that cointegration occurs (with p-values of the null hypothesis of no-cointegration less than 1%), two that no cointegration occurs (with p-values over 10%), and one is more ambiguous (with a p-value of 9%).

To go beyond this ambiguity, we may select our "favored" test. Under the alternative assumption of cointegration, four of Pedroni's tests, called panel-tests, assume homogeneous roots for the first-step estimation residual, and three tests relax this assumption, the group-tests. As discussed in the previous section, we have no reason to assume homogeneity and prefer the group-tests. Among these group-tests, Pedroni's (2004) Monte-Carlo simulations show that, for panels with short time dimensions, the two group-t tests exhibit a strong size, and the group-ρ test a weak size. Therefore, the more cautious rule would be to favor the group-ρ test. This test concludes that a cointegrated relationship occurs.

However, the previous section underlines that Pedroni's cointegration test results are not enough to reach conclusions about the relevance of inferences made using different (not individual specific) coefficient heterogeneity assumptions from these tests. The residuals of the estimated relationship would be non-stationary even if the true data generating process were cointegrated. Therefore, we would like to test the stationarity of our main estimation residuals in addition to Pedroni's cointegration tests. This test could be difficult to implement, as there are many different heterogeneity assumptions leading to various test statistic distributions to compute. But it may be good practice to perform a unit root test using our main estimation residuals. If Pedroni's tests conclude that cointegration occurs but the unit root assumption is not rejected for our main estimation residuals, the cointegration conclusion should be taken with caution. In our application, we perform the Im et al. (2003) unit root test on residuals from the four Table 14.3 specifications. These tests conclude the residuals are non-stationary, except when heterogeneous effects are allowed for both explanatory variables and unobserved common factors (column 4). We see that preventing spurious regression may require taking our data to the limit (or over) of what is feasible.

14.3 Productivity Impact of Non-Manufacturing Regulations

14.3.1 Literature Review

After estimating the impact of ICT and R&D capital on productivity, we extend our analysis to the effects of Non-Manufacturing Regulations (NMR). The literature investigating the impact of competition – and policies affecting it – (see Aghion et al., 2009 for a survey) has evolved during the last 10 years, notably to emphasize the impact of regulations in non-manufacturing industries, producing intermediate inputs on the productivity of industries using these inputs (referred to as upstream and downstream industries, respectively). If firms in downstream industries have to negotiate terms and conditions of their contracts with suppliers having market power thanks to regulations, part of the rents expected downstream from adopting best practices will be grabbed by intermediate input providers. This in turn will reduce incentives to improve efficiency and curb productivity in downstream sectors.

Only a few papers have looked at this issue so far. Some of them are static cross-sectional analyses relating manufacturing productivity outcomes to measures of competition in services (see Allegra et al., 2004, Faini et al., 2006, and Barone and Cingano, 2011). Others use firm level data for one country: Forlani (2010) on France and Arnold et al. (2011) on the Czech Republic. The estimation results of these papers are not directly comparable, but conclude that Non-Manufacturing Regulations (NMR) negatively impact innovation and productivity.

This latter application borrows from our previous papers on country-industry panel data, but discusses and extends the analysis of heterogeneity issues. The for-

mer investigations confirm a negative impact of NMR on productivity and that it undergoes a decline in competition (Cette et al., 2016). Moreover, the negative impact of NMR is stronger close to the technological frontier (Bourlès et al., 2013) and stems largely (30% to 60%) from the effects of NMR on R&D and ICT investments (Cette et al., 2013). In light of these findings, the impact of NMR on productivity estimated in this chapter should be interpreted as the impact on an industry at the average distance from the technological frontier and a lower limit, because the impact of NMR via ICT and R&D investment is ignored.

14.3.2 Model and Data

In this second application, we investigate the effects of regulations in non-manufacturing industries producing intermediate inputs (the upstream industries) on the productivity of industries, using these inputs (the downstream industries). Thanks to this second application, we introduce and discuss an important advantage of country-industry panel data to tackle issues that were previously investigated on national data; the possibility of testing whether the effects of national level variables differ between industries depending on industry-specific characteristics, thus identifying the causality link from NMR to productivity.

In order to focus on the NMR effects, we estimate the impact of regulations on MultiFactor Productivity (MFP), assuming the same Cobb-Douglas knowledge production function as previously and constant returns to scale (country, industry, and year indices omitted to alleviate equations):

$$mfp \equiv (y-l) - \tilde{\gamma}_{NI}(c_{NI}-l) - \tilde{\gamma}_c(c_c-l) - \tilde{\gamma}_I(c_I-l) - \tilde{\gamma}_k(c_k-l)\,, \qquad (14.4)$$

with lower-case letters for logarithms. Labor productivity Y/L, asset a capital intensities C_a/L, and calibrated elasticities $\tilde{\gamma}_a$ are computed with the same data and method as in the first application.[12] This MFP measure takes into account all the capital asset contributions, contrary to the partial MFP defined in the previous application. Our main estimation results are robust when factor elasticities are estimated rather than calibrated.

The anti-competition regulations are taken into account through the OECD NMR indicators. They try to measure the extent to which competition and firm choices are restricted when there are no a priori reasons for government interference, or when regulatory goals could plausibly be achieved by less coercive means. They are based on detailed information on laws, rules, and market and industry settings and they cover energy (gas and electricity), transport (rail, road, and air) and communications (post, fixed and cellular telecommunications), retail distribution, and professional services (see Appendix 1 for a more detailed presentation).

[12] We assume $\tilde{\gamma}_k$ to be zero for the six industries making almost no investment in R&D (wood products, energy, construction, retail distribution, hotels & restaurants, and banking services).

We investigate the impact of upstream NMR on downstream productivity. In a macro-econometric analysis such as ours, lack of data variability means that NMR indicators cannot be used separately to assess this upstream regulatory impact. Therefore, we combine these NMR indicators, considering that their individual impacts are most likely to vary with the respective importance of upstream industries as suppliers of intermediate inputs and computing the following indicator of regulatory impact (*REG*)

$$REG_{ijt} = \sum_{k \neq j} w_j^k \cdot NMR_{it}^k \,, \tag{14.5}$$

where NMR_{it}^k is the NMR indicator of the upstream industry k for country i in year t, and w_j^k stands for the intensity-of-use of intermediate inputs from industry k by industry j, as measured from the US 2000 input-output table as the ratio of the intermediate inputs from industry k to industry j over the total output of industry j.

We investigate the effects of the *REG* indicator on MFP using the following estimated specification[13]

$$mfp_{ijt} = \alpha + \beta \cdot mfp_{jt}^{US} + \theta \cdot REG_{ijt} + \varepsilon_{ijt} \,. \tag{14.6}$$

As in the previous application, this simple equation underlines our focus on the long-term average, but the various estimations broaden this equation by including different sets of fixed effects, taking into account the explanatory variables dynamic and allowing for coefficient heterogeneity. The US MFP, measuring the technological frontier of our sample, can be used for estimating the catch-up effect, but is also particularly relevant for improving the estimation of the NMR impact on productivity, as discussed below.

14.3.3 Estimation Strategy with 2D Explanatory Variables

A major issue when estimating NMR effects on productivity is the endogeneity of regulation because of reverse causality. Public authorities may react to country specific productivity shocks by implementing NMR changes. The 3D structure of our dependent variable helps to address this issue, contrary to empirical investigations on 2D national level data. This structure and the construction of the *REG* indicator, which means the impact of NMR may differ between industries, allows us to include country×year fixed effects in our estimated specifications that offset this reverse causality endogeneity bias. These fixed effects are essential to our estimation strategy, and we focus hereafter on specifications including at least these fixed effects.[14] Therefore, our specifications allow testing whether the impact of upstream

[13] See Appendix 1 for information on this database and more detailed information on data sources and the authors' calculations.

[14] As for the previous application, the country×year fixed effects also act as good proxies for a variety of omitted variables (technical progress, labour-force education and skills, international trade conditions, etc.).

NMR on downstream productivity grows with the intensity of the use of regulated intermediate inputs.

Table 14.4 summarizes the variance analysis of relation (14.6) variables, showing the consequences of the choice of fixed effects. As with Table 14.1, the first column presents the coefficient of variation of each variable, the second column documents the R-squares of the regressions of the variables on the three one-way fixed effects separately, and the following columns document the additional variability lost when we also include two-way fixed effects.

Table 14.4 Variance analysis – application 2

	(1)	(2)	(3)	(4)	(5)	(6)
Variable	Coef. of variation	First step R^2	Second step R^2			
MFP (in logarithm)	0.183	0.984	0.046	0.557	0.468	0.910
US MFP (in logarithm)	0.177	0.990	0.019	[a]	0.091	[a]
Regulatory impact (REG)	0.427	0.925	0.358	0.560	0.767	0.949
Observations	4,624	4,624	4,624	4,624	4,624	4,624
Residual degree of freedom		4,571	4,316	3,956	4,081	3,738
Fixed effects:						
Country, industry, year		Y	-	-	-	-
Country×year		N	Y	Y	Y	Y
Industry×year		N	N	Y	N	Y
Country×industry		N	N	N	Y	Y

Notes: First step R^2: R-squares of the regressions of our model variables on the three one-way effects separately. Second step R^2: R-squares of the regressions of the first step residuals on two-way fixed effects.

[a] The US MFP is industry×year specific and so collinear to the industry×year fixed effects.

We see that the one-way country, industry, and year fixed effects taken alone already account for a large part of data variability, even more so than in the previous application. Fortunately, the addition of country×year fixed effects is relatively inexpensive in terms of variability (column 3). This is not the case when adding industry×year and/or country×industry fixed effects: almost all data variability is taken into account (columns 4 to 6). Therefore, we must avoid introducing the full set of fixed effects, as with the first application, but even the sets including industry×year (or country×industry) fixed effects must be used with caution.

Instead of industry×year fixed effects, the US MFP variable provides a parsimonious way to take into account a variety of omitted variables (such as industry technical progress).[15] It justifies our inclusion of US MFP in the estimated specifications at least insofar as to estimate a catch-up effect. It is worth noting that the omission of the US MFP variable would induce the non-stationarity of the esti-

[15] Inserting the average industry×year MFP rather than US MFP would play the same role.

mation residual (when industry×year fixed effects are also omitted), leading to the fallacious regression issue.[16]

14.3.4 Estimation Results

As with the previous production function estimation, we mobilize the DOLS estimator to estimate the NMR and catch-up effects on productivity.[17] Reflecting the variance analysis, Table 14.5 presents the sensitivity of the estimation results to various sets of fixed effects (columns 1 to 5). Based on the discussion in the previous section, our privileged estimation is with country×year and industry fixed effects (column 2). This table also shows the effects of an alternative indicator to *REG*, which is an unweighted average of the NMR OECD indicators (column 6), in order to emphasize the importance of our estimation strategy.

Table 14.5 *REG* effects estimation sensitivity to fixed effects

	(1)	(2)	(3)	(4)	(5)	(6)
US MFP	0.907^{***}	0.910^{***}	[a]	0.853^{***}	[a]	0.906^{***}
	[0.0315]	[0.0310]		[0.0183]		[0.0315]
Regulatory impact (REG)	-0.142^{**}	-0.285^{***}	-0.141^{*}	-0.384^{***}	0.0116	
	[0.0597]	[0.0727]	[0.0823]	[0.0824]	[0.118]	
Alternative regulatory impact						0.297
						[0.404]
Observations	4,099	4,099	4,099	4,099	4,099	4,099
R-squared	0.992	0.992	0.993	0.998	0.999	0.991
RMSE	0.193	0.193	0.190	0.103	0.084	0.194
Residual degree of freedom	4,044	3,807	3,519	3,519	3,301	4,044
Fixed effects:						
Country, industry, year	Y	Y	Y	Y	Y	Y
Country×year	N	Y	Y	Y	Y	N
Industry×year	N	N	Y	N	Y	N
Country×industry	N	N	N	Y	Y	N

Notes: Estimator: DOLS(1;1), dependent variable: MFP. Newey-West standard errors in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.
[a] US MFP is by definition collinear to the industry×year fixed effects.

[16] As for the previous application, we perform the Im et al. (2003) unit root test and Pedroni's (1999; 2001) cointegration tests. According to these tests, MFP is non-stationary and cointegrated with US MFP. We assume that the *REG* indicator is stationary by nature, as NMR indicator values are limited.

[17] Leads and lags of first difference of the DOLS estimator are introduced for the MFP US alone, as the NMR (and so REG) evolves by a small number of steps in each country. Estimated effects of US MFP and *REG* are robust to the use of the OLS estimator or to the number of leads and lags for the DOLS estimator (see Table 14.10 in Appendix 2). Again, we use a Hausman test to conclude that we must use the DOLS estimator rather than the OLS estimator and that 1 lead and 1 lag for the first differences are enough.

Table 14.5 shows a statistically significant and positive catch-up effect. The estimates are close to one, which is called the "homogeneity assumption" in terms of convergence: any short-term divergence between country productivity should disappear in the long run. The estimated effect of the *REG* indicator is statistically significant and negative, except when all the two-way fixed effects are included, certainly due to the lack of the remaining variability (column 5). The values of *REG* effects differ only a little depending on the various sets of fixed effects introduced (columns 1 to 4). We illustrate further the economic significance of these estimation results via a simulation of the impact of reforms. Finally, when using the alternative regulatory indicator calculated as an unweighted average of NMR OECD indicators, the estimated effect is not statistically significant (country×year fixed effects are excluded from this specification to avoid collinearity). This emphasizes the importance of our estimation strategy: according to the estimation results in Table 14.5, the impact of upstream NMR on downstream productivity increases with the intensity of use of regulated intermediate inputs.

We also test the sensitivity of our estimation results to (i) uncorrected standard errors; (ii) introduction of the direct impact of NMR, i.e., the impact of regulation on the productivity of the regulated industry; and (iii) estimating factors' elasticities simultaneously to *REG* and catch-up effects rather than calibrating MFP. The estimated catch-up and *REG* effects on productivity are robust to these changes, with a slight increase (in absolute terms) of the *REG* effect when the direct impact of NMR is taken into account (see Tables 14.11, 14.12, and 14.13 in Appendix 2). Moreover, (i) we observe a statistically significant direct negative impact of NMR (that must be taken with caution, because of the within industry reverse causality); and (ii) we find believable estimated elasticities for each production factor.

To illustrate the implications of our estimation results, we propose a simple simulation of the impact of NMR reforms on MFP. The reforms considered are switched to the lightest practices, i.e., to the smallest values observed in our country sample of 2013 OECD NMR indicators (these indicators are not available later). Based on our privileged estimates (Table 14.5, column 2), the effects of these NMR reforms are calculated for each downstream industry, then aggregated at the national level using domestic industry shares in national value added.

Figure 14.1 presents the simulated MFP gains from these reforms. These gains are very different between countries, depending on their excess regulation in 2013. They are relatively high for some countries, with a maximum of 5.9% for Italy, but correspond to long-term gains of very ambitious reforms. For other countries with NMR indicators already close to the lightest practices, the expected MFP gains are relatively small, with a minimum of 1.5% for the Netherlands, Sweden, and the United Kingdom.

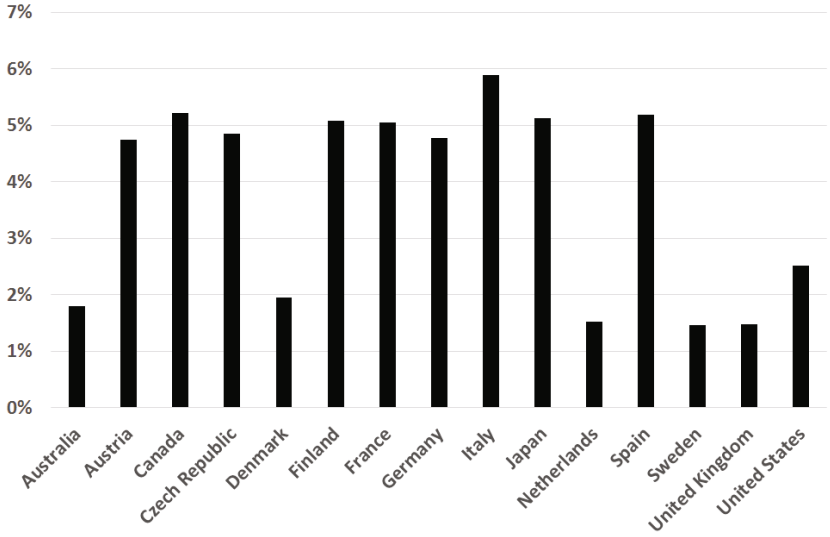

Fig. 14.1 Long-term MFP gains from reforms to the lightest practices in 2013

14.3.5 Discussion of Heterogeneous Effects with 2D Explanatory Variables

As mentioned in the first application, it may be particularly important to take account of the heterogeneous effects of the explanatory variables and unobservable common correlated factors. However, the approach in the second application introduces major differences with respect to this issue: (i) the construction of the *REG* indicator already takes into account part of the heterogeneity of NMR effects on MFP; (ii) the NMR indicators are 2D variables; and (iii) these indicators (and so the *REG* indicator) are stationary, as their values are limited. We discuss the consequences of these differences below.

As a first step, let us assume that we are concerned with only one upstream NMR indicator and there exist country×year unobserved correlated common factors F_{it}. Thus, relation (14.6) may be rewritten as

$$mfp_{ijt} = \alpha + \beta \cdot mfp_{jt}^{US} + \theta \cdot w_j \cdot NMR_{it} + \mu \cdot F_{it} + \varepsilon_{ijt} \ . \tag{14.7}$$

Because the NMR indicator is 2D, this relation shows that we cannot introduce country×year cross-sectional averages of dependent and explanatory variables to take into account the heterogeneous effects of the unobserved common factors. It also shows that our constraint on the impact of NMR (i.e., $\theta_j = \theta \cdot w_j$) is required to identify this impact when country×year fixed effects are introduced. In other words, this constraint on the coefficient may be added to other constraints depending

on various industry specific characteristics, but we cannot introduce unconstrained coefficients θ_j because of collinearity with the fixed effects.

Our analysis focuses not on one, but on four upstream industries and we construct the *REG* indicator in order to estimate the average effect of the four NMR indicators. This *REG* indicator is 3D, as we just show in the example with one NMR, we see that it is not desirable to treat for heterogeneous effects as we did in our first application: the *REG* effects would be very difficult to identify and interpret.

However, we may be able to go one step further in the sensitivity analysis. If we assume that NMR indicators are stationary, and so is the *REG* indicator, the cointegrated relationship concerns only MFP and US MFP variables. So we treat for heterogeneous US MFP effects, assuming industry and country specific effects ($\beta_{ij} = \beta_i + \beta_j$), then we augment the cointegrated relationship with the *REG* indicator and our set of fixed effects. This approach induces a substantial reduction of the average effect of US MFP, but the *REG* effect is robust to this change (see Table 14.14 in Appendix 2).[18]

14.4 Conclusion

Our estimation results are consistent with the empirical literature. More importantly, we show how confident we can be about these results, dealing with various heterogeneity issues despite the relatively short duration of our country-industry panel data.

We first assume the homogeneous effects of the explanatory variables and unobserved common factors. When the specification includes our preferred set of fixed effects, the estimated productivity elasticities of ICT and R&D capital stocks are 10% and 6%, respectively. These elasticities suggest high social surplus returns (in the absence of such returns, our accounting analysis would conclude there are elasticities of 3% and 4%, respectively). Our preferred set of fixed effects includes country×year and industry×year fixed effects to prevent specific omission bias, but not individual (country×industry) fixed effects. Considering our data as a 2D panel with country×industry individuals leads to almost systematically introducing individual fixed effects, resulting in unlikely estimates: in one case a very high R&D elasticity of 15%, in the other negative ICT elasticity.

Second, we use the 3D structure of our data to relax the homogeneous effect assumptions, using specifications derived from the Common Correlated Estimator and the Mean Group Estimator. The estimated elasticities are robust to estimating the heterogeneous effects of the explanatory variables or unobserved common factors separately, but the estimated ICT elasticity falls to 4% when both heterogeneities

[18] Unfortunately, the Im et al. (2003) test concludes to a unit root residual of this estimation, even though earlier Pedroni's (1999; 2001) tests conclude there is cointegration. This may be explained by the heterogeneous effects of the unobserved common factors if they are non-stationary. This underlines the difficulties in making inferences from short panel data with non-stationary variables when at least some of them are only 2D.

are taken into account simultaneously. As there is a priori no methodological reason for this change, this may indicate that we are over the limit of what we can do on our data.

Finally, we benefit also from the 3D structure of our data to prevent a reverse causality issue when estimating the NMR effects on productivity if governments react to economic shocks. According to our estimates, the impact of upstream NMR on downstream productivity grows with the intensity of use of regulated intermediate inputs, which is meaningful in terms of causality. The estimated NMR impact is sizeable: the effect of a switch for each country of 2013 NMR to the "lightest practices" would increase MFP by 3.8% on average. However, it is the long-run effect of extreme reforms.

This chapter shows that our simple approach to country-industry level data allows us to benefit from some macroeconomic panel data econometric methods, such as the common correlated estimator and the mean group estimator. We think that many other macroeconomic panel methods would be used by applied researchers on data such as ours if the three (or more) dimensional structure were taken into account. The second application also shows the limits of our approach. When some explanatory variables are two-dimensional, we may take benefits from the higher dimension of the others to give more insight into causality, as for NMR, but we are unable to relax the homogeneity assumptions on the effects of explanatory variables and unobserved correlated common factors. This seems to underline that the 3D structure is required to estimate heterogeneous effects on short panel data. However, more research is needed to overcome this difficulty as panels mixing two and three-dimensional variables are widely used.

References

Aghion, P., Blundell, R., Griffith, R., Howitt, P., and Prantl, S. (2009). The effects of entry on incumbent innovation and productivity. *Review of Economics and Statistics*, 91(1):20–32.

Allegra, E., Forni, M., Grillo, M., and Magnani, L. (2004). Anti-trust policy and national growth: Some evidence from Italy. *International Macroeconomics Discussion Paper Series, No. 4373, CEPR, London*.

Arnold, J., Javorcik, B., and Mattoo, A. (2011). Does services liberalization benefit manufacturing firms? *Journal of International Economics*, 85(1):136–146.

Barone, G. and Cingano, F. (2011). Service regulation and growth: Evidence from OECD countries. *The Economic Journal*, 121(555):931–957.

Bourlès, R., Cette, G., Lopez, J., Mairesse, J., and Nicoletti, G. (2013). Do product market regulations in upstream sectors curb productivity growth? Panel data evidence for OECD countries. *The Review of Economics and Statistics*, 95(5):1750–1768.

Cette, G., Lopez, J., and Mairesse, J. (2013). Upstream product market regulations, ICT, R&D and productivity. *NBER Working Paper Series, 19488, forthcoming in*

Review of Income and Wealth.

Cette, G., Lopez, J., and Mairesse, J. (2016). Market regulations, prices and productivity. *American Economic Review: Papers and Proceedings*, 106(5):104–108.

Cobb, C. and Douglas, P. (1928). A theory of production. *The American Economic Review, Papers and Proceedings of the Fortieth Annual Meeting of the American Economic Association*, 18(1):139–165.

Conway, P. and Nicoletti, G. (2007). Product market regulation and productivity convergence: OECD evidence and implications for Canada. *International Productivity Monitor*, 15:3–24.

Faini, R., Haskel, J., Navaretti, G., Scarpa, C., and Wey, C. (2006). Contrasting Europe's decline: Do product market reforms help? In T. Boeri, M. Castanheira, R. F. and Galasso, V., editors, *Structural Reforms Without Prejudices*. Oxford University Press.

Forlani, E. (2010). Competition in the service sector and the performances of manufacturing firms: Does liberalization matter? *CESifo Working Paper series, 2942*.

Hall, B., Mairesse, J., and Mohnen, P. (2010). Measuring the returns to R&D. In Hall, B. and Rosenberg, N., editors, *Handbook of the Economics of Innovation*. Elsevier, 1 edition.

Hurlin, C. and Mignon, V. (2007). Une synthèse des tests de cointégration sur données de panel. *Economie et Prévision*, 180–181:241–265.

Im, K., Pesaran, M., and Shin, Y. (2003). Testing for unit roots in heterogeneous panels. *Journal of Econometrics*, 115(1):53–74.

Jorgenson, D., Ho, M., and Samuels, J. (2016). The impact of information technology on postwar US economic growth. *Telecommunications Policy*, 40:398–411.

Kocoglu, Y. and Mairesse, J. (2005). Issues in measuring knowledge: The contribution of R&D and ICT to growth. *CEDERS Working Papers*.

Newey, W. and West, K. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708.

Pedroni, P. (1999). Critical values for cointegration tests in heterogeneous panels with multiple regressors. *Oxford Bulletin of Economics and Statistics*, 61:653–670.

Pedroni, P. (2001). Purchasing power parity tests in cointegrated panels. *The Review of Economics and Statistics*, 83(4):727–731.

Pedroni, P. (2004). Panel cointegration: Asymptotic and finite sample properties of pooled time series tests with an application to the PPP hypothesis. *Econometric Theory*, 20(3):597–625.

Pesaran, H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica*, 74(4):967–1012.

Pesaran, H., Shin, Y., and Smith, R. (1999). Pooled mean group estimation of dynamic heterogeneous panels. *Journal of the American Statistical Association*, 94:289–326.

Stiroh, K. (2005). Reassessing the impact of IT in the production function: A meta-analysis and sensitivity tests. *Annales d'Economie et de Statistique*, 79/80:529–561.

Stock, J. and Watson, M. (1993). A simple estimator of cointegrating vectors in higher order integrated systems. *Econometrica*, 61(4):783–820.
Timmer, M., Inklaar, R., O'Mahony, M., and van Ark, B. (2011). Productivity and economic growth in Europe: A comparative industry perspective. *International Productivity Monitor*, 21:3–23.
Venturini, F. (2015). The modern drivers of productivity. *Research Policy*, 44(2):357–369.
Wooldridge, J. (2002). *Econometric Analysis of Cross Section and Panel Data*. Cambridge, Massachusetts, The MIT Press.

Appendix

Appendix 1: Data

To investigate the impact of ICT and R&D or NMR on productivity, we use an unbalanced panel of 4,750 observations (4,624 when mobilizing the regulatory indicators, see relation (14.6)) from 15 OECD countries and 18 industries, covering the period from 1987 to 2008. This sample is reduced to 3,999 observations when using the Dynamic OLS estimator (4,099 observations for relation 2). The 15 countries are: Australia, Austria, Canada, the Czech Republic, Denmark, Finland, France, Germany, Italy, Japan, the Netherlands, Spain, Sweden, the United Kingdom, and the United States. The 18 industries cover the manufacturing sector (food products, textiles, wood products, paper, chemical products, non-metallic mineral products, metal products, machinery not elsewhere classified, electrical equipment, transport equipment, manufacturing not elsewhere classified) as well as network and market service industries (energy, construction, retail distribution, hotels & restaurants, transport & communication, banking services, and professional services).

We mobilize data from four databases: (i) STAN for value added and total employment in number of persons engaged; (ii) EU KLEMS for physical capital; (iii) ANBERD for R&D investment; and (iv) OECD indicators of NMR for the second application. Since R&D was not treated as investment in the national accounts collected by the OECD, we have to correct both the industry value added by adding (expensing out) the intermediate consumption of their R&D activities, and the industry number of employees by subtracting the number of R&D personnel ("avoiding double counting"). We compute capital stock using the permanent inventory method and assuming constant geometric rates of depreciation: 5% for non-residential constructions, 10% for non-ICT equipment, 15% for communication equipment, 30% for hardware and software, and 25% for R&D.[19]

[19] We compute separately the capital stock of communication equipment, hardware, and software, and then aggregate them into the ICT capital stock. To compute these three ICT capital stocks, we assume for all countries that the ratio of investment prices over the GDP price is the same as for the USA. This is much better for comparability since the USA is the country that most extensively

To compute the partial MFP (MFP^p), we calibrate the non-ICT equipment and non-residential construction elasticities $\tilde{\gamma}_{NI}$ and $\tilde{\gamma}_c$ as industry specific average shares of their user costs computed for the USA over the whole estimation period. Our estimation results are robust to this choice. In order to make a comparison with estimated values, it is interesting to note the values of these cost shares for the four capital production factors, averaged over industries: 0.13 for non-ICT equipment, 0.06 for non-residential construction, 0.04 for ICT equipment, and 0.07 for R&D.

Figure 14.2 presents the industry sample averages of the ICT and R&D coefficients, i.e., the ratio of investment over value added, over the period 1987–1997 and 1998–2008. It shows big differences between industries. In particular, some industries invest almost nothing in R&D: wood products, energy, construction, retail distribution, hotels & restaurants, and banking services. Therefore, we estimate the R&D output elasticity on just the 12 industries investing heavily in R&D (our estimation results are robust to this choice of industry exclusion).

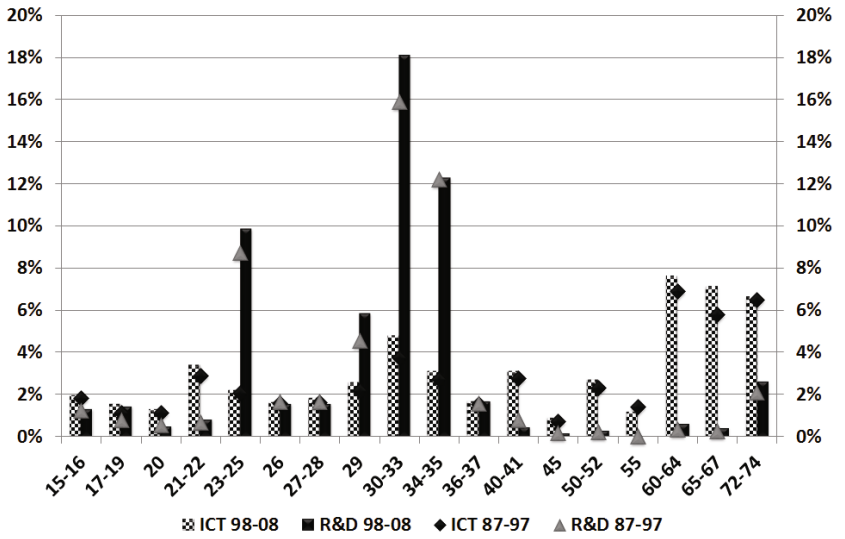

Fig. 14.2 ICT and R&D coefficients (ratio of investment over value added). Notes: The 2-digit (ISIC Rev. 3) industries (with their codes in parentheses) are: food products (15-16), textiles (17-19), wood products (20), paper (21-22), chemical products (23-25), non-metallic mineral products (26), metal products (27-28), machinery n.e.c. (29), electrical equipment (30-33), transport equipment (34-35), manufacturing n.e.c. (36-37), energy (40-41), construction (45), retail distribution (50-52), hotels & restaurants (55), transport & communication (60-64), banking services (65-67), and professional services (72-74).

relies on hedonic methods for measuring these prices. Note also that we have to modify the price index of value added, and hence its value at constant prices for the electrical and optical equipment industry, which includes ICT equipment. We assume for ICT investment that in this industry the ratio of value added prices to the GDP price is the same in all countries in the same way as in the USA.

After estimating the impact of ICT and R&D capital on productivity, we investigate the effects of upstream NMR on downstream productivity. We use the OECD NMR indicators measuring the extent to which competition and firms' choices are restricted where there are no a priori reasons for government interference, or where regulatory goals could plausibly be achieved by less coercive means. These indicators are based on detailed information on laws, rules, and market and industry settings and they cover energy (gas and electricity), transport (rail, road, and air) and communications (post, fixed and cellular telecommunications), retail distribution, and professional services (see Conway and Nicoletti, 2007 for a more detailed presentation).

Because of lack of data variability, we combine the NMR indicators considering that their individual impacts are most likely to vary with the respective importance of upstream industries as suppliers of intermediate inputs, and computing the following indicator of regulatory impact (*REG*)

$$REG_{ijt} = \sum_{k \neq j} w_j^k \cdot NMR_{it}^k \,, \tag{14.8}$$

where NMR_{it}^k is the NMR indicator of the upstream industry k for country i in year t, and w_j^k stands for the intensity-of-use of intermediate inputs from industry k by industry i, as measured from the US 2000 input-output table as the ratio of the intermediate inputs from industry k to industry j over the total output of industry j.

We prefer to use a fixed reference input-output table to compute the intensity-of-use ratios, rather than the different country and year input-output tables, to avoid endogeneity biases that might arise from potential correlations between such ratios and productivity or ICT and R&D, since the importance of upstream regulations may well influence the use of domestic regulated intermediate inputs. We actually used the 2000 input-output table for the USA. For similar endogeneity, as well as measurement error concerns, note also that in estimating *REG* for the upstream industries, we exclude within-industry intermediate consumption. Note that the USA is excluded from the estimation sample.

Figure 14.3 presents the contribution of each upstream industry regulation to the value of the *REG* indicator for 1987 and 2007. It shows that the reduction of the NMR indicators leads to a marked decrease in *REG* over-time in every country, but important differences remain between countries, notably when observing each upstream industry separately.

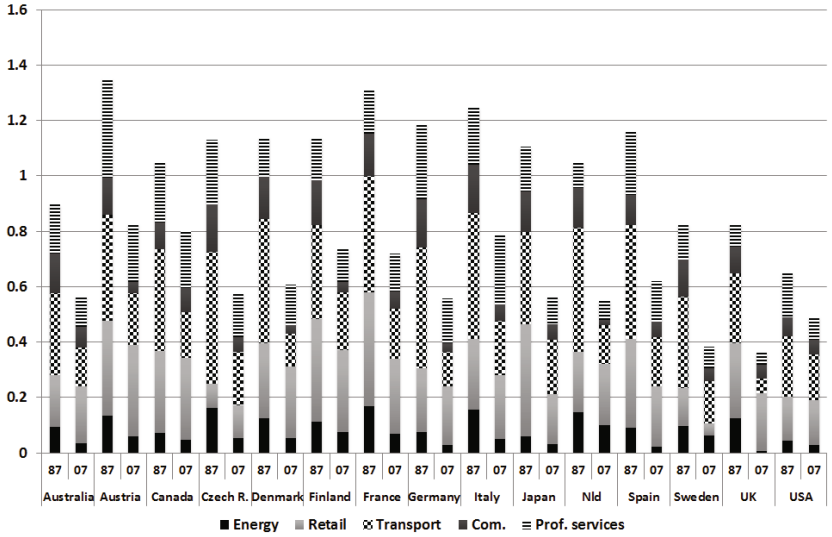

Fig. 14.3 Non-Manufacturing Regulation (NMR) OECD indicators. Notes: Scale 0-6 for each indicator, 0 for the most pro-competitive

Appendix 2: Supplementary Estimation Tables

Table 14.6 Sensitivity to the DOLS estimators – application 1

	(1)	(2)	(3)	(4)	(5)
Estimator	OLS	DOLS			
Number of leads and lags	-	1	2	3	4
ICT capital intensity	0.0802***	0.101***	0.109***	0.116***	0.123***
	[0.00634]	[0.00717]	[0.00772]	[0.00837]	[0.00926]
R&D capital intensity[a]	0.0641***	0.0612***	0.0610***	0.0605***	0.0621***
	[0.00498]	[0.00542]	[0.00577]	[0.00624]	[0.00692]
Observations	4,750	3,991	3,531	3,041	2,548
R-squared	0.991	0.991	0.992	0.992	0.992
RMSE	0.188	0.182	0.179	0.176	0.174
Residual degree of freedom	4,068	3,399	2,998	2,568	2,135

Notes: Dependent variable: Partial MFP. Standard errors in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Country×year and industry×year fixed effects are included in all specifications.
[a] The R&D output elasticity is estimated on 12 industries as six industries make almost no investment in R&D.

Table 14.7 Sensitivity to fixed effects, using uncorrected standard errors – application 1

	(1)	(2)	(3)	(4)	(5)	(6)
ICT capital intensity	0.0934***	0.0999***	0.0925***	0.0683***	0.101***	−0.024***
	[0.00947]	[0.00995]	[0.00692]	[0.0161]	[0.00717]	[0.00891]
R&D capital intensity[a]	0.0740***	0.0691***	0.0645***	0.146***	0.0612***	0.0637***
	[0.00719]	[0.00732]	[0.00543]	[0.0130]	[0.00542]	[0.00889]
Observations	3,991	3,991	3,991	3,991	3,991	3,991
R-squared	0.980	0.981	0.991	0.988	0.991	0.998
RMSE	0.258	0.258	0.185	0.203	0.182	0.0821
Residual degree of freedom	3,933	3,704	3,628	3,714	3,399	3,180
Fixed effects:						
Country, industry, year	Y	Y	Y	Y	Y	Y
Country×year	N	Y	N	N	Y	Y
Industry×year	N	N	Y	N	Y	Y
Country×industry	N	N	N	Y	N	Y

Notes: Dependent variable: Partial MFP, estimator: DOLS(1;1). Standard errors in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

[a] The R&D output elasticity is estimated on 12 industries as six industries make almost no investment in R&D.

Table 14.8 Sensitivity to fixed effects of the labor productivity estimations – application 1

	(1)	(2)	(3)	(4)	(5)	(6)
Non-ICT Eq. capital intensity	0.0950***	0.106***	0.0790***	0.181***	0.0895***	0.169***
	[0.0168]	[0.0178]	[0.0143]	[0.0551]	[0.0149]	[0.0320]
Structure capital intensity	0.0668***	0.0742***	0.0518***	0.288***	0.0520***	0.0501
	[0.0157]	[0.0157]	[0.0121]	[0.0596]	[0.0124]	[0.0313]
ICT capital intensity	0.0850***	0.0838***	0.0996***	0.0396*	0.0989***	−0.0256*
	[0.0136]	[0.0145]	[0.0108]	[0.0221]	[0.0115]	[0.0143]
R&D capital intensity[a]	0.0744***	0.0687***	0.0683***	0.103***	0.0629***	0.0664***
	[0.0102]	[0.0100]	[0.00756]	[0.0228]	[0.00764]	[0.0141]
Observations	3,991	3,991	3,991	3,991	3,991	3,991
R-squared	0.774	0.786	0.874	0.874	0.901	0.982
RMSE	0.259	0.260	0.199	0.199	0.184	0.0814
Residual degree of freedom	3,925	3,696	3,619	3,706	3,391	3,172
Fixed effects:						
Country, industry, year	Y	Y	Y	Y	Y	Y
Country×year	N	Y	N	N	Y	Y
Industry×year	N	N	Y	N	Y	Y
Country×industry	N	N	N	Y	N	Y

Notes: Estimator: DOLS(1;1), dependent variable: Labour Productivity. Newey-West standard errors in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

[a] The R&D output elasticity is estimated on 12 industries as six industries make almost no investment in R&D.

Table 14.9 Sensitivity to the omission of explanatory variables

	(1)	(2)	(3)
ICT capital intensity	0.101***	0.116***	
	[0.0105]	[0.0109]	
R&D capital intensity[a]	0.0612***		0.0764***
	[0.00727]		[0.00799]
Observations	3,991	3,991	3,991
R-squared	0.991	0.991	0.991
RMSE	0.182	0.185	0.188
Residual degree of freedom	3,399	3,403	3,403

Notes: Estimator: DOLS(1;1), dependent variable: Partial MFP. Newey-West standard errors in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Country×year and industry×year fixed effects are included in all specifications.
[a] The R&D output elasticity is estimated on 12 industries as six industries make almost no investment in R&D.

Table 14.10 Sensitivity to the DOLS estimators – application 2

	(1)	(2)	(3)	(4)	(5)
Estimator	OLS	DOLS			
Number of leads and lags		1	2	3	4
US MFP	0.873***	0.910***	0.922***	0.927***	0.936***
	[0.0142]	[0.0168]	[0.0193]	[0.0236]	[0.0349]
Regulatory impact	−0.293***	−0.285***	−0.271***	−0.246***	−0.222***
	[0.0523]	[0.0558]	[0.0590]	[0.0628]	[0.0679]
Observations	4,624	4,099	3,706	3,303	2,865
R-squared	0.992	0.992	0.992	0.992	0.993
RMSE	0.198	0.193	0.190	0.188	0.187
Residual degree of freedom	4,297	3,807	3,438	3,060	2,648

Notes: Dependent variable: Partial MFP. Standard errors in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Country×year and industry fixed effects are included in all specifications.

Table 14.11 Sensitivity to the fixed effects using uncorrected standard errors – application 2

	(1)	(2)	(3)	(4)	(5)	(6)
US MFP	0.907***	0.910***	[a]	0.853***	[a]	0.906***
	[0.0168]	[0.0168]		[0.00886]		[0.0168]
Regulatory impact (REG)	−0.142***	−0.285***	−0.141**	−0.384***	0.0116	
	[0.0451]	[0.0558]	[0.0641]	[0.0525]	[0.0891]	
Alternative regulatory impact						0.297
						[0.299]
Observations	4,099	4,099	4,099	4,099	4,099	4,099
R-squared	0.992	0.992	0.993	0.998	0.999	0.991
RMSE	0.193	0.193	0.190	0.103	0.084	0.194
Residual degree of freedom	4,044	3,807	3,519	3,592	3,301	4,044
Fixed effects:						
Country, industry, year	Y	Y	Y	Y	Y	Y
Country×year	N	Y	Y	Y	Y	N
Industry×year	N	N	Y	N	Y	N
Country×industry	N	N	N	Y	Y	N

Notes: Estimator: DOLS(1;1), dependent variable: MFP. Standard errors in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.
[a] US MFP is by definition collinear to the industry×year fixed effects.

Table 14.12 Estimation of the direct effect of NMR

	(1)	(2)
US MFP	0.910***	0.913***
	[0.0168]	[0.0311]
Regulatory impact (REG)	−0.285***	−0.502***
	[0.0558]	[0.0835]
Direct regulatory impact		−0.0609***
		[0.00970]
Observations	4,099	4,099
R-squared	0.992	0.992
RMSE	0.193	0.191
Residual degree of freedom	3,807	3,806

Notes: Estimator: DOLS(1;1), dependent variable: MFP. Newey-West standard errors in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Country×year and industry fixed effects are included in all specifications.

Table 14.13 Sensitivity to fixed effects of the labor productivity estimations – application 2

Dep. Variable	(1) MFP	(2)	(3) Partial MFP	(4) Labour productivity	(5)
US MFP	0.923***				
	[0.0320]				
US Partial MFP			0.886***		
			[0.0288]		
US LP					0.856***
					[0.0281]
Regulatory impact (REG)	−0.321***		−0.251***		−0.270***
	[0.0762]		[0.0748]		[0.0749]
ICT capital intensity		0.103***	0.0953***	0.0871***	0.0918***
		[0.0136]	[0.0105]	[0.0148]	[0.0115]
R&D capital intensity[a]		0.0698***	0.0541***	0.0691***	0.0550***
		[0.0102]	[0.00715]	[0.0105]	[0.00751]
Non-ICT Eq. capital intensity				0.0999***	0.0863***
				[0.0178]	[0.0148]
Structure capital intensity				0.0803***	0.0558***
				[0.0161]	[0.0123]
Observations	3,857	3,857	3,857	3,857	3,857
R-squared	0.992	0.981	0.990	0.782	0.890
RMSE	0.190	0.261	0.184	0.262	0.186
Residual degree of freedom	3,584	3,581	3,576	3,573	3,568

Notes: Estimator: DOLS(1;1). Newey-West standard errors in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Country×year and industry fixed effects included in all specifications.
[a] The R&D output elasticity is estimated on 12 industries as six industries make almost no investment in R&D.

Table 14.14 Estimation of REG effects allowing for heterogeneous effects

	(1)	(2)
US MFP	0.910***	0.347***
	[0.0168]	[0.0768]
Regulatory impact (REG)	−0.285***	−0.232***
	[0.0558]	[0.0584]
Observations	4,099	4,099
R-squared	0.992	0.994
RMSE	0.193	0.177
Residual degree of freedom	3,807	3,699
Heterogeneous effects of observable variables	N	Y

Notes: Estimator: DOLS(1;1), dependent variable: MFP. Standard errors in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Country×year and industry fixed effects are included in all specifications. Column 2 presents unweighted average of industry and country specific coefficients.

Chapter 15
The Determinants of Consumer Price Dispersion: Evidence from French Supermarkets

Nicoletta Berardi, Patrick Sevestre, and Jonathan Thébault

Abstract In this chapter, we characterize the dispersion of grocery prices in France based on a large original data set of prices in more than 1500 supermarkets across the country. On average across products, the 90^{th} percentile of relative prices is 17 percentage points higher than the 10^{th}. The mean absolute deviation from quarterly average product prices is 5% on average in the French retail sector, and the standard deviation of relative prices is 7%. We show that temporary sales and promotions offer a limited explanation of the observed price dispersion, while the permanent component of price dispersion largely dominates. We find that in France price dispersion across stores essentially results from persistent heterogeneity in retail chains' national pricing. Indeed, consumer prices are largely determined at a national level by retail groups' bargaining power with producers and by retail chains' positioning. We also show, however, that local conditions regarding demand and local competition between supermarkets do explain prices observed in local markets, though to a much lower extent.

15.1 Introduction

Why do prices of the same good often significantly differ across shops? Deviations from the law of one price have attracted the attention of economists for a long time. However, due to the scarcity of adequate microeconomic price data, the first studies on price dispersion could only focus on specific markets. For instance, Borenstein

Nicoletta Berardi
Banque de France, Paris, France

Patrick Sevestre
Greqam-AMSE, Université Aix-Marseille, Marseille, France.

Jonathan Thébault
French Embassy in Brazil, Brasilia, Brazil

L. Matyas (ed.), *The Econometrics of Multi-dimensional Panels*, Advanced Studies in Theoretical and Applied Econometrics 50, DOI 10.1007/978-3-319-60783-2_15

and Rose (1994) and Gerardi and Shapiro (2009) zoomed in on the airline industry in the US. Sorensen (2000) analyzed the distribution of prices for several drugs across different pharmacies in two cities in upstate New York. Hong and Shum (2006) documented the distribution of prices posted by online booksellers for four academic textbooks. Moraga-González and Wildenbeest (2008) did the same for online sellers of several computer memory chips, while Woodward and Hall (2012) for mortgage brokerage services.

The last fifteen years have seen a significant improvement in researchers' access to microeconomic data on prices. While the access to CPI price records allowed thorough analyses of the dynamics of consumer prices and of the link between overall inflation and price dispersion,[1] the characteristics of CPI data, as released to researchers, most often do not allow a precise analysis of price dispersion.[2] Fortunately, household surveys or scanner data collected by marketing companies, such as Nielsen or Kantar, have recently been made available. These data sets generally contain prices of very precisely defined products in different outlets at different dates, thus allowing a precise quantification of price dispersion.

Asplund and Friberg (2002) use price data from a household survey carried out by the Swedish Pensioners' National Organization for five grocery products (sugar, washing-up detergent, crispbread, spread, and cocoa) sold in about 1000 supermarkets in Sweden between 1993 and 1997 and find that most of the price dispersion is explained by supermarket specific factors (like supermarket size and retail chain) and, to a lesser extent, by factors related to sales space and transport costs. Kaplan and Menzio (2015), based on the Kilts-Nielsen Consumer Panel Data set in the US over the period 2004-2009, find that in the US the variance of prices is mainly accounted for by idiosyncratic supermarket characteristics and time variation, and only to a limited extent by the retailer to which the supermarket belongs. They also find that the typical distribution of normalized prices is characterized by a unique mode and by a 19% average standard deviation for identical goods, and that it is symmetric and leptokurtic. As far as the French retailing sector is concerned, Dubois and Perrone (2015) have been the only ones to study price dispersion at barcode level. Their analysis is based on a household survey about consumers' purchases of food products over 3 years (1999, 2000, and 2001). Based on four product categories (beer, coffee, cola, and whisky), they show that price dispersion is prevalent in the French food market.

This chapter assesses and analyzes consumer price dispersion in France. We are able to characterize the overall shape and structure of price dispersion in the French retail sector based on an original data set containing almost 40 million weekly price records from geo-localized medium and large supermarkets in France. After trimming, there are about one and a half million price trajectories regarding one thousand products (individually identified by a barcode) sold in more than 1500 supermarkets from October 2011 to September 2012.

---

[1] E.g., see Nakamura et al. (2016) for the US, Dixon et al. (2014) for the UK, Alvarez et al. (2013) for Argentina, Moen et al. (2014) for Norway, and Eyal and Eden (2004) for Israel.

[2] One exception is Cheung and Fujii (2008) for Japan.

Our contribution in this chapter is twofold. First, we characterize the overall dispersion of consumer prices in the French retail sector. On average across products, the 90^{th} percentile of relative prices is 17 percentage points higher than the 10^{th}. The mean absolute deviation with quarterly average product prices as the measure of central tendency is 5% on average in the French retail sector, and the standard deviation of relative prices is 7%. Second, we decompose the observed price dispersion into several components. We first assess the relative contribution of sales and temporary promotions and that of permanent price differences across shops to the observed price dispersion at the barcode level. In a second stage we then show that the latter strongly depends on retail chains, while local markets' characteristics play a significant role but to a much lower extent.

The remainder of the chapter is structured as follows. A detailed presentation of the data sets is given in Sect. 15.2. Section 15.3 characterizes descriptively price dispersion in France and investigates its structure, while Sect. 15.4 disentangles the permanent and transitory components of price dispersion and further investigates the role played by national and local factors as far as the former is concerned. Section 15.5 concludes.

15.2 Data Description

Our analysis is based on the combination of two original data sets. The first provides millions of price records regarding a thousand products sold in more than 1500 medium and large supermarkets, together with precise identifiers of both products (through their EAN barcodes)[3] and supermarkets (through their names and addresses). The second is an exhaustive data base of all supermarkets in France, including their names and locations.

15.2.1 Grocery Price Data

Our analysis of price dispersion in the grocery sector is based on millions of price records from geo-localized medium and large supermarkets in France. The price data was collected by Prixing, a start-up that offers a mobile price comparison app and website (http://www.prixing.fr/). Prixing has developed automatic procedures enabling it to collect price lists from supermarkets offering "click&collect" services, known as "drives" in France (which refers to the fact that customers drive close the supermarket). This retail channel is a relatively recent

[3] The EAN barcodes (originally European Article Number, now renamed International Article Number even though the abbreviation EAN has been retained) are international product identification numbers.

service developed by French supermarkets over the past 10 years.[4] It has two distinctive features: customers order their shopping on the supermarket's click&collect website and then collect it at a drive-through facility.[5] More importantly for our analysis, prices are exactly the same as those of the brick-and-mortar supermarket associated with the drive-through.[6] Therefore, our price data base corresponds to grocery prices in medium and large supermarkets, which represent more than 80% of grocery sales in France (Anderton et al., 2011).

The original database contains slightly more than 45 million price spells corresponding to 2.3 billion daily prices of about 90 thousand products in about 1600 supermarkets from October 2011 to September 2012.[7] Due to computational challenges, we reduced the size of the data set. First, we based our analysis on one price observation per week, chosen as the most frequently observed price over a week for each specific product sold in a specific supermarket (i.e., the mode of the weekly price distribution). This does not induce any significant loss of information regarding price dispersion, as the within week price variance is null in almost 99% of the cases. Second, to keep our econometric estimation feasible, we also restricted the sample to the one thousand most widely sold products (i.e., 1000 barcodes). The resulting weekly modal price data set contains more than 37 million observations (almost 1.5 million trajectories) of the most widely sold products.[8]

One advantage of our price data with respect to those collected from scanners or from household surveys is that each price does correspond to what the consumer really paid for a product on the day the data was collected and not to an average unit value computed from several transactions recorded over a period.

Table 15.1 shows the category of available products based on the Classification of Individual Consumption by Purpose (COICOP). Both in terms of the number of

[4] There were around 500 drives in France at the end of 2010, their number doubled by the end of 2011, to reach almost 2000 by the end of 2012, and about 2700 by the end of 2013 (Dauvers, 2013a,b).

[5] Most click&collect drive-throughs are associated with a supermarket, but there are also a few stand alone drive-throughs (known as "drives-entrepôt" in French).

[6] One of the major retail chains in France at some point tested a different pricing strategy whereby a few products are cheaper when bought through the click&collect than in the associated supermarkets. However, this practice was not yet implemented at the time of the collection of the data we use here. Generally, it may happen that some discounts are available in the brick-and-mortar store and not in click&collect and vice-versa.

[7] A "price spell" is made up of the following three elements:

- the price of a precisely defined product *i*, identified by its barcode (e.g., a 1 liter glass bottle of brand *b* pure orange juice) in a given store *s* (e.g., the supermarket from retail chain *r*, located at a given address),
- the date when this price was first set (start date of the price spell),
- the date when this price was changed, even if only temporarily (end date of the price spell).

For instance, if the price of a product in a store decreases from 1 euro to 80 euro cents and then increases to 90 euro cents, this will define 3 price spells.

[8] We discarded spells with inconsistent start and end dates (e.g., a start date posterior to the end date). We also dropped null prices (4 in the whole sample), as well as observations of prices whenever their log-difference to the average national price was larger than 2 in absolute value.

products and corresponding price observations, there is a strong predominance of food and beverages (almost 77% of barcodes); about 13% of barcodes are "miscellaneous" goods like personal care products (e.g., toothpaste, shampoo, etc.) and about 6% "furnishings and household equipment and routine maintenance" (e.g., washing-up liquid or dishwasher detergent) and 4% are items of "recreation and culture" (e.g., CDs, toys, etc.). The last two columns report the number of brands and producers by COICOP-level 1 product category.

Table 15.1 Numbers and percentages of products by COICOP product categories and brand/manufacturer

COICOP-level 1	Prices		Products		Brands	Manufact.
Food & non-alc.bev.	26231005	70.6%	703	70.3%	250	76
Alc.beverages	2320956	6.3%	66	6.6%	42	22
HH eqpt.&maint.	1992881	5.4%	55	5.5%	26	8
Recreation&culture	1580217	4.2%	42	4.2%	14	3
Miscellaneous	5005850	13.5%	134	13.4%	39	12
Total	37130909	100%	1000	100%	371	121

15.2.2 Supermarket Data and Competition Measures

The second source of information that we exploit is an exhaustive data base of all medium and large supermarkets in France.[9] For each supermarket it includes in particular their name, location, retail chain and regional branch/wholesaler.[10] In this data set, we identified the supermarkets for which we have price records.[11]

Table 15.2 shows the retail chains of the supermarkets for which we have price records. Those supermarkets belong mostly to the retail groups Carrefour, Systeme U and Les Mousquetaires with more than 18% of supermarkets each (second column of Table 15.2).[12] Some retail groups are over-represented with respect to their share in the total population of medium and large supermarkets. For instance, almost 30% of the sample of supermarkets for which we have price records belong to the group Systeme U and 14% to Leclerc, while at the national level Systeme U has 16.5% and Leclerc 8.4% of the supermarkets. However, Carrefour is under-represented

[9] The data was bought from LSA (http://expert.lsa-conso.fr/).

[10] In France, retail groups have regional branches, which in some cases operate as wholesalers, and in other cases do not directly bargain with producers, but have a role in setting prices in their region.

[11] In about 50 cases, finding the right correspondence proved to be problematic. We then discarded price observations relative to those stores, in order to avoid making mistakes in attributing the wrong store specific characteristics to our data set of price observations.

[12] Note that we have omitted retail chains for which no price records are available, and normalized the percentages of the remaining ones to 100%.

among the supermarkets with price observations with respect to the total number of Carrefours in France. Note also that, since click&collect are more frequently associated with large rather than with medium size supermarkets, retail chains of large supermarkets are over-represented with respect to those typically having smaller supermarkets. Indeed, about 40% of supermarkets for which we have price records are large supermarkets. This proportion is close to their market share in the French retail sector (see Anderton et al., 2011). The last column of Table 15.2 reports the retail chains' market shares in 2012 as reported by Kantar Worldpanel-LSA.

Table 15.2 Stores' retail chain/group and market shares

Retail chain	N.store (w/prices)	%	N.all stores (med.-large)	%	Mrkt shares [Kantar]
Group AUCHAN:	*92*	*5.9%*	*443*	*7.0%*	*14.5%*
Auchan (large supermkts)	55	3.5%	139	1.8%	11.2%
Simply Market (medium)	37	2.4%	304	4.0%	3.3%
Group CARREFOUR:	*286*	*18.2%*	*1733*	*27.7%*	*24.9%*
Carrefour (large)	186	11.9%	226	3.0%	14.8%
Carrefour Market (large&med.)	100	6.4%	1507	19.8%	10.1%
Group CASINO:	*126*	*8.0%*	*506*	*8.1%*	*6.7%*
Geant Casino (large)	87	5.5%	150	2.0%	3.5%
Casino (large&medium)	39	2.5%	356	4.7%	3.2%
Group SYSTEME U:	*468*	*29.8%*	*1033*	*16.5%*	*12.2%*
Hyper U (large)	55	3.5%	64	0.8%	
Marché U (medium)	7	0.5%	8	0.1%	
Super U (large&medium)	390	24.8%	717	9.4%	
U Express (medium)	16	1.0%	244	3.2%	
Group LECLERC:					
Leclerc (large)	222	14.1%	641	8.4%	24.8%
Group MOUSQUETAIRES:					
Intermarché (large&medium)	377	24.0%	1837	24.1%	16.8 %
Total	1571	100%	7596	100%	100.0%

The supermarkets for which we have price records are a rather representative sample in terms of their geographical distribution (see Fig. 15.11 in the Appendix as compared to the universe of medium and large supermarkets in France represented in Fig. 15.10). We also consider their location in smaller geographical areas or local markets, which correspond to French "arrondissements" (which may roughly be translated into English as districts). Prices in our data set concern supermarkets located in 286 local markets (out of 330 in mainland France). On average, there

are about 13 local markets per region and, in each local market, 6 supermarkets for which we have price records.

15.3 Assessing Price Dispersion in the French Retail Sector

In this chapter, we define price dispersion as price differences within a quarter for exactly the same product (i.e., barcode) sold in different supermarkets.[13]

In order to measure price dispersion, we first compute percentage price deviations from a quarterly reference price for each product. In particular, we define the relative price of a product i in a quarter q as $p_{ist}^{rel(iq)} = \ln(p_{ist}) - \overline{\ln(p_{iq})}$, where p_{ist} is the price of a product i sold in supermarket s in week t and $\overline{\ln(p_{iq})}$ is the average of the quarterly log price of product i over supermarkets. Figure 15.1 shows the distribution of the full set of relative prices in our data.

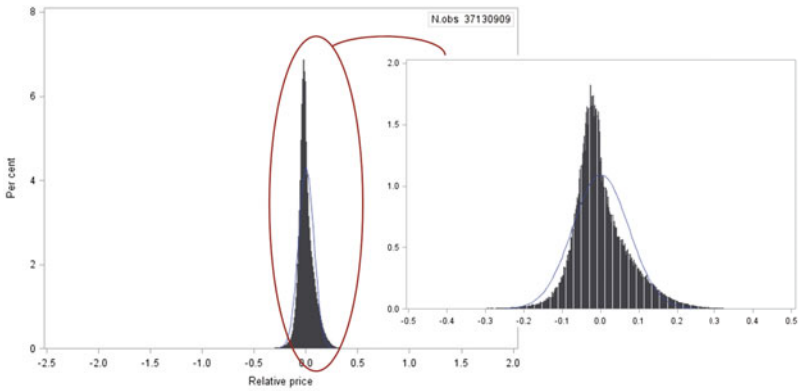

Fig. 15.1 Distribution of relative prices

Price dispersion can then be summarized in different ways. Among the possible measures, there are

- the standard deviation of relative prices: $\sigma(p_{ist}^{rel(iq)})$,
- the interquartile range of relative prices: $IQ(p_{ist}^{rel(iq)}) = Q3(p_{ist}^{rel(iq)}) - Q1(p_{ist}^{rel(iq)})$,
- the interdecile range of relative prices: $ID(p_{ist}^{rel(iq)}) = P90(p_{ist}^{rel(iq)}) - P10(p_{ist}^{rel(iq)})$,
- the mean of absolute deviation : $MAD(p_{ist}^{rel(iq)}) = 1/4n \sum_{i=1}^{n} \sum_{q=1}^{4} |p_{ist}^{rel(iq)}|$.

Table 15.3 provides the standard deviation, interquartile range, interdecile range, and the mean of absolute relative prices computed over all observations. The mean

[13] Price dispersion may also be considered using less strict definitions of a product. Some studies, for instance, also look at price differences for the same product categories (e.g., Kaplan and Menzio, 2015).

absolute deviation of relative prices is about 5% in France and the interquartile range 8 is percentage points, suggesting a non-trivial overall price dispersion. The distribution of the mean absolute deviation of relative prices by product is shown in Fig. 15.2.

Table 15.3 Different measures of price dispersion

Measures of price dispersion		
Standard deviation of relative prices	$\sigma(p_{ist}^{rel(iq)})$	0.07
Interquartile range of relative prices	$Q3(p_{ist}^{rel(iq)}) - Q1(p_{ist}^{rel(iq)})$	0.08
Interdecile range of relative prices	$P90(p_{ist}^{rel(iq)}) - P10(p_{ist}^{rel(iq)})$	0.17
Mean absolute deviation	$1/4n \sum_{i=1}^{n} \sum_{q=1}^{4} \lvert p_{ist}^{rel(iq)} \rvert$	0.05
N.obs		37,130,909
N.products		1,000

Note: Statistics are computed at the barcode level and then averaged across barcodes.

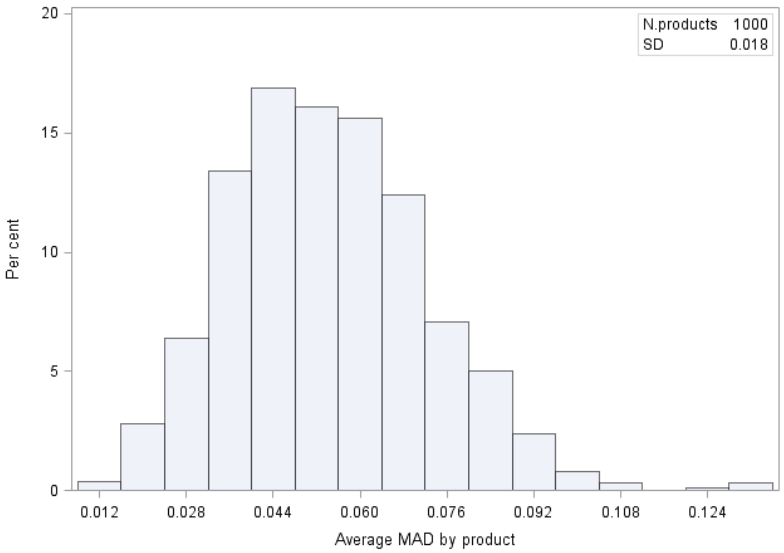

Fig. 15.2 Distribution of MAD by products

For the sake of comparison, Table 15.4 reports several measures of price dispersion provided by other studies. The only assessment of price dispersion at the barcode level available for France is provided by Dubois and Perrone (2015). The rest of Table 15.4 reports measures of price dispersion computed for the US, the

UK, and Canada (Kaplan and Menzio, 2015; Gorodnichenko et al., 2014; Gorodnichenko and Talavera, 2017). The last column corresponds to our calculation of the same measures of price dispersion as in those studies, but based on our data set.

Table 15.4 Comparison with other measures of price dispersion in the retail sector

Dubois and Perrone (2015):		
France; 1999-2001; 2 barcodes of beer, coffee, cola, whisky		with our data:
Coefficient of variation of prices	0.11	0.08
Interquartile ratio of prices	1.14	1.09
$95^{th}/5^{th}$ centile ratio of prices	1.37	1.26
Kaplan and Menzio (2015):		
US; 2004-2009		
Standard deviation of normalized prices	0.19	0.06
$90^{th}/10^{th}$ centile ratio of normalized prices	1.72	1.13
$90^{th}/50^{th}$ centile ratio of normalized prices	1.26	1.07
$50^{th}/10^{th}$ centile ratio of normalized prices	1.35	1.06
Gorodnichenko et al. (2014):		
US; May2010-Feb2012; goods sold on online shopping platforms		
Standard deviation of log prices	0.24	0.07
Coefficient of variation of prices	0.22	0.07
Interquartile range of log prices	0.35	0.09
Gorodnichenko et al. (2014):		
UK; May2010-Feb2012; goods sold on online shopping platforms		
Standard deviation of log prices	0.23	0.07
Coefficient of variation of prices	0.19	0.07
Interquartile range of log prices	0.31	0.09
Gorodnichenko and Talavera (2017):		
US; Nov2008-Sep2012; goods sold on online price comparators		
Standard deviation of log prices	0.16	0.07
Interquartile range of log prices	0.17	0.09
Gorodnichenko and Talavera (2017):		
Canada; Nov2008-Sep2012; goods sold on online price comparators		
Standard deviation of log prices	0.13	0.07
Interquartile range of log prices	0.11	0.09

Note: Statistics are computed at the barcode level and then averaged across barcodes.

The first conclusion that can be drawn from Table 15.4 is that price dispersion in our data is quite close to that obtained by Dubois and Perrone (2015). On average across products, the 95^{th} percentile of observed prices is 26% higher than the 5^{th} (versus 37% computed by Dubois and Perrone, 2015). The interquartile ratios are even closer to each other (1.09 in our data, versus 1.14). The second conclusion

emerging from comparing statistics of price dispersion computed for the US, the UK and Canada with the same statistics computed on our data is that in France prices appear to be less dispersed. In particular, the comparison with price dispersion across brick-and-mortar stores measured by Kaplan and Menzio (2015) suggests that US prices are at least 20 to 50% more dispersed than in France, depending on the statistic considered.

Price dispersion may differ across many dimensions due to factors related to the product, store or time. Among the dimensions that characterize a product i, there are its brand b, its manufacturer p, and its product category (COICOP-level4) k. A supermarket s can be characterized in terms of its regional branch/wholesaler, denoted wholesaler w as a shortcut, and retail chain r, as well as by its location in market m and region g.

In the product dimension, Fig. 15.3 suggests that price dispersion varies depending on the product's brand and manufacturer. The distribution of the mean absolute deviation is asymmetric, with a long right tail, indicating that products of some brands (and manufacturers) are characterized by deviations from the product's quarterly average log price of an order of more than 10%. The extent of price dispersion also differs across product categories. The upper panel of Fig. 15.4 shows that some product categories at COICOP-level 4 are characterized by a larger price dispersion than others. The lower panel of the same figure suggests that the median relative price is below the mean for all wide product categories (i.e., COICOP level 1).

Regarding the store dimension, the upper panel of Fig. 15.5 shows that price dispersion is quite heterogenous across supermarkets. The lower panel of the figure suggests that the regional branch/wholesaler affiliation of supermarkets accounts for a significant proportion of the observed price dispersion. Looking more specifically at retail chains, each line in Fig. 15.6 represents the boxplot of relative prices for a French retail chain. The average relative prices, as well as their dispersion, largely vary across retail chains.

Another dimension characterizing stores is their location. The long right tail of the upper panel of Fig. 15.7 suggests that some local markets are characterized by rather large average price dispersion, and the lower panel shows that relative prices are more dispersed in some regions, like for instance, in Île-de-France (region number 11, where Paris is), Corse (region number 94), or Provence-Alpes-Côte d'Azur (region number 93), than in others.[14]

Finally, as far as time is concerned, a simple variance analysis of relative prices entailing product, store and time fixed effects suggests that the most important factors relate to stores, followed by factors related to products, while week fixed effects basically do not explain the variance of deviations from the product quarterly average log price.

[14] French regions' code (till 2015) were: 11 Île-de-France, 21 Champagne-Ardenne, 22 Picardie, 23 Haute-Normandie, 24 Centre-Val de Loire, 25 Basse-Normandie, 26 Bourgogne, 31 Nord-Pas-de-Calais, 41 Lorraine, 42 Alsace, 43 Franche-Comté, 52 Pays de la Loire, 53 Bretagne, 54 Poitou-Charentes, 72 Aquitaine, 73 Midi-Pyrénées, 74 Limousin, 82 Rhône-Alpes, 83 Auvergne, 91 Languedoc-Roussillon, 93 Provence-Alpes-Côte d'Azur, 94 Corse.

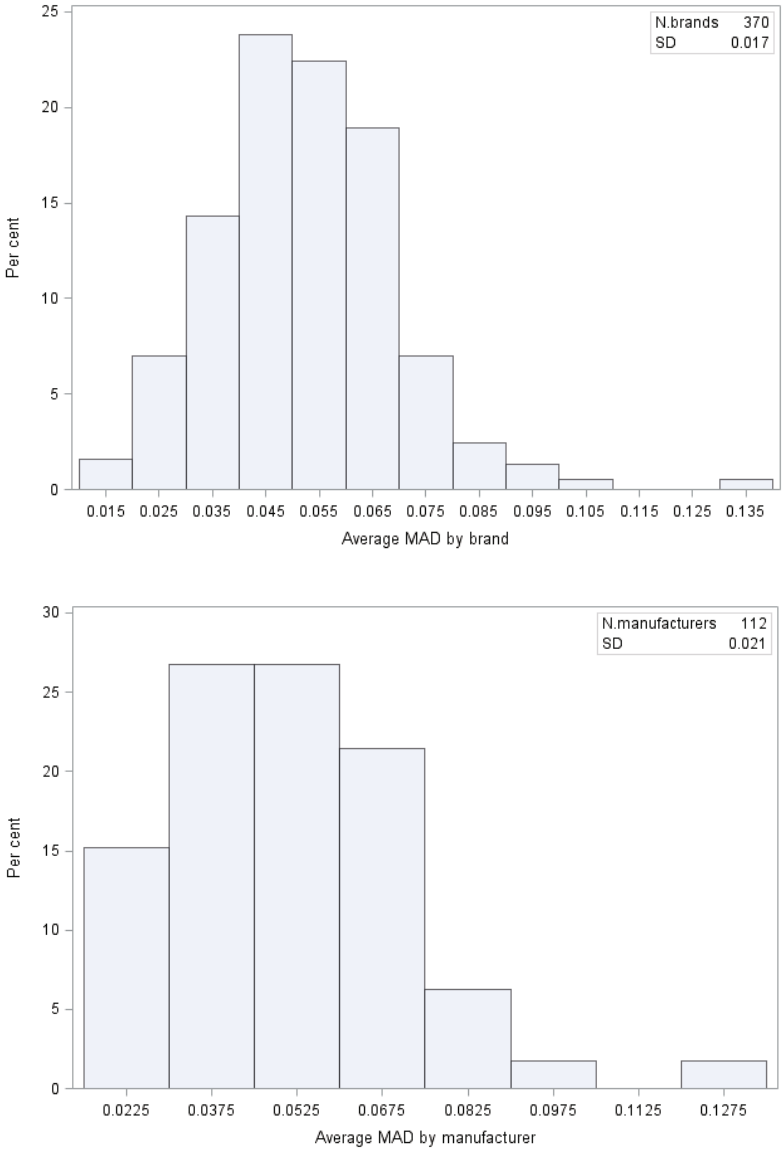

Fig. 15.3 Distribution of MAD by products' brand and manufacturer

15.4 Disentangling the Sources of Price Dispersion

The descriptive analyses presented above may give us some indications about the potential sources of price dispersion. However, a more structured analysis is needed

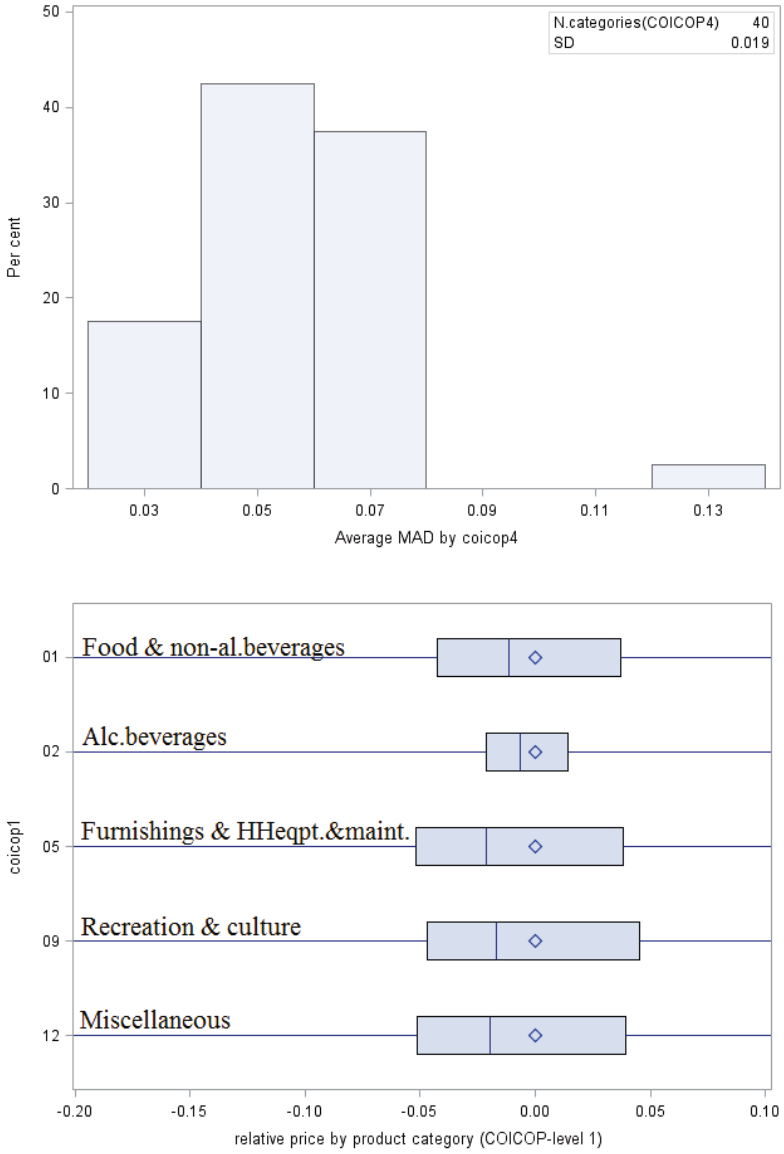

Fig. 15.4 Distribution of MAD by product category COICOP-level 4 and boxplot of relative prices by product category COICOP-level 1

to properly assess their relative importance. In particular, this section provides an answer to a highly debated issue in the theoretical literature on price dispersion: is price dispersion spatial (i.e., some shops persistently sell at lower prices) or rather

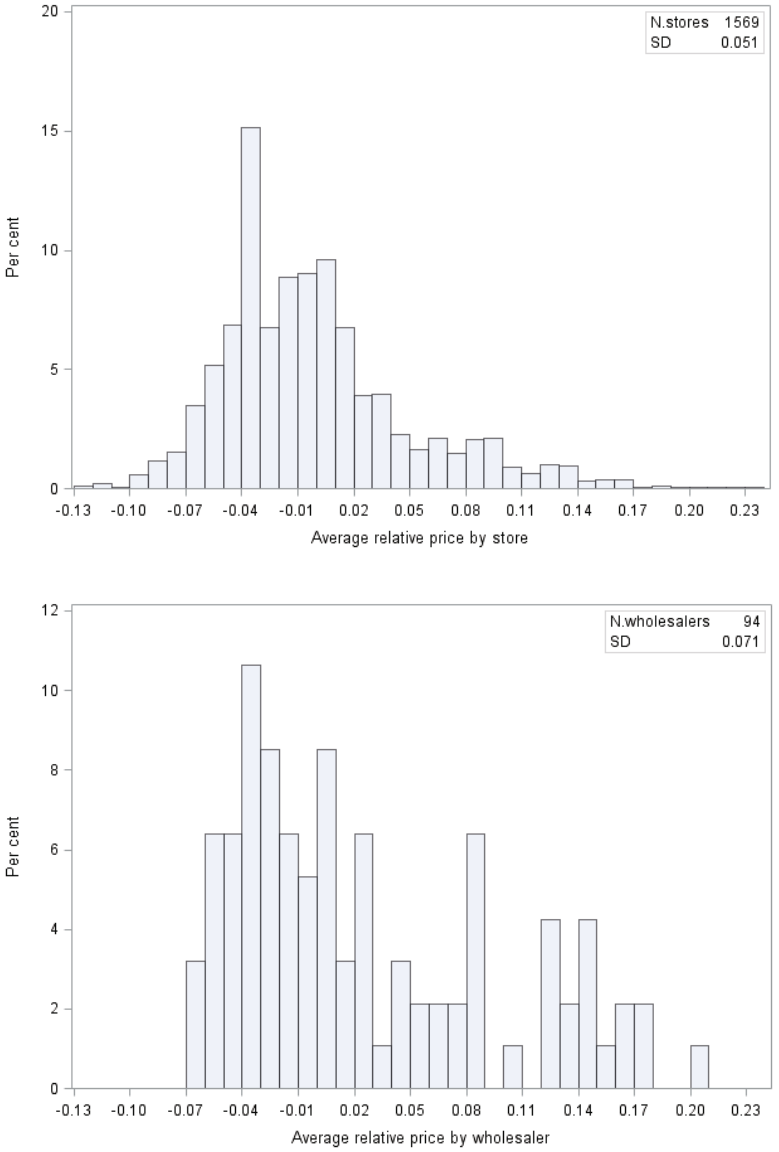

Fig. 15.5 Distribution of average relative prices by stores and wholesalers

transitory (i.e., each supermarket changes its prices over time, so that consumers cannot learn by experience which shops provide the best price)? In order to disentangle the two dimensions of price dispersion, for each product we separately estimate a fixed effect model including supermarket fixed effects, as well as fixed

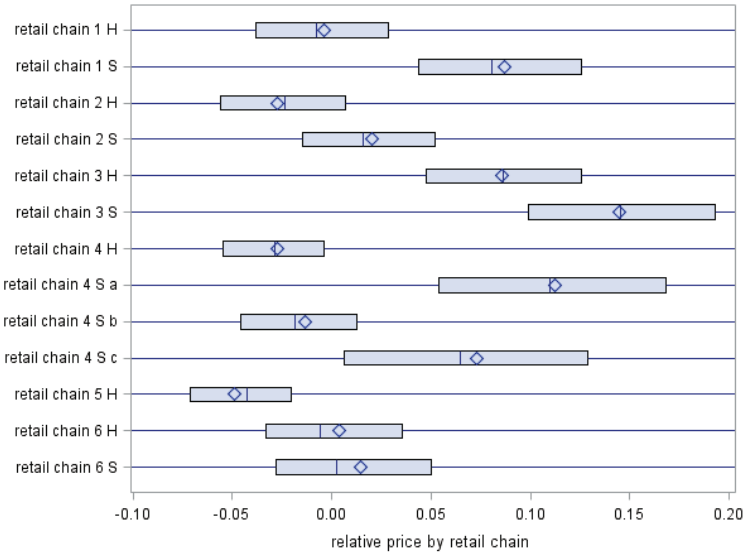

Fig. 15.6 Boxplot of relative prices by retail chains

effects for the combinations of weeks and regional branches. The former aim at capturing all persistent characteristics of supermarkets determining price setting, while the latter account for temporary discounts. Indeed, in France bargaining with producers is typically done by retail groups, while sales and promotions are chain or regional branch-specific, and are thus implemented and advertised at the national or regional level. We then estimate the following model, product by product

$$p_{ist}^{rel(iq)} = \alpha_{is} + \alpha_{iwt} + \varepsilon_{iswt}\,, \tag{15.1}$$

where $p_{ist}^{rel(iq)}$ is the relative price (defined at the beginning of Sect. 15.3), α_{is} are supermarket s fixed effects, α_{iwt} are combinations of week t and regional branch w fixed effects, and ε_{iswt} are the error terms.

Figure 15.8 shows the distribution of the overall variance explained by the above model estimated by product. It suggests that store and temporary discount fixed effects suffice to explain on average 90% of the observed dispersion of prices across stores and time. The model explains 42% of price dispersion in the worst case, and up to more than 98% in the case of the product with the best fit.

Moreover, Fig. 15.9 shows that the estimated variance of supermarket fixed effects α_{is} represents by far the most important component of the total variance explained by the model. On average, the correlation between store fixed effects and relative prices is 0.84, ranging from 0.39 to more than 0.98. As far as temporary discount fixed effects are concerned, on average, the correlation with the observed relative prices is 0.26. This shows that temporary discounts happen on average more

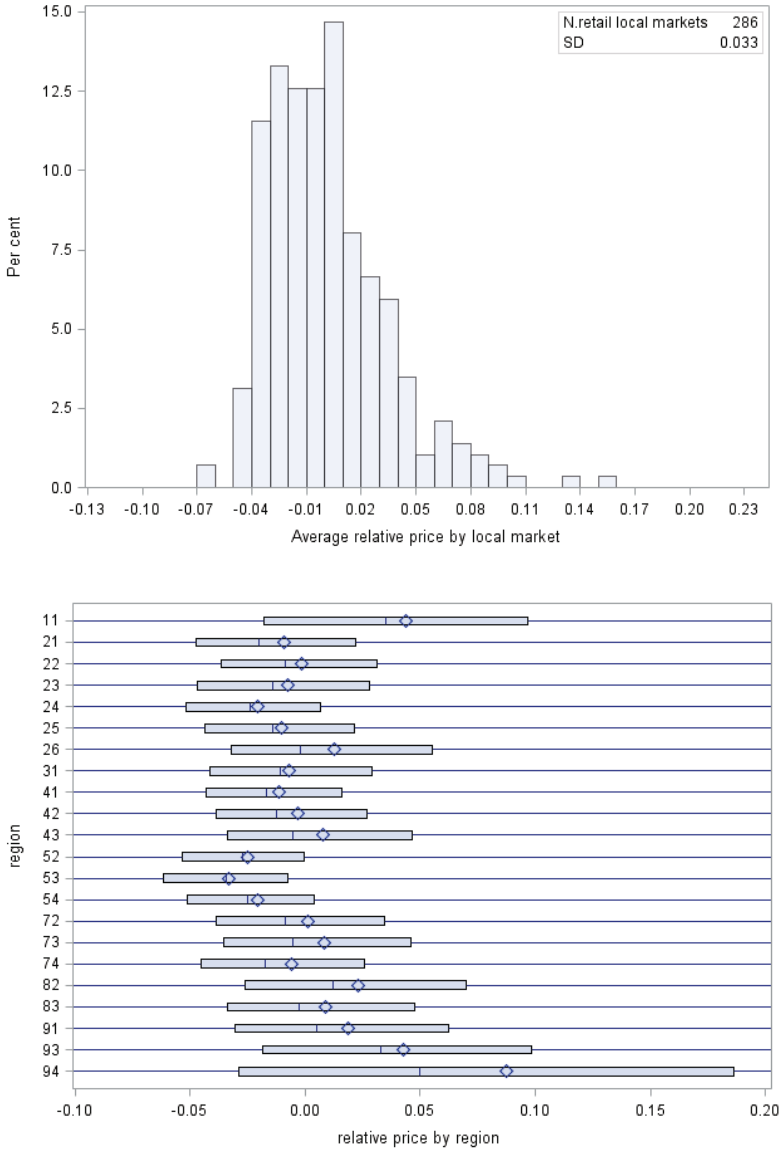

Fig. 15.7 Distribution of relative prices by local markets and boxplot of relative prices by region

frequently for products sold at a relatively high price. However, these correlations range from −0.38 to 0.68, suggesting that discounts may nevertheless also concern products sold at a relatively low price.

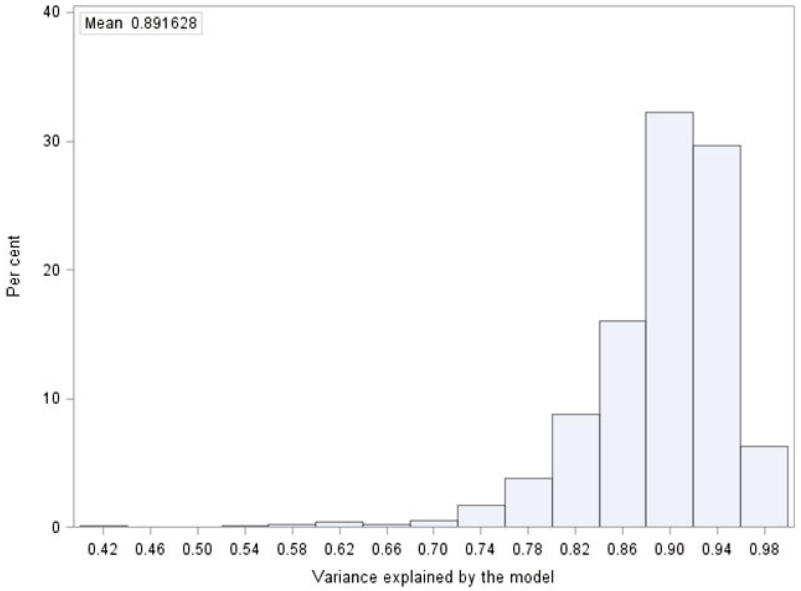

Fig. 15.8 Distribution of the overall variance explained by the model by product

Once reached the conclusion that in France price dispersion is mainly spatial, it is possible to go one step further and estimate a second stage that investigates its determinants. In particular, it is important to understand the relative role played by retail chains at the national level and factors at the local level. It seems likely that the urban density of the area where a supermarket is located may play a role, as well as the characteristics of local demand and competition.[15] We approximate local demand by log per capita income and the population of the district *m*. Competition in the local market is captured by two variables. The first is the number of supermarkets selling product *i* in local market *m*. The second is the distance (in kilometers) to the closest large supermarket, computed exploiting the geo-localization of supermarkets and the exhaustive data base of all medium and large supermarkets in France.[16]

We thus estimate the following models:

1. additive fixed effects model

[15] Urban density categories are based on a combination of population density and absolute population of INSEE "canton-ou-ville" (see Fig. 15.12 in the Appendix for details).

[16] We follow the definition of competitors adopted by the French Competition Authority (2010) and assume that large supermarkets are only in competition with other large ones, while medium size supermarkets and discount stores also are also in competition with large supermarkets. Therefore, not all supermarkets are competing with some medium size competitor, but for all supermarkets we can compute the distance with respect to their closest large size competitor. Distances and driving time to the closest competitor are calculated using two internet applications: GoogleMap and YourNavigation. Since the computed distances are similar, in what follows, we only present measures calculated from YourNavigation.

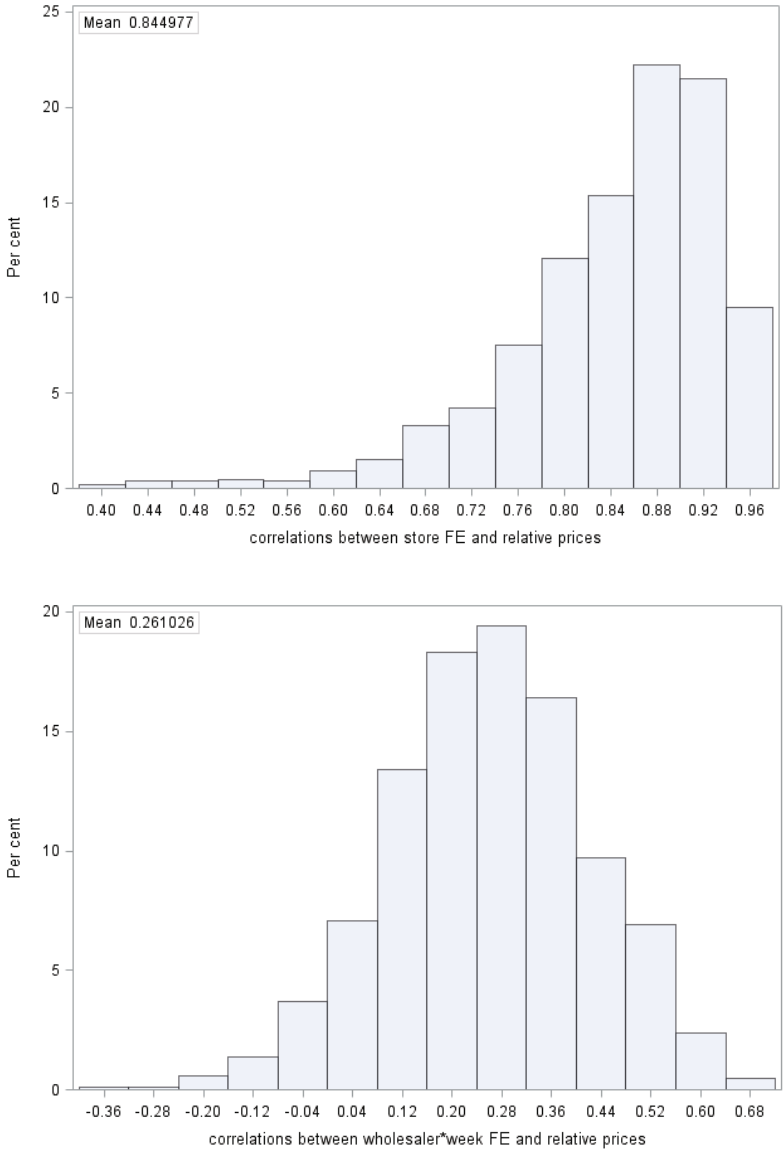

Fig. 15.9 Distribution of correlations between store fixed effects and relative prices by product

$$\alpha_{is} = \delta_i + \delta_r + \gamma_1 \cdot \text{urban} + \gamma_2 \cdot \text{income} + \gamma_3 \cdot \text{pop} + \gamma_4 \cdot \text{sameprod} + \gamma_5 \cdot \text{closest} + \upsilon_{is}\,, \quad (15.2)$$

2. multiplicative fixed effects model

$$\alpha_{is} = \delta_i \cdot \delta_r + \gamma_1 \cdot \text{urban} + \gamma_2 \cdot \text{income} + \gamma_3 \cdot \text{pop} + \gamma_4 \cdot \text{sameprod} + \gamma_5 \cdot \text{closest} + \upsilon_{is}\,, \quad (15.3)$$

where α_{is} are the product×supermarket fixed effects estimated in the first stage, δ_i are product fixed effects, δ_r are fixed effects for (anonymized) retail chains to which the supermarket belongs, *urban* is a 4-level categorical variable for urban density, *income* and *pop* are respectively log per capita income and population of district *m* (which approximates a local market), *sameprod* corresponds to the number of supermarkets selling product *i* in the local market *m* and *closest* is the distance, in kilometers, to the closest large supermarket.

Table 15.5 National and local determinants of spatial price dispersion

	additive FE		multiplicative FE	
Regressor	estimate	SE	estimate	SE
retail chain 1 H	−0.02***	0.0003		
retail chain 1 S	0.04***	0.0004		
retail chain 2 H	−0.04***	0.0002		
retail chain 2 S	0.00***	0.0002		
retail chain 3 H	0.07***	0.0002		
retail chain 3 S	0.10***	0.0004		
retail chain 4 H	−0.03***	0.0003		
retail chain 4 S a	0.09***	0.0008		
retail chain 4 S b	−0.02***	0.0002		
retail chain 4 S c	0.05***	0.0006		
retail chain 5 H	−0.06***	0.0002		
retail chain 6 H	−0.01***	0.0002		
retail chain 6 S (ref)	.	.		
rural	−0.05***	0.0005	−0.05***	0.0004
semi-urban	−0.05***	0.0004	−0.05***	0.0003
urban	−0.04***	0.0004	−0.04***	0.0003
metropolitan (ref)	.	.		
log local per capita income	0.01480***	0.0002	0.01461***	0.0002
log local population	0.00217***	0.0000	0.00222***	0.0000
n.stores selling product	−0.00002***	0.0000	−0.00002***	0.0000
closest large supermkt	0.00030***	0.0000	0.00030***	0.0000
product FE	yes		no	
product×retail chain FE	no		yes	
R squared	0.34		0.58	

Notes: *** means significant at 1%.

Table 15.5 reports the estimated coefficients of the regressions (15.2) and (15.3). The estimation results show that the multiplicative fixed effect model does a significantly better job in fitting the data than the additive fixed effect model. The *R* squared is almost doubled when considering multiplicative fixed effects. The comparison of the regressors' contributions to the variance explained by the model re-

veals that retail chains are by far the most relevant factor. Conditionally on all other regressors, retail chains account for 76% of the variance explained by the model in the additive fixed effects model. The second most important factor is represented by product fixed effects, which account for 8% of the variance explained by the model. In the multiplicative effects model, retail chain and product fixed effects represent 97% of the explained variance, suggesting that not all products exhibit the same price dispersion across retail chains. Not surprisingly, retail chain fixed effects show that prices in large supermarkets (denoted by the inclusion in the anonymized retail chain name in Table 15.5 of an "H" for "Hypermarche" in French) tend to be lower than those in medium-size supermarkets (denoted by an "S").[17] Regarding the impact of local factors on prices, it is worth noticing that the estimates in the additive and multiplicative effects model are very similar. First, urban density increases price levels in supermarkets. Second, supermarkets facing favorable local demand conditions in terms of a larger population and per capita income also exhibit higher prices. Finally, stronger local competition tends to decrease price levels in supermarkets. Indeed, the further away the nearest large supermarket is located, the higher the prices. At the same time, if a product sold by a supermarket s is also available in several other supermarkets located in the same local market, price levels decrease. However, despite their statistical significance, these local factors have a rather limited quantitative impact on prices.

The fact that in France centralized price setting strategies dominate local factors in retail prices is the main message that can be drawn from our analysis of price dispersion. This finding is robust to alternative definitions of the level of centralization in price setting determination. If we include retail group (instead of chain) dummies in model (15.2), the R squared decreases only to 0.30 and, conditionally on all other regressors, retail groups still account for 72% of the variance explained by the model. These results suggest that a large chunk of price setting is actually already determined at this level. Similarly, if we include regional branch/wholesaler dummies in model (15.2), the R squared increases only to 0.37 and, conditionally on all other regressors, regional branches account for 77% of the variance explained by the model. Therefore, there seems to be an additional stage of price setting happening at the regional level, but national price setting strategies are dominant.

15.5 Conclusion

Based on a large and original data set containing almost 40 million weekly price records from more than 1500 medium and large size supermarkets in France from October 2011 to September 2012, we characterize the overall shape and structure of price dispersion in the French retail sector. We show that temporary sales and

[17] The only exception is retail group 3, which has higher prices than some medium-sized supermarket retail chains. Note, however, that prices in its large supermarkets tend anyway to be lower than in its own smaller supermarkets.

promotions offer a limited explanation of the observed price dispersion, while the permanent component of price dispersion largely dominates.

In fact, in France price dispersion across stores is essentially the result of persistent heterogeneity in retail chains' national pricing. First, retail groups bargain with producers and the market share of the corresponding retail groups is likely to affect their bargaining power. Second, retail groups set national prices at the retail chain level (i.e., retail groups owning more than one retail chain set different prices across them) or at the regional branch level. Within a retail group, for instance, prices are lower in chains characterized by larger stores. More generally, the average level of prices depends on the positioning and the customers' target of the retail chain. Despite this rather centralized price-setting behavior, we show that local conditions regarding demand or local competition between supermarkets also contribute to explaining the observed prices in local markets, though to a much lower extent.

Acknowledgements

The authors would like to thank Prixing, the start-up that provided the price data used in this chapter and in particular Eric Larchevêque, the owner and CEO of the company at the time data were released. We also thank the Aix-Marseille School of Economics and the Banque de France for jointly funding the LSA database. We are extremely grateful to Etienne Gagnon for his very helpful comments. We wish to express our special thanks to Alexandre Vigneron, Marine Tépaut, Sherly Jean-Charles, Sophie Saleh, Sylvie Tarrieu, and Yue Zhu for their precious assistance with complementary data collection and treatment. Finally, we thank Vincent Guegan and Victor Yammouni for life-saving IT assistance. The views expressed in this chapter are those of the authors and do not necessarily represent those of their institutions.

References

Alvarez, F., Golzález-Rodaza, M., Neumeyer, A., and Beraja, M. (2013). From hyperinflation to stable prices: Argentina's evidence on menu cost models. Working papers, Stanford Center for International Development.

Anderton, R., Meyler, A., Gattini, L., Izquierdo, M., Jarvis, V., Kaarup, R., Komzakova, M., Landau, B., Mohr, M., Page, A., Sondermann, D., Vermeulen, P., Cornille, D., Tsalinski, T., Vladova, Z., Hartmann, C., Stahl, H., Linehan, S., Balfoussia, C., Panagiotou, S., de los Llanos Matea, M., Alvarez, L., Bardet-Fremann, P., Berardi, N., Sevestre, P., Ciapanna, E., Rondinelli, C., Kapatais, D., Walch, E., Lunnemann, P., Zerafa, S., Pace, C., Kieft, J., Fritzer, F., Cardoso, F., Gabrijelcic, M., Karma, B., and Kivisto, J. (2011). Structural features of distributive trades and their impact on prices in the Euro area. European Central Bank Occasional Paper Series 128.

Asplund, M. and Friberg, R. (2002). Food prices and market structure in Sweden. *Scandinavean Journal of Economics*, 104(4):547–566.

Borenstein, S. and Rose, N. L. (1994). Competition and price dispersion in the U.S. airline industry. *Journal of Political Economy*, 102(4):653–83.

Cheung, Y.-W. and Fujii, E. (2008). Deviations from the law of one price in Japan. CESifo Working Paper Series 2275, CESifo Group Munich.

Dauvers, O. (2013a). Drive insights.

Dauvers, O. (2013b). Drive insights.

Dixon, H. D., Luintel, K. B., and Tian, K. (2014). The impact of the 2008 crisis on UK prices: what we can learn from the CPI microdata. Cardiff Economics Working Papers E2014/7, Cardiff University, Cardiff Business School, Economics Section.

Dubois, P. and Perrone, H. (2015). Price dispersion and informational frictions: Evidence from supermarket purchases. CEPR Discussion Papers 10906, C.E.P.R. Discussion Papers.

Eyal, B. and Eden, B. (2004). Price rigidity and price dispersion: Evidence from micro data. *Review of Economic Dynamics*, 7(3):613–641.

French Competition Authority (2010). La concentration des zones de chalandise.

Gerardi, K. S. and Shapiro, A. H. (2009). Does competition reduce price dispersion? New evidence from the airline industry. *Journal of Political Economy*, 117(1):1–37.

Gorodnichenko, Y., Sheremirov, V., and Talavera, O. (2014). Price setting in online markets: Does IT click? NBER Working Papers 20819, National Bureau of Economic Research, Inc.

Gorodnichenko, Y. and Talavera, O. (2017). Price setting in online markets: Basic facts, international comparisons, and cross-border integration. *American Economic Review*, 107(1):249–82.

Hong, H. and Shum, M. (2006). Using price distributions to estimate search costs. *The RAND Journal of Economics*, 37:257–275.

Kaplan, G. and Menzio, G. (2015). The morphology of price dispersion. *International Economic Review*, 56(4):1165–206.

Moen, E., Wulfsberg, F., and Aas, O. (2014). Dispersion in retail prices. Working paper.

Moraga-González, J. L. and Wildenbeest, M. R. (2008). Maximum likelihood estimation of search costs. *European Economic Review*, 52(5):820–848.

Nakamura, E., Steinsson, J., Sun, P., and Villar, D. (2016). The elusive costs of inflation: Price dispersion during the U.S. great inflation. NBER Working Papers 22505, National Bureau of Economic Research, Inc.

Sorensen, A. T. (2000). Equilibrium price dispersion in retail markets for prescription drugs. *Journal of Political Economy*, 108(4):833–850.

Woodward, S. E. and Hall, R. E. (2012). Diagnosing consumer confusion and suboptimal shopping effort: Theory and mortgage-market evidence. *American Economic Review*, 102(7):3249–3276.

Appendix

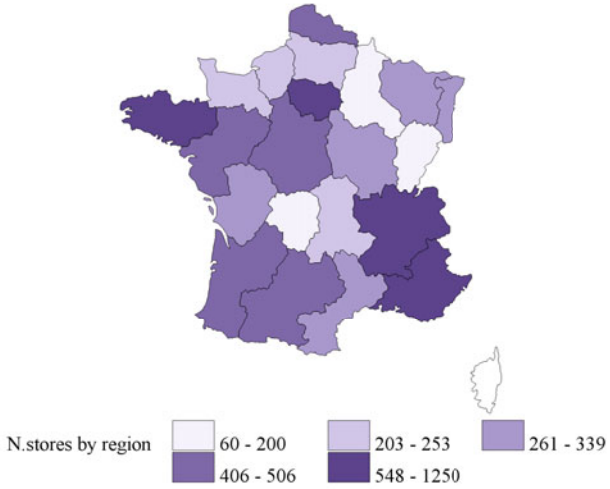

Fig. 15.10 Geographical distribution by region of stores. Note: Darker colored regions correspond to larger number of stores

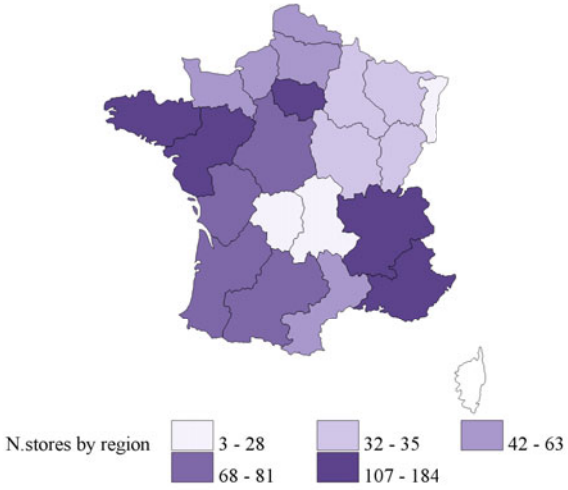

Fig. 15.11 Geographical distribution by department of stores for which price data are available. Note: Darker colored regions correspond to larger number of stores

	Population in 'canton-ou-ville'			
Pop. density	0 to 10,000	10,000 to 50,000	50,000 to 300,000	300,000 +
0 to 200	**0**	**1**	**2**	**3**
200 to 500	**1**			
500 to 3,000		**2**		
3,000 to 15,000				
15,000 +			**3**	

Fig. 15.12 Criteria defining urban density

Index

L. Matyas (ed.), *The Econometrics of Multi-dimensional Panels*, Advanced Studies in Theoretical and Applied Econometrics 50, DOI 10.1007/978-3-319-60783-2

Z