Chapter 1
Fixed Effects Models

László Balázsi, László Mátyás, 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 some

László Balázsi
Morgan Stanley, Budapest, Hungary, e-mail: Official

László Mátyás ✉
Central European University, Budapest, Hungary and Vienna, Austria, e-mail: matyas@ceu.edu

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

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, Mayer and Ries (2010)	Colonial Trade Linkages		618 000	γ_{ij}	No
S. Baier and Bergstrand (2002)	Endogeneity of Trade Flows		1 400	$\alpha_i + \gamma_j$	No
S. L. 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, Larch, E. and Winkelmann (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, Mayer and Thoenig (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, Egger and Pfaffermayr (2003)	Gravity Model Spec.		10 000	$\gamma_{ij} + \alpha_{it} + \alpha^*_{jt}$	No
S. L. 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, Larch and Yotov (2015)	Border Effects		24 000	$\gamma_{ij} + \alpha_{it} + \alpha^*_{jt}$	No
Chapter 12 – Housing and Prices					
Fu, Zhu and Ren (2015)	Housing Tenure Choices	household - prefecture - type	2 500 000	λ_t	No
Syed, Hill and Melser (2008)	House Prices Indices	house - region - quarter	418 000	α_{jt}	No
Gayer, Hamilton and Viscusi (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, McMillan, Murphy and Timmins (2016)	Demand for Houses	household - neighbourhood - time	1 000 000	$\alpha_i + \alpha^*_{jt}$	No
Baltagi, Bresson and Etienne (2015)	Neighbor's Prices	year - 'arrondissement' - quartier - block - flat	157 000	$\alpha_{ta} + \gamma_{taq} + \lambda_{taqi}$	No

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

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 replacing 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, Matyas and Wansbeek (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

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

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, Brücker and Fernández-Huertas Moraga (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, Alimi, Cameron and Maré (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, Lima and Orefice (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, Sheremirov and Talavera (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, Boutin and Verge (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

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, Alba and Park (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, Milet and Mirza (2016)	Domestic Trade Regulations	origin country - destination country - firm - time	115 000	$\alpha_s + \lambda_t$; α_{st}	No
Bussiere, Fidrmuc and Schnatz (2005)	Trade Integration	origin country - destination country - year	50 000	$\gamma_{ij} + \lambda_t$	No
Bellak, Leibrecht and Riedl (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, Konya and Matyas (2000)	Environmental Regulations		3 800	$\alpha_i + \gamma_j + \lambda_t$	No
Heyman, Sjoholm and Tingvall (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, Burgess, Redding and Zilibotti (2008)	Indian Trade Liberalization	industry - state - year	18 000	$\gamma_{ij} + \alpha_{it} + \alpha^*_{jt}$	Yes

the fastest,[2] such as

$$(y_{111}, \ldots, y_{11T}, \ldots, y_{1N_2 1}, \ldots, y_{1N_2 T}, \ldots, y_{N_1 11}, \ldots, y_{N_1 1T}, \ldots, y_{N_1 N_2 1}, \ldots, y_{N_1 N_2 T})'.$$

We assume in general that the index sets $i \in \{1, \ldots, N_1\}$ and $j \in \{1, \ldots, N_2\}$ are (completely or partially) different. When dealing with economic flows, such as trade, capital, investment (FDI), etc., there is some kind of reciprocity, in such cases it is assumed that $N_1 = N_2 = N$. The main question is how to formalize the individual and time heterogeneity – in our case, the fixed effects. In standard 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

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

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 $(N_1 N_2 T \times 1)$ and $(N_1 N_2 T \times K)$, β being the vector of the slope parameters of size $(K \times 1)$, π the composite fixed effects parameters, D the matrix of dummy variables, and finally, ε the vector of the disturbance terms.

The first attempt to properly extend the standard fixed effects panel data model to a multi-dimensional setup was proposed by Matyas (1997) (see for more, for example, Baltagi, 2013 and Balestra & 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 & van Wincoop, 2001; Magee, 2003; Parsley, 2003; Heyman et al., 2007; Bellak et al., 2008; Fourie & Santana-Gallego, 2011; Ortega & Peri, 2013; Adserà & Pytliková, 2015; and Turnbull & 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 & Rose, 2002; Smith & Yetman, 2007; Head et al., 2010; Hur et al., 2010; and 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 & Klaassen, 2002, Eilat & Einav, 2004, Bussiere et al., 2005, Romalis, 2007, Shin & Serlenga, 2007; Martin et al., 2008; Bertoli et al., 2016 or Bertoli & Fernández-Huertas Moraga, 2013; Beine & 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 S. L. 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 & Schnier, 2010; and (Crozet et al., 2016)). It is reasonable to present the symmetric version of this model (with α_{it} fixed effects); however, as it has exactly the same properties, we consider the two models together.[3]

A variation of this model is

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

(Olivero & Yotov, 2012 and S. L. Baier & 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; S. L. Baier & Bergstrand, 2007; Aghion et al., 2008; Melitz & Toubal, 2012; Nuroglu & Kunst, 2014; and 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)	$\big((I_{N_1} \otimes \iota_{N_2T}),\ (\iota_{N_1} \otimes I_{N_2} \otimes \iota_T),\ (\iota_{N_1N_2} \otimes I_T)\big)$
(1.3)	$\big(I_{N_1N_2} \otimes \iota_T\big)$
(1.4)	$\big((I_{N_1N_2} \otimes \iota_T),\ (\iota_{N_1N_2} \otimes I_T)\big)$
(1.5)	$\big(I_{N_1} \otimes \iota_{N_2} \otimes I_T\big)$
(1.6)	$\big((I_{N_1} \otimes \iota_{N_2} \otimes I_T),\ (\iota_{N_1} \otimes I_{N_2T})\big)$
(1.7)	$\big((I_{N_1N_2} \otimes \iota_T),\ (I_{N_1} \otimes \iota_{N_2} \otimes I_T),\ (\iota_{N_1} \otimes I_{N_2T})\big)$

[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.

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)

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 $(K + N_1 N_2)$, 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

$$\hat{\beta} = (X' M_D X)^{-1} X' M_D y \tag{1.8}$$
$$\hat{\pi} = (D'D)^{-1} D'(y - X\hat{\beta}) , \tag{1.9}$$

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

[4] Since $\iota_{N_1 N_2 T}$ is spanned by all given specifications for D, there is intercept in X.

$$D'D\hat{\pi} + D'X\hat{\beta} = D'y \tag{1.10}$$
$$X'D\hat{\pi} + X'X\hat{\beta} = X'y\,. \tag{1.11}$$

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 $\hat{\beta}$.

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

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

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

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

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

with the appropriate M_D, and

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

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

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

where

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

is the transformed residual square, and $(\mathsf{rank}(M_D) - K)$ 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, $N_1 \mathsf{var}(\hat{\beta})$ is finite in the limit, as

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

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

$$\hat{\beta}_k / \sqrt{\widehat{\mathsf{var}}(\hat{\beta}_k)}$$

where $\widehat{\mathsf{var}}(\hat{\beta}_k)$ is the k-th diagonal element of $\widehat{\mathsf{var}}(\hat{\beta})$. 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 R_U^2 is the R^2 of the *unrestricted model* (that is the full model (1.2)), while R_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 $(N_1 N_2 T - N_1 - N_2 - T + 1 - K)$. This statistic then has an F-distribution with $(N_1 - 1, N_1 N_2 T - N_1 - N_2 - T + 1 - K)$ degrees of freedom.

1.4 Incomplete Panels

As in the case of the usual 2D panel data sets (see Wansbeek & 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 $t \in \{1, \dots, T\}$, 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 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 & 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, Guimaraes and Portugal (2012). Let us show the procedure on model (1.2), the rest is a direct analogy. Model (1.2) in matrix form reads as

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

[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.

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 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

[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.

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

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 variables, (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:

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

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

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

and

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

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

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

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:

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

and finally

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

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 $(t = 1 \ldots T)$ 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,

$$
\begin{aligned}
D_1^b &= \left(W_1', W_2', \ldots, W_T'\right)' \quad \text{of size} \quad (R \times N_1 N_2) \,, \\
D_2^b &= D_3^a \quad \text{of size} \quad (R \times T) \,.
\end{aligned}
$$

The first step in the projection is now

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

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

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

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,

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

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

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

leading to

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

Note that for M_{D_c}, we have to invert an order $\min\{N_1 T, N_2 T\}$ 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:

$$
\begin{aligned}
D_1^d &= D_1^b \quad \text{of size} \quad (R \times N_1 N_2)\,, \\
D_2^d &= D_1^c \quad \text{of size} \quad (R \times N_1 T)\,, \\
D_3^d &= D_2^c \quad \text{of size} \quad (R \times N_2 T)\,.
\end{aligned}
$$

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 $N_1 T$, and $N_2 T$) 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

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

where

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

---

[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)	$\max\{N_1T, N_2T\}$

Now the scalar representation of transformation (1.29) is

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

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

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

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

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

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 $\mathsf{min}\{N_1,N_2\}$ 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

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

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

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

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

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

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:

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

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

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

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

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

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

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

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

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

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 $\mathsf{var}(\hat{\beta})$. Then we get

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

where "∼" 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

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

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

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

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 $i = 1\dots N_1$, $j = 1\dots N_2$, $s = 1\dots N_3$, and $t = 1\dots T$. This is what we consider from now on as 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

$$D = \big((I_{N_1N_2N_3} \otimes \iota_T),\ (I_{N_1N_2} \otimes \iota_{N_3} \otimes I_T),\ (\iota_{N_1} \otimes I_{N_1N_3T}),\ (I_{N_1} \otimes \iota_{N_2} \otimes I_{N_3T})\big)$$

with size $(N_1N_2N_3T \times (N_1N_2N_3 + N_1N_2T + N_2N_3T + N_1N_3T))$ and column rank $(N_1N_2N_3T - (N_1-1)(N_2-1)(N_3-1)(T-1))$, leading to

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

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

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

(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

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

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

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

[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.

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:

$$\sum_{ij} \gamma_{ij} = 0 \, ; \quad \sum_t \lambda_t = 0 \, . \tag{1.42}$$

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

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

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

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

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

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

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

[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.

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

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

$$\begin{aligned} \hat{\beta} &= \tfrac{1}{N_1N_2}\sum_{ij}\gamma_{ij}^0 + \frac{1}{T}\sum_t \lambda_t^0 \\ \hat{\gamma}_{ij} &= \gamma_{ij}^0 - \tfrac{1}{N_1N_2}\sum_{ij}\gamma_{ij}^0 \qquad (i, j = 1 \ldots N_1, N_2) \\ \hat{\lambda}_t &= \lambda_t^0 - \tfrac{1}{T}\sum_t \lambda_t^0 \qquad (t = 1 \ldots T) \end{aligned} \tag{1.47}$$

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

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

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

Adserà, A. & 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. & 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. & Bergstrand, J. (2002). *On the endogeneity of international trade flows and free trade agreements.* Unpublished Manuscript. University of Notre Dame.

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

Baier, S. L. & 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. & Wansbeek, T. (2015). The estimation of multi-dimensional fixed effects panel data models. *Econometric Reviews*, *forthcoming, published online: April 2015*.

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

Balestra, P. & Krishnakumar, J. (2008). Fixed effects and fixed coefficients models. In Matyas & Sevestre (Eds.), *The econometrics of panel data, 3rd edition* (pp. 23–48). Springer Verlag.

Baltagi, B. H. (2013). *Econometric analysis of panel data* (5th ed.). John Wiley & Sons.

Baltagi, B. H., Bresson, G. & 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. & Pfaffermayr, M. (2003). A generalized design for bilateral trade flow models. *Economics Letters*, *80*, 391–397.

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

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

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

Bellak, C., Leibrecht, M. & 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. & Ederveen, S. (2012). Cultural barriers in migration between oecd countries. *Journal of Population Economics*, *25*(3), 1077–1105.

Bergstrand, J. H., Larch, M. & 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. & Fernández-Huertas Moraga, J. (2016). The european crisis and migration to Germany. *Regional Science and Urban Economics*, *60*, 61–72.

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

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

Borenstein, S. & Rose, N. L. (1994). Competition and price dispersion in the U.S. airline industry. *Journal of Political Economy*, *102*(4), 653–683.

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

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

Echevarria, J. & 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. & Winkelmann, R. (2011). The trade effects of endogenous preferential trade agreements. *American Economic Journal: Economic Policy*, *3*(3), 113–43.

Egger, P. & 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. & Pfaffermayr, M. (2011). *Structural estimation of gravity models with path-dependent market entry* (Tech. Rep.). C.E.P.R. Discussion Papers.

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

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

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

Fu, Q., Zhu, Y. & 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. & 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. & 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. & Rose, A. (2002). Does a currency union affect trade? The time series evidence. *European Economic Review*, *46*, 1125–1151.

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

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

Haller, L. E. & 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. & 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. & Ries, J. (2010). The erosion of colonial trade linkages after independence. *Journal of International Economics*, *81*(1), 1–14.

Heyman, F., Sjoholm, F. & 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. & 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* (3rd ed.). Econometric Society Monographs, Cambridge University Press.

Hur, J., Alba, J. D. & 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. & Hornok, C. (2017). Winners and losers of globalisation: Sixteen challenges for measurement and theory. In R. Blundell et al. (Eds.), *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. & 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. & Toubal, F. (2012). Native language, spoken language, translation and trade. *Journal of International Economics*, *93*(2).

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

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

Perkins, R. & 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. & 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. & 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. & 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. & Yetman, J. (2007). The curse of Irving Fisher (professional forecasters' version). *Queen's Economics Department Working Paper No. 1144*.

Syed, I., Hill, R. J. & 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. & 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. & van der Vlist, A. J. (2015). Uninformed house buyers and foreclosures. *University of Central Florida, Department of Finance, Working Paper 1503*.

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