Central European University
Department of Economics

Work in progress, please do not quote!

Modelling Multi-dimensional Panel Data: A Random Effects Approach

Laszlo Balazsi
Central European University

Laszlo Matyas[1]
Central European University

Daria Pus
University of Texas at Austin

May 2015

Abstract: This paper presents the empirically most relevant three-way error components models to be used in three-dimensional panel data sets. Appropriate Feasible GLS estimators are derived for complete data, as well as incomplete, and unbalanced ones. Special attention is paid to the relevant asymptotic results, as these these panels are usually very large. Some extensions are also considered, in particular, the estimation of mixed fixed effects–random effects models, to be used when the data is of limited size in one or several dimensions. It is also shown how to extend the main results to four-dimensions and beyond. The results are illustrated empirically on a World trade data set, highlighting some important characteristics of the Random effect vs. Fixed effects estimation.

Key words: panel data, multidimensional panel data, incomplete data, random effects, mixed effects, error components model, trade model, gravity model.

JEL classification: C1, C2, C4, F17, F47.

Acknowledgement: Comments by Tom Wansbeek were highly appreciated.

[1] Corresponding author; Central European University, Department of Economics, Budapest 1052, Nador u. 9, Hungary; matyas@ceu.hu.

1. Introduction

The last few years new large data sets have been emerging in economics and finance almost on a weekly basis. Often these take the form of multi-dimensional panels, like, for example, in the case of salary data of firms' employees over time (three dimensions), which can then be linked together between regions or countries (four dimensions), or trade data between countries or regions observed at sector level. Several model specifications have been proposed in the literature to deal with the heterogeneity of this type of data, but most of them treated these as fixed effects, i.e., fixed unknown, but observable, parameters (see, for example, *Baltagi et al.* [2003], *Egger and Pfaffermayr* [2003], *Baldwin and Taglioni* [2006], and *Baier and Bergstrand* [2007]). As it is pretty well understood from the theory and practice of the "usual" two-dimensional (2D) panel data econometrics, fixed effects formulations are more suited to deal with cases when panels are, at least in one dimension, short. On the other hand, for large data sets, when observations can be considered as sample(s) from an underlying population, random effects specifications seems to be more suited, where the specific effects are considered as random.

More than two decade ago *Moulton* [1990] draw the attention to the fact that when dealing with disaggregated observations to answer macro type questions the covariance structure of the model is of paramount importance. When no repeated observations are available heavy assumptions are needed to get treatable models (see, for example, *Gelman* [2006], amongst others). On the other hand, when working with panel data sets we can take advantage of the repeated observations, and generalize them to higher dimensional setups, in fact deriving multi-dimensional random effects panel data models.

Historically multi-dimensional random effects (or error components) models can be traced back to the variance component analysis literature (see *Rao and Kleffe* [1980], or the seminal results of *Laird and Ware* [1982] or *Leeuw and Kreft* [1986]) and can be considered special cases of multi-level models, well known in statistics (see, for example, *Scott et. al.* [2013], *Luke* [2004], *Goldstein* [1995], and *Bryk and Raudenbush* [1992]). We, however, in most of the paper, assume fixed slopes for the regressors (rather than a composition of fixed and random elements), and zero means for the random components. This allows for special covariance structures simplifying estimation, but at the same time, we have to deal with some issues mostly specific to economic data.

In this paper we introduce several types of three-dimensional random effects model specifications and their extensions to higher dimensions. We derive proper estimation methods for each of them and analyze their properties under some data problems. The general form of these random effects formulations can be casted as

$$y = X\beta + u\,, \tag{1}$$

where y and X are respectively the vector and matrix of observations of the dependent and explanatory variables, β is the vector of unknown parameters, and we want to exploit the structure embedded in the random disturbance terms u. As it is well known from the Gauss-Markov theorem, the General Least Squares (GLS) estimator is BLUE for β. To make it operational, in principle, we have to perform three steps. First, using the specific structure of u, we have to derive the variance-covariance matrix of model (1), $E(uu') = \Omega$, then, preferably using spectral decomposition, we have to calculate its inverse. This is important, as multi-dimensional data often tend to be very large, leading to some Ω-s of extreme order. And finally, we need to estimate the unknown variance components of Ω to arrive to the well known Feasible GLS (FGLS) formulation.

2. Different Model Specifications

In this section we present the most relevant three-dimensional model formulations, paying special attention to the different interaction effects.

2.1. Various Heterogeneity Formulations

The most general model we can think of in the three-dimensional setup encompassing all pairwise random effects is

$$y_{ijt} = x'_{ijt}\beta + \mu_{ij} + \upsilon_{it} + \zeta_{jt} + \varepsilon_{ijt}\,, \tag{2}$$

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

$$E(\mu_{ij}\mu_{i'j'}) = \begin{cases} \sigma_\mu^2 & i = i' \text{ and } j = j' \\ 0 & \text{otherwise} \end{cases}$$

$$E(\upsilon_{it}\upsilon_{i't'}) = \begin{cases} \sigma_\upsilon^2 & i = i' \text{ and } t = t' \\ 0 & \text{otherwise} \end{cases}$$

$$E(\zeta_{jt}\zeta_{j't'}) = \begin{cases} \sigma_\zeta^2 & j = j' \text{ and } t = t' \\ 0 & \text{otherwise} \end{cases}$$

The covariance matrix of such error components structure is simply

$$\Omega = E(uu') = \sigma^2_\mu (I_{N_iN_j} \otimes J_T) + \sigma^2_\upsilon (I_{N_i} \otimes J_{N_j} \otimes I_T) + \sigma^2_\zeta (J_{N_i} \otimes I_{N_jT}) + \sigma^2_\varepsilon I_{N_iN_jT} \ , \quad (3)$$

where I_{N_i} and J_{N_i} are the identity, and the square matrix of ones respectively, with the size in the index.

All other relevant model specifications, are obtained by applying some restrictions on the random effects structure, that is all covariance structures are nested into that of model (2). The model which only uses individual-time-varying effects reads as

$$y_{ijt} = x'_{ijt}\beta + \upsilon_{it} + \zeta_{jt} + \varepsilon_{ijt} \ , \quad (4)$$

together with the appropriate assumptions enlisted for model (2). Now

$$\Omega = \sigma^2_\upsilon (I_{N_i} \otimes J_{N_j} \otimes I_T) + \sigma^2_\zeta (J_{N_i} \otimes I_{N_j} \otimes I_T) + \sigma^2_\varepsilon I_{N_iN_jT} \ . \quad (5)$$

A further restriction on the above model is

$$y_{ijt} = x'_{ijt}\beta + \zeta_{jt} + \varepsilon_{ijt} \ , \quad (6)$$

which in fact is a generalization of the approach used in multi-level modeling, see for example, *Ebbes, Bockenholt and Wedel* [2004] or *Hubler* [2006].[2] The covariance matrix now is

$$\Omega = \sigma^2_\zeta (J_{N_i} \otimes I_{N_jT}) + \sigma^2_\varepsilon I_{N_iN_jT} \ . \quad (7)$$

An other restrictions of model (2) is to leave in the pair-wise random effects, and restrict the individual-time-varying terms. Specifically, model

$$y_{ijt} = x'_{ijt}\beta + \mu_{ij} + \lambda_t + \varepsilon_{ijt} \quad (8)$$

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

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

Now

$$\Omega = \sigma^2_\mu (I_{N_iN_j} \otimes J_T) + \sigma^2_\lambda (J_{N_iN_j} \otimes I_T) + \sigma^2_\varepsilon I_{N_iN_jT} \ . \quad (9)$$

[2] The symmetric counterpart of model (6), with υ_{it} random effects, could also be listed here, however, as it has the exact same properties as model (6), we take the two models together.

A restriction of the above model, when we assume, that $\mu_{ij} = \upsilon_i + \zeta_j$ is[3]

$$y_{ijt} = x'_{ijt}\beta + \upsilon_i + \zeta_j + \lambda_t + \varepsilon_{ijt} \tag{10}$$

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

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

Its covariance structure is

$$\Omega = \sigma_\upsilon^2 (I_{N_i} \otimes J_{N_j T}) + \sigma_\zeta^2 (J_{N_i} \otimes I_{N_j} \otimes J_T) + \sigma_\lambda^2 (J_{N_i N_j} \otimes I_T) + \sigma_\varepsilon^2 I_{N_i N_j T} \,. \tag{11}$$

Lastly, the simplest model is

$$y_{ijt} = x'_{ijt}\beta + \mu_{ij} + \varepsilon_{ijt} \tag{12}$$

with

$$\Omega = \sigma_\mu^2 (I_{N_i N_j} \otimes J_T) + \sigma_\varepsilon^2 I_{N_i N_j T} \,. \tag{13}$$

Note, that model (12) can be considered in fact as a straight panel data model, where the individuals are now the (ij) pairs (so essentially it does not take into account the three-dimensional nature of the data).

2.2. Spectral Decomposition of the Covariance Matrixes

To estimate the above models, the inverse of Ω is needed, a matrix of size $(N_i N_j T \times N_i N_j T)$. For even moderately large samples, this can be practically unfeasible without further elaboration. The common practise is to use the spectral decomposition of Ω, which in turn gives the inverse as a function of fairly standard matrixes (see *Wansbeek and Kapteyn* [1982]). We derive the algebra for model (2), Ω^{-1} for all other models can de derived likewise, so we only present the final results. First, consider a simple rewriting of the identity matrix

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

[3] This model has in fact been introduced in *Matyas* [1998].

with $\bar{J}_{N_i} = \frac{1}{N_i} J_{N_i}$. Now Ω becomes

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

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

$$B_{ij.} = (Q_{N_i} \otimes Q_{N_j} \otimes \bar{J}_T)$$

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

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

Notice, that the covariance matrix of all three-way error components model can be represented by these B building blocks. For model (2), this means

$$\begin{aligned}\Omega =&\sigma_\varepsilon^2 B_{ijt} + (\sigma_\varepsilon^2 + T\sigma_\mu^2)B_{ij.} + (\sigma_\varepsilon^2 + N_j\sigma_\upsilon^2)B_{i.t} + (\sigma_\varepsilon^2 + N_i\sigma_\zeta^2)B_{.jt} \\ &+ (\sigma_\varepsilon^2 + T\sigma_\mu^2 + N_j\sigma_\upsilon^2)B_{i..} + (\sigma_\varepsilon^2 + T\sigma_\mu^2 + N_i\sigma_\zeta^2)B_{.j.} \\ &+ (\sigma_\varepsilon^2 + N_j\sigma_\upsilon^2 + N_i\sigma_\zeta^2)B_{..t} + (\sigma_\varepsilon^2 + T\sigma_\mu^2 + N_j\sigma_\upsilon^2 + N_i\sigma_\zeta^2)B_{...}\,.\end{aligned} \tag{14}$$

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

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

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

$$\begin{aligned}\sigma_\varepsilon^2\Omega^{-1} =&I_{N_iN_jT} - (1-\theta_1)(\bar{J}_{N_i} \otimes I_{N_jT}) - (1-\theta_2)(I_{N_i} \otimes \bar{J}_{N_j} \otimes I_T) \\ &- (1-\theta_3)(I_{N_iN_j} \otimes \bar{J}_T) + (1-\theta_1-\theta_2+\theta_4)(\bar{J}_{N_iN_j} \otimes I_T) \\ &+ (1-\theta_1-\theta_3+\theta_5)(\bar{J}_{N_i} \otimes I_{N_j} \otimes \bar{J}_T) + (1-\theta_2-\theta_3+\theta_6)(I_{N_i} \otimes \bar{J}_{N_jT}) \\ &- (1-\theta_1-\theta_2-\theta_3+\theta_4+\theta_5+\theta_6-\theta_7)\bar{J}_{N_iN_jT}\,,\end{aligned} \tag{15}$$

with

$$\theta_1 = \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + N_i\sigma_\zeta^2}\,, \quad \theta_2 = \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + N_j\sigma_\upsilon^2}\,, \quad \theta_3 = \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T\sigma_\mu^2}$$

$$\theta_4 = \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + N_j\sigma_\upsilon^2 + N_i\sigma_\zeta^2}\,, \quad \theta_5 = \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T\sigma_\mu^2 + N_i\sigma_\zeta^2}\,,$$

$$\theta_6 = \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T\sigma_\mu^2 + N_j\sigma_\upsilon^2}\,, \quad \text{and} \quad \theta_7 = \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T\sigma_\mu^2 + N_j\sigma_\upsilon^2 + N_i\sigma_\zeta^2}\,.$$

The good thing is that we can fully get rid of the matrix notations, as $\sigma_\varepsilon^2\Omega^{-1}y$ can be written up in scalar form as well. In particular, this is the same as transforming y as

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

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

For other models, the job is essentially the same. For model (4),

$$\begin{aligned}\sigma_\varepsilon^2\Omega^{-1} =& I_{N_iN_jT} - (I_{N_i}\otimes\bar{J}_{N_j}\otimes I_T) - (\bar{J}_{N_i}\otimes I_{N_jT}) + (\bar{J}_{N_iN_j}\otimes I_T) \\ &+ \frac{\sigma_\varepsilon^2}{N_i\sigma_\zeta^2+\sigma_\varepsilon^2}((\bar{J}_{N_i}\otimes I_{N_jT}) - (\bar{J}_{N_iN_j}\otimes I_T)) \\ &+ \frac{\sigma_\varepsilon^2}{N_j\sigma_\upsilon^2+\sigma_\varepsilon^2}((I_{N_i}\otimes\bar{J}_{N_j}\otimes I_T)) - (\bar{J}_{N_iN_j}\otimes I_T)) \\ &+ \frac{\sigma_\varepsilon^2}{N_j\sigma_\upsilon^2+N_i\sigma_\zeta^2+\sigma_\varepsilon^2}(\bar{J}_{N_iN_j}\otimes I_T),\end{aligned}$$

which translates into

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

with

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

For model (6), the inverse of the covariance matrix is even simpler,

$$\sigma_\varepsilon^2\Omega^{-1} = I_{N_iN_jT} - (\bar{J}_{N_i}\otimes I_{N_jT}) + \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+N_i\sigma_\zeta^2}(\bar{J}_{N_i}\otimes I_{N_jT})\,,$$

which is the same as transforming y as

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

For Model (8), it is

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

or simply seen, how it transforms y,

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

with

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

For model (10), we get

$$\begin{aligned}\sigma_\varepsilon^2\Omega^{-1} =& I_{N_iN_jT} - (\bar{J}_{N_iN_j} \otimes I_T) - (\bar{J}_{N_i} \otimes I_{N_j} \otimes \bar{J}_T) - (I_{N_i} \otimes \bar{J}_{N_jT}) + 2\bar{J}_{N_iN_jT} \\ &+ \frac{\sigma_\varepsilon^2}{N_jT\sigma_\upsilon^2 + \sigma_\varepsilon^2}((I_{N_i} \otimes \bar{J}_{N_jT}) - \bar{J}_{N_iN_jT}) \\ &+ \frac{\sigma_\varepsilon^2}{N_iT\sigma_\zeta^2 + \sigma_\varepsilon^2}((\bar{J}_{N_i} \otimes I_{N_j} \otimes \bar{J}_T) - \bar{J}_{N_iN_jT}) \\ &+ \frac{\sigma_\varepsilon^2}{N_iN_j\sigma_\lambda^2 + \sigma_\varepsilon^2}((\bar{J}_{N_iN_j} \otimes I_T) - \bar{J}_{N_iN_jT}) \\ &+ \frac{\sigma_\varepsilon^2}{N_jT\sigma_\upsilon^2 + N_iT\sigma_\zeta^2 + N_iN_j\sigma_\lambda^2 + \sigma_\varepsilon^2}\bar{J}_{N_iN_jT}\,,\end{aligned}$$

or applied to y,

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

where

$$\theta_{15} = \frac{\sigma_\varepsilon^2}{N_j T \sigma_\upsilon^2 + \sigma_\varepsilon^2}\,, \quad \theta_{16} = \frac{\sigma_\varepsilon^2}{N_i T \sigma_\zeta^2 + \sigma_\varepsilon^2}\,, \quad \theta_{17} = \frac{\sigma_\varepsilon^2}{N_i N_j \sigma_\lambda^2 + \sigma_\varepsilon^2}\,, \quad \text{and}$$

$$\theta_{18} = \frac{\sigma_\varepsilon^2}{N_j T \sigma_\upsilon^2 + N_i T \sigma_\zeta^2 + N_i N_j \sigma_\lambda^2 + \sigma_\varepsilon^2}$$

whereas, for model (12), the inversion gives

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

leading to the simplest transformation

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

Table 1 summarizes the key elements in each models' inverse covariance matrix in the finite case.

Table 1 : Structure of the Ω^{-1} matrices.

Model	(2)	(4)	(6)	(8)	(10)	(12)
I_{N^2T}	+	+	+	+	+	+
$(I_{N^2} \otimes \bar{J}_T)$	+			+		+
$(I_N \otimes \bar{J}_N \otimes I_T)$	+	+				
$(\bar{J}_N \otimes I_{NT})$	+	+	+			
$(I_N \otimes \bar{J}_{NT})$	+				+	
$(\bar{J}_N \otimes I_N \otimes \bar{J}_T)$	+				+	
$(\bar{J}_{N^2} \otimes I_T)$	+	+		+	+	
$\bar{J}_{N^2T}$	+			+	+	

A "+" sign in a column says which building element is part of the given model's Ω^{-1}. If the "+"-s in the column of model A cover that of model B's means that model B is nested into model A. It can be seen that, for example, all models are in fact nested into (2), or that model (12) is nested into model (8).

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

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

which is the covariance matrix of the within estimator. Table 2 collects the asymptotic conditions, when (F)GLS estimator is converging to the Within one.

Table 2 : Asymptotic Conditions when the FGLS converges to the Within

Model	Condition
(2)	$N_i\to\infty,\ N_j\to\infty,\ T\to\infty$
(4)	$N_i\to\infty,\ N_j\to\infty$
(6)	$N_i\to\infty$
(8)	$(N_i\to\infty,\ T\to\infty)$ or $(N_j\to\infty,\ T\to\infty)$
(10)	$(N_i\to\infty,\ N_j\to\infty)$ or $(N_i\to\infty,\ T\to\infty)$ or $(N_j\to\infty,\ T\to\infty)$
(12)	$T\to\infty$

3. FGLS Estimation

To make the FGLS estimator operational, we need estimators for the variance components. Let us start again with model (2), for the other models, the job is essentially the same. Using that the error components are pairwise uncorrelated,

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

By introducing different Within transformations and so projecting the error components into different subspaces of the original three-dimensional space, we can derive further identifying equations. The appropriate Within transformation for model (2) (see for details *Balazsi et. al.* [2015]) is

$$\tilde{u}_{ijt} = u_{ijt} - \bar{u}_{.jt} - \bar{u}_{i.t} - \bar{u}_{ij.} + \bar{u}_{..t} + \bar{u}_{.j.} + \bar{u}_{i..} - \bar{u}_{...}\,. \tag{16}$$

Note, that this transformation corresponds to the projection matrix

$$\begin{aligned}M =& I_{N_iN_jT} - (I_{N_iN_j}\otimes \bar{J}_T) - (I_{N_i}\otimes \bar{J}_{N_j}\otimes I_T) - (\bar{J}_{N_i}\otimes I_{N_jT}) \\ &+ (I_{N_i}\otimes \bar{J}_{N_jT}) + (\bar{J}_{N_i}\otimes I_{N_j}\otimes \bar{J}_T) + (\bar{J}_{N_iN_j}\otimes I_T) - \bar{J}_{N_iN_jT}\,,\end{aligned}$$

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

$$\begin{aligned} E[\tilde{u}^2_{ijt}] = E[\tilde{\varepsilon}^2_{ijt}] &= E[(\varepsilon_{ijt} - \bar{\varepsilon}_{.jt} - \bar{\varepsilon}_{i.t} - \bar{\varepsilon}_{ij.} + \bar{\varepsilon}_{..t} + \bar{\varepsilon}_{.j.} + \bar{\varepsilon}_{i..} - \bar{\varepsilon}_{...})^2] \\ &= \frac{(N_i - 1)(N_j - 1)(T - 1)}{N_i N_j T} \sigma^2_\varepsilon \,, \end{aligned}$$

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

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

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

$$E[(\tilde{u}^a_{ijt})^2] = E[(\tilde{\mu}^a_{ij} + \tilde{\varepsilon}^a_{ijt})^2] = E[(\tilde{\mu}^a_{ij})^2] + E[(\tilde{\varepsilon}^a_{ijt})^2] = \frac{(N_i - 1)(N_j - 1)}{N_i N_j} (\sigma^2_\mu + \sigma^2_\varepsilon) \,.$$

Transforming according to

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

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

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

Finally, using

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

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

$$E[(\tilde{u}^c_{ijt})^2] = E[(\tilde{\zeta}^c_{jt} + \tilde{\varepsilon}^c_{ijt})^2] = E[(\tilde{\zeta}^c_{jt})^2] + E[(\tilde{\varepsilon}^c_{ijt})^2] = \frac{(N_j - 1)(T - 1)}{N_j T} (\sigma^2_\zeta + \sigma^2_\varepsilon) \,.$$

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

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

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

Note, however, that the FGLS estimator of model (2) is only consistent if the data grows in at least two dimensions, that is, any two of $N_i \to \infty$, $N_j \to \infty$, and $T \to \infty$ has to hold. This is, because σ^2_{μ} (the variance of μ_{ij}) cannot be estimated consistently, when only $T \to \infty$, σ^2_{υ}, when only $N_i \to \infty$, and so on. For the consistency of the FGLS we need all variance components to be estimated consistently, something which holds only if the data grows in at least two dimensions. Table 3 collects the conditions needed for consistency for all models considered. So what if, for example, the data is such that N_i is large, but N_j and T are small (like in case of a employee-firm data with an extensive number of workers, but with few hiring firms observed annually)? This would mean, that σ^2_{μ} and σ^2_{υ} is estimated consistently, unlike σ^2_{ζ}. In such cases, it makes more sense to assume ζ_{jt} to be fixed instead of random (while still assuming the randomness of μ_{ij} and υ_{it}), arriving to the so-called mixed-effects models, something we explore in Section 5.

We can estimate the variance components of the other models in a similar way. As the algebra is essentially the same, we only present here the main results. For model (4),

$$
E[\tilde{u}^2_{ijt}] = \frac{(N_i-1)(N_j-1)}{N_iN_j}\sigma^2_{\varepsilon} \, , \quad E[(\tilde{u}^a_{ijt})^2] = \frac{N_i-1}{N_i}(\sigma^2_{\upsilon}+\sigma^2_{\varepsilon}) \quad \text{and}
$$

$$
E[(\tilde{u}^b_{ijt})^2] = \frac{N_j-1}{N_j}(\sigma^2_{\zeta}+\sigma^2_{\varepsilon}) \, ,
$$

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

$$\begin{aligned}
M &= I_{N_iN_jT} - (\bar{J}_{N_i} \otimes I_{N_jT}) - (I_{N_i} \otimes \bar{J}_{N_j} \otimes I_T) + (\bar{J}_{N_iN_j} \otimes I_T) \\
M^a &= I_{N_iN_jT} - (\bar{J}_{N_i} \otimes I_{N_jT}) \\
M^b &= I_{N_iN_jT} - (I_{N_i} \otimes \bar{J}_{N_j} \otimes I_T)
\end{aligned}$$

respectively. The estimators for the variance components then are

$$\hat{\sigma}^2_\varepsilon = \frac{1}{(N_i-1)(N_j-1)T}\sum_{ijt}\tilde{\hat{u}}^2_{ijt}\,, \quad \hat{\sigma}^2_\upsilon = \frac{1}{(N_i-1)N_jT}\sum_{ijt}(\tilde{\hat{u}}^a_{ijt})^2 - \hat{\sigma}^2_\varepsilon\,, \quad \text{and}$$
$$\hat{\sigma}^2_\zeta = \frac{1}{N_i(N_j-1)T}\sum_{ijt}(\tilde{\hat{u}}^b_{ijt})^2 - \hat{\sigma}^2_\varepsilon\,,$$

where again, $\tilde{\hat{u}}_{ijt}$, $\tilde{\hat{u}}^a_{ijt}$ and $\tilde{\hat{u}}^b_{ijt}$ are obtained by transforming the residual $\hat{u}_{ijt}$ according to M, M^a, and M^b respectively. For model (6), as

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

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

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

For model (8),

$$E[\tilde{u}^2_{ijt}] = \frac{(N_iN_j-1)(T-1)}{N_iN_jT}\sigma^2_\varepsilon\,, \quad E[(\tilde{u}^a_{ijt})^2] = \frac{N_iN_j-1}{N_iN_j}(\sigma^2_\mu + \sigma^2_\varepsilon)\,, \quad \text{and}$$
$$E[(\tilde{u}^b_{ijt})^2] = \frac{T-1}{T}(\sigma^2_\lambda + \sigma^2_\varepsilon)\,,$$

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

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

respectively. The estimators for the variance components are

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

$$\hat{\sigma}^2_\lambda = \frac{1}{N_iN_j(T-1)}\sum_{ijt}(\tilde{\hat{u}}^b_{ijt})^2 - \hat{\sigma}^2_\varepsilon\,.$$

For model (10), as

$$\begin{aligned}
E[\tilde{u}^2_{ijt}] &= \frac{(N_iN_j-1)T-(N_i-1)-(N_j-1)}{N_iN_jT}\sigma^2_\varepsilon \\
E[(\tilde{u}^a_{ijt})^2] &= \frac{(N_iN_j-1)T-(N_j-1)}{N_iN_jT}(\sigma^2_\upsilon+\sigma^2_\varepsilon) \\
E[(\tilde{u}^b_{ijt})^2] &= \frac{(N_iN_j-1)T-(N_i-1)}{N_iN_jT}(\sigma^2_\zeta+\sigma^2_\varepsilon) \\
E[(\tilde{u}^c_{ijt})^2] &= \frac{N_iN_jT-N_i-N_j+1}{N_iN_jT}(\sigma^2_\mu+\sigma^2_\varepsilon)
\end{aligned}$$

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

$$\begin{aligned}
M &= I_{N_iN_jT} - (\bar{J}_{N_iN_j}\otimes I_T) - (\bar{J}_{N_i}\otimes I_{N_j}\otimes \bar{J}_T) - (I_{N_i}\otimes \bar{J}_{N_jT}) + 2\bar{J}_{N_iN_jT} \\
M^a &= I_{N_iN_jT} - (\bar{J}_{N_iN_j}\otimes I_T) - (\bar{J}_{N_i}\otimes I_{N_j}\otimes \bar{J}_T) + \bar{J}_{N_iN_jT} \\
M^b &= I_{N_iN_jT} - (\bar{J}_{N_iN_j}\otimes I_T) - (I_{N_i}\otimes \bar{J}_{N_jT}) + \bar{J}_{N_iN_jT} \\
M^c &= I_{N_iN_jT} - (\bar{J}_{N_i}\otimes I_{N_j}\otimes \bar{J}_T) - (I_{N_i}\otimes \bar{J}_{N_jT}) + \bar{J}_{N_iN_jT}
\end{aligned}$$

respectively. The estimators for the variance components are

$$\begin{aligned}
\hat{\sigma}^2_\varepsilon &= \frac{1}{(N_iN_j-1)T-(N_i-1)-(N_j-1)}\sum_{ijt}\tilde{\hat{u}}^2_{ijt} \\
\hat{\sigma}^2_\upsilon &= \frac{1}{(N_iN_j-1)T-(N_j-1)}\sum_{ijt}(\tilde{\hat{u}}^a_{ijt})^2 - \hat{\sigma}^2_\varepsilon \\
\hat{\sigma}^2_\zeta &= \frac{1}{(N_iN_j-1)T-(N_i-1)}\sum_{ijt}(\tilde{\hat{u}}^b_{ijt})^2 - \hat{\sigma}^2_\varepsilon \\
\hat{\sigma}^2_\lambda &= \frac{1}{N_iN_jT-N_i-N_j+1}\sum_{ijt}(\tilde{\hat{u}}^c_{ijt})^2 - \hat{\sigma}^2_\varepsilon\,.
\end{aligned}$$

Lastly, for model (12) we get

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

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

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

Standard errors are computed accordingly, using $\text{Var}(\hat{\beta}_{FGLS}) = (X'\hat{\Omega}^{-1}X)^{-1}$. In the limiting cases, the usual normalization factors are needed to obtain finite variances. If, for example N_i and T are growing, $\sqrt{N_iT}(\hat{\beta}_{FGLS} - \beta)$ has a normal distribution with zero mean, and $Q_{X\Omega X}^{-1}$ variance, where $Q_{X\Omega X}^{-1} = \text{plim}_{N_i,T\to\infty} \frac{X'\hat{\Omega}^{-1}X}{N_iT}$ is assumed to be a finite, positive definite matrix. This holds model-wide. However, we have no such luck with the OLS estimator. The issue is best illustrated with model (12). It can be shown, just as with the usual 2D panel models, $V(\hat{\beta}_{OLS}) = (X'X)^{-1}X'\hat{\Omega}X(X'X)^{-1}$ (with $\hat{\Omega}$ being model-specific, but let us assume for now, that it corresponds to (13)).

Table 3 : Sample conditions for the consistency of the FGLS Estimator

Model	Consistency requirements
(2)	$(N_i \to \infty,\ N_j \to \infty)$ or $(N_i \to \infty,\ T \to \infty)$ or $(N_j \to \infty,\ T \to \infty)$
(4)	$(T \to \infty)$ or $(N_i \to \infty,\ N_j \to \infty)$
(6)	$(N_j \to \infty)$ or $(T \to \infty)$
(8)	$(N_i \to \infty,\ T \to \infty)$ or $(N_j \to \infty,\ T \to \infty)$
(10)	$(N_i \to \infty,\ N_j \to \infty,\ T \to \infty)$
(12)	$(N_i \to \infty)$ or $(N_j \to \infty)$

In the asymptotic case, when $N_i, N_j \to \infty$, $\sqrt{N_iN_j}(\hat{\beta}_{OLS} - \beta)$ has a normal distribution with finite variance, but this variance grows without bound (at rate $O(T)$) once $T \to \infty$. That is, an extra $1/\sqrt{T}$ normalization factor has to be added to regain a normal distribution with bounded variance. Table 4 collects normalization factors needed for a finite $V(\hat{\beta}_{OLS})$ for the different models considered. As it is uncommon to normalize with 1, or with expression like $\frac{\sqrt{N_i N_j}}{\sqrt{A}}$, some insights into the normalizations are given in Appendix 1.

Table 4 : Normalization factors for the finiteness of $\hat{\beta}_{OLS}$

Model	(2)	(4)	(6)	(8)	(10)	(12)
$N_i \to \infty$	1	1	1	1	1	$\sqrt{N_i}$
$N_j \to \infty$	1	1	$\sqrt{N_j}$	1	1	$\sqrt{N_j}$
$T \to \infty$	1	$\sqrt{T}$	$\sqrt{T}$	1	1	1
$N_i, N_j \to \infty$	$\frac{\sqrt{N_i N_j}}{\sqrt{A}}$	$\frac{\sqrt{N_i N_j}}{\sqrt{A}}$	$\sqrt{N_j}$	1	1	$\sqrt{N_i N_j}$
$N_i, T \to \infty$	$\frac{\sqrt{N_i T}}{\sqrt{A}}$	$\sqrt{T}$	$\sqrt{T}$	$\frac{\sqrt{N_i T}}{\sqrt{A}}$	1	$\sqrt{N_i}$
$N_j, T \to \infty$	$\frac{\sqrt{N_j T}}{\sqrt{A}}$	$\sqrt{T}$	$\sqrt{N_j T}$	$\frac{\sqrt{N_j T}}{\sqrt{A}}$	1	$\sqrt{N_j}$
$N_i, N_j, T \to \infty$	$\frac{\sqrt{N_i N_j T}}{\sqrt{A}}$	$\frac{\sqrt{N_i N_j}}{\sqrt{A}}\sqrt{T}$	$\sqrt{N_j T}$	$\frac{\sqrt{N_i N_j T}}{\sqrt{A}}$	$\frac{\sqrt{N_i N_j T}}{\sqrt{A_1 A_2}}$	$\sqrt{N_i N_j}$

where A is the sample size which grows with the highest rate, (N_i, N_j, or T), and A_1, A_2 are the two sample sizes which grow with the highest rates.

An other interesting aspect is revealed by comparing Tables 2 and 3, that is the consistency requirements for the estimation of the variance components (Table 2) and the asymptotic results, when the FGLS converges to the Within estimator (Table 3).

Table 5 : Asymptotic Results when the OLS should be used

Model	(2)	(4)	(6)	(8)	(10)	(12)
$N_i \to \infty$			+			−
$N_j \to \infty$			−			−
$T \to \infty$		−	−			+
$N_i, N_j \to \infty$	−	+	+		+	−
$N_i, T \to \infty$	−	−	+	+	+	+
$N_j, T \to \infty$	−	−	−	+	+	+
$N_i, N_j, T \to \infty$	+	+	+	+	+	+

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

As can be seen from Table 5, for all models the FGLS is consistent if all N_i, N_j, T go to infinity, but in these cases the (F)GLS estimator converges to the Within. This is problematic, as some parameters, previously estimable, become suddenly unidentified.

In such cases, we have to rely on the OLS estimates, rather than the FGLS. This is generally the case whenever a "+" sign is found in Table 5, most significant for models (8) and (10). For them, the FGLS is only consistent, when it is in fact the Within Estimator, leading to likely severe identification issues. The best case scenarios are indicated with a "-" sign, where the respective asymptotics are already enough for the consistency of the FGLS, but do not yet cause identification problems. Lastly, blank spaces are left in the table if, under the given asymptotic, the FGLS is not consistent. In such cases we can again rely on the consistency of the OLS, but its standard errors are inconsistent, just as with the FGLS.

4. Unbalanced Data

4.1 Structure of the Covariance Matrixes

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

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

For the all-encompassing model (2), u_{ijt} can be stacked into vector u. Remember, that in the complete case it is

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

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

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

which is identical to (3). However, in the case of missing data, we have to modify the underlying D_k dummy matrixes to reflect the unbalanced nature of the data. For every (ij) pair, let V_{ij} denote the size $(|T_{ij}| \times T)$ matrix, which we obtain from the $(T \times T)$ identity matrix by deleting rows corresponding to missing observations.[4] With this, the incomplete D_k dummies are

$$\begin{aligned} D_1 &= \text{diag}\{V_{11}\iota_T, V_{12}\iota_T, \dots, V_{N_iN_j}\iota_T\} \quad \text{of size} \quad (\sum_{ij}|T_{ij}| \times \tilde{n}_{ij})\,, \\ D_2 &= \text{diag}\left\{(V_{11}', V_{12}', \dots, V_{1N_j}')', \dots, (V_{N_i1}', V_{N_i2}', \dots, V_{N_iN_j}')'\right\} \\ &\qquad\qquad\qquad\qquad\qquad\qquad \text{of size} \quad (\sum_{ij}|T_{ij}| \times \tilde{n}_{it}) \\ D_3 &= \left(\text{diag}\{V_{11}', V_{12}', \dots, V_{1N_j}'\}', \dots, \text{diag}\{V_{N_i1}', V_{N_i2}', \dots, V_{N_iN_j}'\}'\right)' \\ &\qquad\qquad\qquad\qquad\qquad\qquad \text{of size} \quad (\sum_{ij}|T_{ij}| \times \tilde{n}_{jt})\,, \end{aligned}$$

where, remember, $\tilde{n}_{ij}$, $\tilde{n}_{it}$, and $\tilde{n}_{jt}$ denote the number of individual pairs, the number of *individual i–time* pairs, and the number of *individual j–time* pairs respectively (in complete panels, these are N_iN_j, N_iT, and N_jT respectively). This in turn can be used to construct the covariance matrix as

$$\Omega = E(uu') = I_{\sum_{ij}|T_{ij}|}\sigma^2_\varepsilon + D_1D_1'\sigma^2_\mu + D_2D_2'\sigma^2_\upsilon + D_3D_3'\sigma^2_\zeta$$

[4] If, for example, $t = 1, 4, 10$ are missing for some (ij), we delete rows 1, 4, and 8 from I_T to get V_{ij}.

of size $\left(\sum_{ij}|T_{ij}| \times \sum_{ij}|T_{ij}|\right)$. If the data is complete, the above covariance structure in fact gives back (3). The job is the same for other models. For models (4) and (6),

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

and

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

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

$$\Omega = I_{\sum_{ij}|T_{ij}|}\sigma_\varepsilon^2 + D_2 D_2' \sigma_\upsilon^2 + D_3 D_3' \sigma_\zeta^2$$

for model (4), and

$$\Omega = I_{\sum_{ij}|T_{ij}|}\sigma_\varepsilon^2 + D_3 D_3' \sigma_\zeta^2$$

for model (6). Again, if the panel were in fact complete, we would get back (5) and (7). The incomplete data covariance matrix of model (8) is

$$\Omega = I_{\sum_{ij}|T_{ij}|}\sigma_\varepsilon^2 + D_1 D_1' \sigma_\mu^2 + D_4 D_4' \sigma_\lambda^2 \,,$$

with

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

The covariance matrix for Model (10) is

$$\Omega = I_{\sum_{ij}|T_{ij}|}\sigma_\varepsilon^2 + D_5 D_5' \sigma_\upsilon^2 + D_6 D_6' \sigma_\zeta^2 + D_4 D_4' \sigma_\lambda^2 \,,$$

where

$$D_5 = \operatorname{diag}\left\{(V_{11}'\iota_T, V_{12}'\iota_T, \ldots, V_{1N_j}'\iota_T)', \ldots, (V_{N_i1}'\iota_T, V_{N_i2}'\iota_T, \ldots, V_{N_iN_j}'\iota_T)'\right\}$$
$$D_6 = \left(\operatorname{diag}\{V_{11}'\iota_T, V_{12}'\iota_T, \ldots, V_{1N_j}'\iota_T\}', \ldots, \operatorname{diag}\{V_{N_i1}'\iota_T, V_{N_i2}'\iota_T, \ldots, V_{N_iN_j}'\iota_T\}'\right)' \,.$$

of sizes $(\sum_{ij}|T_{ij}| \times N_i)$, and $(\sum_{ij}|T_{ij}| \times N_j)$. Lastly, for model (12),

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

simply, with everything defined above.

An important practical difficulty is that the spectral decomposition of the covariance matrixes introduced in Section 2.2 are no longer valid, so the inversion of Ω for very large data sets can be forbidding. To go around this problem, let us construct the *quasi-spectral decomposition* of the incomplete data covariance matrixes, which is simply done by leaving out the missing rows from the appropriate B. Specifically, let us call B^* the incompleteness-adjusted versions of any B, which we get by removing the rows corresponding to the missing observations. For example, the spectral decomposition (14) for the all-encompassing model reads as

$$\begin{aligned}\Omega^* =&\sigma_\varepsilon^2 B^*_{ijt} + (\sigma_\varepsilon^2 + T\sigma_\mu^2)B^*_{ij.} + (\sigma_\varepsilon^2 + N_j\sigma_\upsilon^2)B^*_{i.t} + (\sigma_\varepsilon^2 + N_i\sigma_\zeta^2)B^*_{.jt}\\ &+ (\sigma_\varepsilon^2 + T\sigma_\mu^2 + N_j\sigma_\upsilon^2)B^*_{i..} + (\sigma_\varepsilon^2 + T\sigma_\mu^2 + N_i\sigma_\zeta^2)B^*_{.j.}\\ &+ (\sigma_\varepsilon^2 + N_j\sigma_\upsilon^2 + N_i\sigma_\zeta^2)B^*_{..t} + (\sigma_\varepsilon^2 + T\sigma_\mu^2 + N_j\sigma_\upsilon^2 + N_i\sigma_\zeta^2)B^*_{...} \,,\end{aligned}$$

where now all B^* have number of rows equal to $\sum_{ij}|T_{ij}|$. Of course, this is not a correct spectral decomposition of Ω, but helps to define the following conjecture.[5] Namely, when the number of missing observations relative to the total number of observations is small, the inverse of Ω based on the quasi-spectral decomposition of it, Ω^{*-1}, approximate arbitrarily well Ω^{-1}. More precisely, if $[N_iN_jT - \sum_i\sum_j|T_{ij}|]/[N_iN_jT] \rightarrow 0$, then $(\Omega^{-1} - \Omega^{*-1}) \rightarrow 0$. This means that in large data sets, when the number of missing observation is small relative to the total number of observations, Ω^{*-1} can safely be used in the GLS estimator instead of Ω^{-1}. Let us seen an example. Multi-dimensional panel data are often used to deal with trade (gravity) models. In these cases, however, when country i trade with country j, there are no (ii) (or (jj)) observations, there is no self-trade. Then the total number of observations is $N^2T - NT$ with NT being the number of missing observations due to no self-trade. Given that $[N^2T - NT]/NT \rightarrow 0$ as the sample size increases, the quasi-spectral decomposition can be used in large data.

4.2 The Inverse of the Covariance Matrixes

The solution proposed above, however, suffers from two potential drawbacks. First, the inverse, though reached at very low cost, may not be accurate enough, and second, when the "holes" in the data are substantial this method cannot be used. These reasons spurs us to derive the analytically correct inverse of the covariance matrixes at the lowest possible cost. To do that, we have to generalize the results of *Wansbeek and Kapteyn* [1989]. This leads us, for model (2), to

$$\sigma_\varepsilon^2\Omega^{-1} = P^b - P^bD_3(R^c)^{-1}D_3'P^b \tag{17}$$

[5] This can be demonstrated by simulation.

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

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

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

where D_1, D_2, D_3 are the incompleteness-adjusted dummy variable matrixes, and are used to construct the P and R matrixes sequentially: first, construct R^a to get P^a, then construct R^b to get P^b, and finally, construct R^c to get P^c. Proof of (17) can be found in Appendix 2. Note, that to get the inverse, we have to invert $\min\{N_iT;\, N_jT;\, N_iN_j\}$ matrixes. The quasi-scalar form of (17) (which corresponds to the incomplete data version of transformation (15)) is

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

with

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

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

For the other models, the job is essentially the same, only the number of steps in obtaining the inverse is smaller (as the number of different random effects decreases). For model (4), it is, with appropriately redefining P and R,

$$\sigma_\varepsilon^2 \Omega^{-1} = P^a - P^a D_3 (R^b)^{-1} D_3' P^a \,, \tag{18}$$

where now

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

with the largest matrix to inverted now of size $\min\{N_iT;\, N_jT\}$. For model (6), it is even more simple,

$$\sigma_\varepsilon^2 \Omega^{-1} = I - D_3 (R^a)^{-1} D_3' \quad \text{with} \quad R^a = D_3' D_3 + \frac{\sigma_\varepsilon^2}{\sigma_\zeta^2} I \,, \tag{19}$$

defining the scalar transformation

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

with n_{jt} being the number of observations for a given (jt) pair. For model (8), the inverse is

$$\sigma_\varepsilon^2 \Omega^{-1} = P^a - P^a D_4 (R^b)^{-1} D_4' P^a \tag{20}$$

where

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

and we have to invert a $\min\{N_i N_j; T\}$ sized matrix. For model (10), the inverse is again the result of a three-step procedure:

$$\sigma_\varepsilon^2 \Omega^{-1} = P^b - P^b D_4 (R^c)^{-1} D_4' P^b \,, \tag{21}$$

where

$$R^c = D_4' P^b D_4 + \frac{\sigma_\varepsilon^2}{\sigma_\lambda^2} I \,, \quad P^b = P^a - P^a D_6 (R^b)^{-1} D_6' P^a \,, \quad R^b = D_6' P^a D_6 + \frac{\sigma_\varepsilon^2}{\sigma_\zeta^2} I \,,$$

$$P^a = I - D_5 (R^a)^{-1} D_5' \quad \text{and} \quad R^a = D_5' D_5 + \frac{\sigma_\varepsilon^2}{\sigma_\upsilon^2} I \,,$$

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

$$\sigma_\varepsilon^2 \Omega^{-1} = I - D_1 (R^a)^{-1} D_1' \quad \text{with} \quad R^a = D_1' D_1 + \frac{\sigma_\varepsilon^2}{\sigma_\mu^2} I \,, \tag{22}$$

defining the scalar transformation

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

on a typical y_{ijt} variable.

4.3 Estimation of the Variance Components

Let us proceed to the estimation of the variance components. The estimators we used for the complete data cases are not applicable here, as, for example, transformation (16) does not eliminate μ_{ij}, υ_{it}, and ζ_{jt} from the composite disturbance term $u_{ijt} =$

$\mu_{ij} + \upsilon_{it} + \zeta_{jt} + \varepsilon_{ijt}$, when the data is incomplete. This problem can be tackled in two ways. We can derive incompleteness-robust alternative of (16), i.e., a transformation which clears the non-idiosyncratic random effects from u_{ijt}, in the case of incomplete data (see *Balazsi et. al.* [2015]). The problem is that most of these transformations involve the manipulation of large matrixes resulting in heavy computational burden. To avoid this we propose simple linear transformations, which on the one hand, are robust to incomplete data, and on the other hand, identify the variance components. Let us see, how this works for model (2). As before

$$E[u_{ijt}^2] = \sigma_\mu^2 + \sigma_\upsilon^2 + \sigma_\zeta^2 + \sigma_\varepsilon^2 \,, \tag{23}$$

but now, let us define

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

It can be seen that

$$\begin{aligned} E[(\tilde{u}_{ijt}^a)^2] &= \frac{|T_{ij}|-1}{|T_{ij}|}(\sigma_\upsilon^2 + \sigma_\zeta^2 + \sigma_\varepsilon^2) \,, \quad E[(\tilde{u}_{ijt}^b)^2] = \frac{n_{it}-1}{n_{it}}(\sigma_\mu^2 + \sigma_\zeta^2 + \sigma_\varepsilon^2) \,, \\ \text{and} \quad E[(\tilde{u}_{ijt}^c)^2] &= \frac{n_{jt}-1}{n_{jt}}(\sigma_\mu^2 + \sigma_\upsilon^2 + \sigma_\varepsilon^2) \,. \end{aligned} \tag{24}$$

Combining (23) with (24) identifies all four variance components. The appropriate estimators are then

$$\begin{aligned} \hat{\sigma}_\mu^2 &= \frac{1}{\sum_{ij}|T_{ij}|}\sum_{ijt}\hat{u}_{ijt}^2 - \frac{1}{\tilde{n}_{ij}}\sum_{ij}\frac{1}{|T_{ij}|-1}\sum_t(\tilde{\hat{u}}_{ijt}^a)^2 \\ \hat{\sigma}_\upsilon^2 &= \frac{1}{\sum_{ij}|T_{ij}|}\sum_{ijt}\hat{u}_{ijt}^2 - \frac{1}{\tilde{n}_{it}}\sum_{it}\frac{1}{n_{it}-1}\sum_j(\tilde{\hat{u}}_{ijt}^b)^2 \\ \hat{\sigma}_\zeta^2 &= \frac{1}{\sum_{ij}|T_{ij}|}\sum_{ijt}\hat{u}_{ijt}^2 - \frac{1}{\tilde{n}_{jt}}\sum_{jt}\frac{1}{n_{jt}-1}\sum_i(\tilde{\hat{u}}_{ijt}^c)^2 \\ \hat{\sigma}_\varepsilon^2 &= \frac{1}{\sum_{ij}|T_{ij}|}\sum_{ijt}\hat{u}_{ijt}^2 - \hat{\sigma}_\mu^2 - \hat{\sigma}_\upsilon^2 - \hat{\sigma}_\zeta^2 \,, \end{aligned} \tag{25}$$

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

The estimation strategy of the variance components is exactly the same for all the other models. Let us keep for now the definitions of $\tilde{u}^b_{ijt}$, and $\tilde{u}^c_{ijt}$. For model (4), with $u_{ijt} = \upsilon_{it} + \zeta_{jt} + \varepsilon_{ijt}$, the estimators read as

$$\begin{aligned}
\hat{\sigma}^2_\upsilon &= \frac{1}{\sum_{ij}|T_{ij}|}\sum_{ijt}\hat{u}^2_{ijt} - \frac{1}{\tilde{n}_{it}}\sum_{it}\frac{1}{n_{it}-1}\sum_j(\tilde{\hat{u}}^b_{ijt})^2 \\
\hat{\sigma}^2_\zeta &= \frac{1}{\sum_{ij}|T_{ij}|}\sum_{ijt}\hat{u}^2_{ijt} - \frac{1}{\tilde{n}_{jt}}\sum_{jt}\frac{1}{n_{jt}-1}\sum_i(\tilde{\hat{u}}^c_{ijt})^2 \\
\hat{\sigma}^2_\varepsilon &= \frac{1}{\sum_{ij}|T_{ij}|}\sum_{ijt}\hat{u}^2_{ijt} - \hat{\sigma}^2_\upsilon - \hat{\sigma}^2_\zeta\,,
\end{aligned} \tag{26}$$

whereas for model (6), with $u_{ijt} = \zeta_{jt} + \varepsilon_{ijt}$, they are

$$\begin{aligned}
\hat{\sigma}^2_\zeta &= \frac{1}{\sum_{ij}|T_{ij}|}\sum_{ijt}\hat{u}^2_{ijt} - \frac{1}{\tilde{n}_{jt}}\sum_{jt}\frac{1}{n_{jt}-1}\sum_i(\tilde{\hat{u}}^c_{ijt})^2 \\
\hat{\sigma}^2_\varepsilon &= \frac{1}{\sum_{ij}|T_{ij}|}\sum_{ijt}\hat{u}^2_{ijt} - \hat{\sigma}^2_\zeta\,,
\end{aligned} \tag{27}$$

Note, that these latter two estimators can be obtained from (25), by assuming $\hat{\sigma}^2_\mu = 0$ for model (4), and $\hat{\sigma}^2_\mu = \hat{\sigma}^2_\upsilon = 0$ for model (6).

For model (8), let us redefine the $\tilde{u}^k_{ijt}$-s, as

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

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

$$E[(\tilde{u}^a_{ijt})^2] = \frac{|T_{ij}|-1}{|T_{ij}|}(\sigma^2_\lambda + \sigma^2_\varepsilon)\,, \quad E[(\tilde{u}^b_{ijt})^2] = \frac{n_t-1}{n_t}(\sigma^2_\mu + \sigma^2_\varepsilon)\,,$$
$$\text{and} \quad E[u^2_{ijt}] = \sigma^2_\mu + \sigma^2_\lambda + \sigma^2_\varepsilon\,.$$

From this set of identifying equations, the estimators are simply

$$\begin{aligned}
\hat{\sigma}^2_\mu &= \frac{1}{\sum_{ij}|T_{ij}|}\sum_{ijt}\hat{u}^2_{ijt} - \frac{1}{\tilde{n}_{ij}}\sum_{ij}\frac{1}{|T_{ij}|-1}\sum_t(\tilde{\hat{u}}^a_{ijt})^2 \\
\hat{\sigma}^2_\lambda &= \frac{1}{\sum_{ij}|T_{ij}|}\sum_{ijt}\hat{u}^2_{ijt} - \frac{1}{T}\sum_t\frac{1}{n_t-1}\sum_{ij}(\tilde{\hat{u}}^b_{ijt})^2 \\
\hat{\sigma}^2_\varepsilon &= \frac{1}{\sum_{ij}|T_{ij}|}\sum_{ijt}\hat{u}^2_{ijt} - \hat{\sigma}^2_\mu - \hat{\sigma}^2_\lambda\,.
\end{aligned} \tag{28}$$

For model (12), with $u_{ijt} = \mu_{ij} + \varepsilon_{ijt}$, keeping the definition of $\tilde{u}^a_{ijt}$,

$$
\begin{aligned}
\hat{\sigma}^2_\mu &= \frac{1}{\sum_{ij} |T_{ij}|} \sum_{ijt} \hat{u}^2_{ijt} - \frac{1}{\tilde{n}_{ij}} \sum_{ij} \frac{1}{|T_{ij}| - 1} \sum_t (\hat{\tilde{u}}^a_{ijt})^2 \\
\hat{\sigma}^2_\varepsilon &= \frac{1}{\sum_{ij} |T_{ij}|} \sum_{ijt} \hat{u}^2_{ijt} - \hat{\sigma}^2_\mu \,.
\end{aligned} \tag{29}
$$

Finally, for model (10), as now $u_{ijt} = \upsilon_i + \zeta_j + \lambda_t + \varepsilon_{ijt}$, using

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

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

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

in turn leading to

$$
\begin{aligned}
\hat{\sigma}^2_\upsilon &= \frac{1}{\sum_{ij} |T_{ij}|} \sum_{ijt} \hat{u}^2_{ijt} - \frac{1}{\tilde{n}_i} \sum_{ij} \frac{1}{n_i - 1} \sum_{jt} (\hat{\tilde{u}}^a_{ijt})^2 \\
\hat{\sigma}^2_\upsilon &= \frac{1}{\sum_{ij} |T_{ij}|} \sum_{ijt} \hat{u}^2_{ijt} - \frac{1}{\tilde{n}_j} \sum_{it} \frac{1}{n_j - 1} \sum_{it} (\hat{\tilde{u}}^b_{ijt})^2 \\
\hat{\sigma}^2_\zeta &= \frac{1}{\sum_{ij} |T_{ij}|} \sum_{ijt} \hat{u}^2_{ijt} - \frac{1}{T} \sum_{jt} \frac{1}{n_t - 1} \sum_{ij} (\hat{\tilde{u}}^c_{ijt})^2 \\
\hat{\sigma}^2_\varepsilon &= \frac{1}{\sum_{ij} |T_{ij}|} \sum_{ijt} \hat{u}^2_{ijt} - \hat{\sigma}^2_\upsilon - \hat{\sigma}^2_\zeta - \hat{\sigma}^2_\lambda \,.
\end{aligned} \tag{30}
$$

5. Extensions

So far we have seen how to formulate and estimate three-way error components models. However, it is more and more typical to have data sets which require an even higher dimensional approach. As the number of feasible model formulations grow exponentially along with the dimensions, there is no point to attempt to collect all of them. Rather, we will take the 4D representation of the all-encompassing model (2), and show how the extension to higher dimensions can be carried out.

5.1. 4D and beyond

The baseline 4D model we use reads as, with $i = 1 \dots N_i$, $j = 1 \dots N_j$, $s = 1 \dots N_s$, and $t = 1 \dots T$,

$$y_{ijst} = x'_{ijst}\beta + \mu_{ijs} + \upsilon_{ist} + \zeta_{jst} + \lambda_{ijt} + \varepsilon_{ijst} = x'_{ijst}\beta + u_{ijst}\,, \tag{31}$$

where we keep assuming, that u (and its components individually) have zero mean, the components are pairwise uncorrelated, and further,

$$E(\mu_{ijs}\mu_{i'j's'}) = \begin{cases} \sigma_\mu^2 & i = i' \text{ and } j = j' \text{ and } s = s' \\ 0 & \text{otherwise} \end{cases}$$
$$E(\upsilon_{ist}\upsilon_{i's't'}) = \begin{cases} \sigma_\upsilon^2 & i = i' \text{ and } s = s' \text{ and } t = t' \\ 0 & \text{otherwise} \end{cases}$$
$$E(\zeta_{jst}\zeta_{j's't'}) = \begin{cases} \sigma_\zeta^2 & j = j' \text{ and } s = s' \text{ and } t = t' \\ 0 & \text{otherwise} \end{cases}$$
$$E(\lambda_{ijt}\zeta_{i'j't'}) = \begin{cases} \sigma_\lambda^2 & i = i' \text{ and } j = j' \text{ and } t = t' \\ 0 & \text{otherwise} \end{cases}$$

The covariance matrix of such error components formulation is

$$\begin{aligned}\Omega = E(uu') =& \sigma_\mu^2(I_{N_iN_jN_s} \otimes J_T) + \sigma_\upsilon^2(I_{N_i} \otimes J_{N_j} \otimes I_{N_sT}) + \sigma_\zeta^2(J_{N_i} \otimes I_{N_jN_sT}) \\ &+ \sigma_\lambda^2(I_{N_iN_j} \otimes J_{N_s} \otimes I_T) + \sigma_\varepsilon^2 I_{N_iN_jN_sT}\,.\end{aligned} \tag{32}$$

Its inverse can be simply calculated, following the method developed in Section 2.2, and the estimation of the variance components can also be derived as in Section 3, see for details Appendix 3.

The estimation procedure is not too difficult in the incomplete case either, at least not theoretically. Taking care of the unbalanced nature of the data in four dimensional panels has nevertheless a growing importance, as the likelihood of having missing and/or incomplete data increases dramatically in higher dimensions. Conveniently, we keep assuming, that our data is such, that, for each (ijs) individual, $t \in T_{ijs}$, where T_{ijs} is a subset of the index-set $\{1, \dots, T\}$, that is, we have $|T_{isj}|$ identical observations for each (ijs) pair. First, let us write up the covariance matrix of (31) as

$$\Omega = E(uu') = \sigma_\varepsilon^2 I + \sigma_\mu^2 D_1D_1' + \sigma_\upsilon^2 D_2D_2' + \sigma_\zeta^2 D_3D_3' + \sigma_\lambda^2 D_4D_4'\,, \tag{33}$$

where, in the complete case,

$$D_1 = (I_{N_iN_jN_s} \otimes \iota_T)\,, \quad D_2 = (I_{N_i} \otimes \iota_{N_j} \otimes I_{N_sT})\,, \quad D_3 = (\iota_{N_i} \otimes I_{N_jN_sT})\,, \quad \text{and}$$
$$D_4 = (I_{N_iN_j} \otimes \iota_{N_s} \otimes I_T)\,,$$

all being $(N_iN_jN_sT \times N_iN_jN_s)$, $(N_iN_jN_sT \times N_iN_sT)$, $(N_iN_jN_sT \times N_jN_sT)$, and $(N_iN_jN_sT \times N_iN_jT)$ sized matrixes respectively, but now we delete, from each D_k, the rows corresponding to the missing observations to reflect the unbalanced nature of the data. The inverse of such covariance formulation can be reached in steps, that is, one has to derive

$$\Omega^{-1}\sigma_\varepsilon^2 = P^c - P^c D_4 (R^d)^{-1} D_4' P^c \tag{34}$$

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

$$R^d = D_4' P^c D_4 + \frac{\sigma_\varepsilon^2}{\sigma_\lambda^2}\,, \quad P^c = P^b - P^b D_3 (R^c)^{-1} D_3' P^b\,, \quad R^c = D_3' P^b D_3 + \frac{\sigma_\varepsilon^2}{\sigma_\zeta^2}\,,$$

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

$$\text{and} \quad \mathrm{R^a} = \mathrm{D}_1'\mathrm{D}_1 + \frac{\sigma_\varepsilon^2}{\sigma_\mu^2}\,.$$

Even though the calculation above alleviates some of the "dimensionality curse"[6], to perform the inverse we still have to manipulate potentially large matrixes. The last step in finishing the FGLS estimation of the incomplete 4D models is to estimate the variance components. Fortunately, this is one of the easiest tasks, as we can do it in linear time, however, due to its size, we present the results in Appendix 3.

5.2. Mixed FE-RE Models

As briefly mentioned in Section 3, when one of the indexes is small, it may be better to treat the related effects as fixed rather than random. As an example, consider an *employee i, employer j, time t* type dataset, where we usually have large i, but relatively small j and t. All this means, that the all-encompassing model (2) should now be rewritten as

$$y_{ijt} = x_{ijt}'\beta + \alpha_{jt} + \mu_{ij} + \upsilon_{it} + \varepsilon_{ijt}\,, \tag{35}$$

or, similarly,

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

with $D_1 = (\iota_{N_i} \otimes I_{N_jT})$, $D_2 = (I_{N_iN_j} \otimes \iota_T)$, and $D_3 = (I_{N_i} \otimes \iota_{N_j} \otimes I_T)$. We assume, that α_{jt} effects enter the model as fixed, related to appropriate dummy variables,

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

and that $u_{ijt} = \mu_{ij} + \upsilon_{it} + \varepsilon_{ijt}$ are the random error components. To estimate this model specification, we have to follow a two step procedure. First, to get rid of the fixed effects, we define a projection orthogonal to the space of α_{jt}. Then, on this transformed model, we perform FGLS. The resulting estimator is not too complicated, and although restricted x_{jt} regressors can not be estimated from (35), $\hat{\beta}_{Mixed}$ is identified and consistent for the rest of the variables. This is a considerable improvement relative to the FGLS estimation of model (2), when N_j and T are both small, as in such cases, as shown in the asymptotic results of Section 3, the lack of consistency when estimating the variance components carries over to the model parameters'. The projection needed to eliminate α_{jt} is

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

Notice, that the resulting transformed (35),

$$\tilde{y}_{ijt} = \tilde{x}_{ijt}\beta + \tilde{u}_{ijt}\,, \tag{36}$$

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

$$\begin{aligned}\Omega = E(\tilde{u}\tilde{u}') = E(M_{D_1}uu'M_{D_1}) &= M_{D_1}D_2D_2'M_{D_1}\sigma_\mu^2 + M_{D_1}D_3D_3'M_{D_1}\sigma_\upsilon^2 + M_{D_1}\sigma_\varepsilon^2 \\ &= ((I_{N_i} - \bar{J}_{N_i}) \otimes I_{N_j} \otimes \bar{J}_T)T\sigma_\mu^2 + ((I_{N_i} - \bar{J}_{N_i}) \otimes \bar{J}_{N_j} \otimes I_T)N_j\sigma_\upsilon^2 \\ &\quad + ((I_{N_i} - \bar{J}_{N_i}) \otimes I_{N_jT})\sigma_\varepsilon^2\,,\end{aligned}$$

while its inverse can be derived using the trick introduced in Section 2 (the substitution $I_{N_i} = Q_{N_i} + \bar{J}_{N_i}$), giving

$$\begin{aligned}\Omega^{-1}\sigma_\varepsilon^2 = \left[I_{N_iN_jT} - (\bar{J}_{N_i} \otimes I_{N_jT})\right] &- (1-\theta_1)\left[(I_{N_i} \otimes \bar{J}_{N_j} \otimes I_T) - (\bar{J}_{N_iN_j} \otimes I_T)\right] \\ &- (1-\theta_2)\left[(I_{N_iN_j} \otimes \bar{J}_T) - (\bar{J}_{N_i} \otimes I_{N_j} \otimes \bar{J}_T)\right] \\ &+ (1-\theta_1-\theta_2+\theta_3)\left[(I_{N_i} \otimes \bar{J}_{N_jT}) - \bar{J}_{N_iN_jT}\right]\end{aligned}$$

with

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

Finally, the mixed effects estimation of (35) leads to the FGLS estimation of (36).The only step remaining is to estimate the variance components. In particular,

$$\begin{aligned}\hat{\sigma}_\varepsilon^2 &= \frac{1}{(N_i-1)(N_j-1)(T-1)}\sum_{ijt}(\tilde{\hat{u}}_{ijt}^a)^2 \\ \hat{\sigma}_\mu^2 &= \frac{1}{(N_i-1)(N_j-1)T}\sum_{ijt}(\tilde{\hat{u}}_{ijt}^b)^2 - \hat{\sigma}_\varepsilon^2 \\ \hat{\sigma}_\upsilon^2 &= \frac{1}{(N_i-1)N_j(T-1)}\sum_{ijt}(\tilde{\hat{u}}_{ijt}^c)^2 - \hat{\sigma}_\varepsilon^2\end{aligned}$$

where $\hat{u}_{ijt}$ are the OLS residuals, and now

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

Other types of mixed models can dealt with in a similar way.

6. Empirics

In this section we illustrate through an empirical application how the models and estimators introduced in this paper fare vis–a–vis the within (fixed effects) approach[7] and OLS. Although here we focus mostly on the general performance of the models, we also hope that the estimation outcome itself will be insightful in some ways.

We estimate a general gravity-type model, using different kinds of effects to formalise heterogeneity, relying on the *Konya et al.* [2013] specification. Our dataset involves 182 trading countries worldwide, observed annually over 53 years (for the period 1960-2012), with over one million ijt-type observations.[8] The panel is highly unbalanced, as around 25% of the observations are missing (relative to a complete, fully balanced data). Let us note here, that even for the most simple model (12), if the μ_{ij} interaction effects were not considered random, but fixed, as it has been the practice in most applied studies, it would mean the explicit or implicit estimation of $N_i N_j$ parameters, in this case about $182 \times 182 = 33,124$. This would look very much like a textbook over-specification case.

The model specification we use follows that of *Rose* [2004], and *Konya et al.* [2013], with the central difference that the individual heterogeneity is treated as random, rather than fixed:

$$\begin{aligned}\log(\text{RT})_{ijt} &= \beta_0 + \beta_1 \text{BOTHIN}_{ijt} + \beta_2 \log(\text{rGDP})_{it} + \beta_3 \log(\text{rGDP})_{jt} + \beta_4 \log\left(\frac{\text{rGDP}}{\text{POP}}\right)_{it} \\ &+ \beta_5 \log\left(\frac{\text{rGDP}}{\text{POP}}\right)_{jt} + \beta_6 \log(\text{DIST})_{ij} + \beta_7 \log(\text{LAND})_i + \beta_8 \log(\text{LAND})_j \\ &+ \beta_9 \text{CLANG}_{ij} + \beta_{10}\text{CBORD}_{ij} + \beta_{11}\text{LLOCK}_i + \beta_{12}\text{LLOCK}_j + \beta_{13}\text{ISAND}_i \\ &+ \beta_{14}\text{ISLAND}_j + \beta_{15}\text{EVCOL}_{ij} + \beta_{16}\text{COMCOL}_{ij} + \beta_{17}\text{MUN}_{ijt} + \beta_{18}\text{TA}_{ijt} + u_{ijt}\end{aligned} \tag{37}$$

[7] See *Balazsi et al* [2015].

[8] Raw net import-export data were collected from IMF's Direction of Trade Statistics Yearbook, and were deflated to 2000 USD using US CPI from IMF's International Financial Statistics Yearbook. Population and GDP measures were obtained from the World Bank's World Development Indicators. Other country- and country-pair specific data was collected from World Trade Organization, CIA's Factbook, and Wikipedia.

where RT is real trade activity (export or import) measured uniformly in 2000 US $, BOTHIN is the dummy variable, taking 1, if country i and j are both GATT/WTO members at time t; rGDP is a country's GDP in real terms at time t; rGDP/POP is the real GDP per capita; DIST is the great circle distance in miles; LAND is land areas of the country; CLANG is a dummy variable taking 1, if the country-pair has common language; CBORD is a dummy taking 1, if the country-pair has common border; LLOCK and ISALND is 1 if the country is landlocked or an island respectively; EVCOL and COMCOL are dummies taking 1, if the country has ever been colonized or if there is common colonizer, respectively; MUNI takes 1 if the country-pair is part of a monetary union at time t; TA is a dummy for existing trade agreement; and finally, u is some random effects structure corresponding to models (2) to (12).

Random Effects estimation results are collected in Table 6. We used exports as an indicator of real trade activity (as dependent variable), results for imports are very similar. As seen, all control variables are identified, and although the parameter estimates vary somewhat across model specifications, they all lead to same qualitative outcomes. That is, being a GATT/WTO member raises trade flows with around 0.5-0.9 log points. GDP also has a strong positive effects on export: a 1% jump in the former amounts to a 1-2% increase in trade activity, whereas per capita GDP estimates are mixed and much smaller in magnitude. The sign of the coefficients are what as one would expect: trade activity mitigates with distance and land area, but grows in "commons": language, border, monetary- and trade union, and colonizer. In this sense, we provided some evidence (in accordance with the findings of *Konya et al.* [2013]), supporting the strong positive effects of GATT/WTO membership on real trade activity.

Table 7 collects OLS estimates of the above models, which naturally gives the same parameter estimates for all models, but different standard errors, depending on the structure of the disturbance terms. As one should expect, the parameter estimates and the overall R^2 stand close to the random effect estimates, but the standard errors are significantly higher now, reflecting the loss of efficiency from not using the extra information about the error components structure for the estimation. Nevertheless, OLS gives a decent first estimate of the random effects models, and while not being optimal, it can be reached at very low computational cost.

Finally, Table 8 contains within estimates of models (2) to (12), that is, when we treat heterogeneity as fixed effects. Estimation procedures are detailed in *Balazsi et al.* [2015]. As can be seen from the table, several variables are not identified under within estimation, simply because some variables are fixed in some dimensions and are eliminated during the orthogonal projection of the within transformation.

For example, the fixed effects representation of model (12) consists γ_{ij} fixed effects, whereas, dist_{ij} is fixed in a similar manner. When we eliminate γ_{ij}, we also clear the latter. This is even more visible in case of the fixed effects version of the all-encompassing model (2), where all variables are fixed in one way or an other, and so are eliminated. Comparing the fixed effects to random effects estimates (that is Table 8 to Table 6), the parameter estimates fall very close to each other. This is appealing, as under random effects specification, all variables are identified, and can in fact be estimated, as opposed to the serious identification problems of the fixed effects models and the within transformation.

7. Conclusion

For large panel data sets, when observations can be considered as samples from an underlying population, random effects specifications seems to be appropriate to formalized the heterogeneity of the data. The most relevant three-dimensional three-way error components models and their FGLS estimators were presented in this paper. As seen, the resulting asymptotic requirements for the consistency of the FGLS are unusual. In fact, as here the data can grow in three directions at the same time, partial asymptotics (when not all dimensions grow infinitely) may be sufficient for consistency. Interestingly, as in the 2D case, for some error components specifications, consistency itself also carries a convergence property, namely the convergence of the FGLS estimator to the Within one. This is utterly important, as under Within estimation, unlike in the 2D case, the parameters of some fixed regressors are unidentified, which is in fact carried over to the FGLS as well. To go around this problem, we shown how the simple OLS can be used (of course, bearing the price of inefficiency), wherever this identification problem arises. Empirical investigation on a World trade data set shows, precisely, that the loss of information in case of OLS increases standard errors only slightly, and in fact leads to very similar parameter estimates than FGLS. There are empirically relevant cases, when the dimensionality of some indexes is small, meaning that this effect may be better formalized as a fixed one. The resulting mixed effects model, is also analyzed in the paper. The main results of the paper are also extended to treat incomplete, unbalanced data, and shown how to generalize them to higher dimensions.

References

Scott L. Baier and Jeffrey H. Bergstrand [2007]: Do Free Trade Agreements Actually Increase Members International Trade? *Journal of International Economics*, 71, 72-95.

Laszlo Balazsi, Laszlo Matyas and Tom Wansbeek [2015]: The Estimation of Multi-dimensional Fixed Effects Panel Data Models, *CEU Economics Working Papers*

Badi H. Baltagi, Peter Egger and Michael Pfaffermayr [2003]: A generalized Design for Bilateral Trade Flow Models, *Economic Letters*, 80, 391-397.

Badi H. Baltagi, Laszlo Matyas and Patrick Sevestre [2008]: *Error Components Models*, in Matyas and Sevetre (eds.) The Econometrics of Panel Data, Springer Verlag, 2008, 49-88.

Peter A. G. van Bergeijk and Steven Brakman, Rijksuniversiteit Groningen(eds) [2010]: *The Gravity Model in International Trade; Advances and Applications*, Cambridge University Press. **Paolo Bientinesi, Brian Gunter and Robert A. van de Geijn** [2000]: Families of Algorithms Related to the Inversion of a Symmetric Positive Definite Matrix, mimeo.

Anthony S. Bryk and Stephen W. Raudenbush [1992]: *Hierarchical Linear Models: Applications and Data Analysis Methods (First Edition)*, Newbury Park, CA: Sage Publications.

Richard Baldwin and Daria Taglioni [2006]: Gravity for Dummies and Dummies for the Gravity Equations, NBER Working Paper 12516.

Peter Ebbes, Ulf Bockenholt and Michel Wedel [2004]: Regressor and Random-effects Dependencies in Multilevel Models, *Statistica Neerlandica*, 58(2), 161178.

Peter Egger and Michael Pfaffermayr [2003]: The Proper Econometric Specification of the Gravity Equation: 3-Way Model with Bilateral Interaction Effects, *Empirical Economics*, 28, 571-580.

Robert A. van de Geijn (ed.) [2008]: A Systematic Approach to Matrix Computations, The University of Texas at Austin, manuscript.

Andrew Gelman [2006]: Multilevel (Hierarchical) Modeling: What It Can and Cannot Do, *Technometrics*, 432-435.

Harvey Goldstein [1995]: *Multilevel Statistical Models*, 2nd Edition, John Wiley & Sons, Ltd, Chichester, UK.

Michael Healy [1968]: Algorithm AS 7: Inversion of a Positive Semi-Definite Symmetric Matrix, *Applied Statistics*, 17(2), 198-199.

Cecilia Hornok [2011]: Gravity or Dummies? The Limits of Identification in Gravity Estimations, *CeFiG Working Papers* No. 15. (September 2011).

Olaf Hubler [2006]: Multilevel and Nonlinear Panel Data Models, *Allgemeines Statistisches Archiv*, 90, 121-136.

Laszlo Konya and Laszlo Matyas and Mark N. Harris [2013]: Modelling Trade Relationships: The Random Effects vs. Fixed Effects Approach, *CEU Economics Working Papers???*.

Nan M. Laird and James H. Ware [1982]: Random-Effects Models for Longitudinal Data, *Biometrics*, 38(4), 963-974.

Jan De Leeuw and Ita Kreft [1986]: Random Coefficient Models for Multilevel Analysis, *Journal of Educational and Behavioral Statistics*, 11(1), 57-85.

Laszlo Matyas [1998]: Proper Econometric Specification of the Gravity Model, *The World Economy*, 20, pp. 363-369.

Brent Moulton [1990]: An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Units, *The Review of Economics and Statistics*, 72, 334-338.

Radhakrishna Rao and Jurgen Kleffe [1980]: Estimation of Variance Components, in P.R. Krishnaiah (ed.) *Handbook of Statistics 1*, North-Holland Publisher, pp. 1-40.

Andrew K. Rose [2004]: Do We Really Know That the WTO Increases Trade? *American Economic Review*, 94, 98-114.

Simone Juhasz Silva and Douglas Nelson [2012]: Does Aid Cause Trade? Evidence from an Asymmetric Gravity Model, *The World Economy*, 545-577.

Tom Wansbeek and Arie Kapteyn [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.

Tom Wansbeek and Arie Kapteyn [1989]: Estimation of the Error-Components Model with Incomplete Panels , *Journal of Econometrics*, 41, 341-361.

Appendix 1

Example for normalizing with 1: Model (12), $T \to \infty$.

$$\operatorname*{plim}_{T\to\infty} V(\hat{\beta}_{OLS}) = \operatorname*{plim}_{T\to\infty} (X'X)^{-1}X'\Omega X(X'X)^{-1} = \operatorname*{plim}_{T\to\infty} \left(\frac{X'X}{T}\right)^{-1} \frac{X'\Omega X}{T^2} \left(\frac{X'X}{T}\right)^{-1}$$

We assume, that $\operatorname*{plim}_{T\to\infty} X'X/T = Q_{XX}$ is a finite, positive definite matrix, and further, we use that $\Omega = \sigma_\varepsilon^2 I_{N_i N_j T} + \sigma_\mu^2 (I_{N_i N_j} \otimes J_T)$. With this,

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

where we know, that $\operatorname*{plim}_{T\to\infty} \frac{\sigma_\varepsilon^2 X'X}{T^2} = 0$, and we assume, that

$$\operatorname*{plim}_{T\to\infty} \frac{\sigma_\mu^2 X'(I_{N_i N_j} \otimes J_T)X}{T^2} = Q_{XBX}$$

is a finite, positive definite matrix. Then the variance is finite, and takes the form

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

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

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

which grows at $O(T)$ because of $\frac{X'\Omega X}{T}$. We have to correct for it with the $1/\sqrt{T}$ factor, leading to the overall normalization factor $\sqrt{T}/\sqrt{T} = 1$. The reasoning is the similar for all other cases and other models.

Example for normalizing with $\frac{\sqrt{N_i N_j}}{A}$: Model (2), $N_i, N_j \to \infty$.

Using the standard $\sqrt{N_i N_j}$ normalization factor gives

$$\begin{aligned}\operatorname*{plim}_{N_i,N_j\to\infty} V(\sqrt{N_i N_j}\hat{\beta}_{OLS}) &= \operatorname*{plim}_{N_i,N_j\to\infty} N_i N_j \cdot (X'X)^{-1}X'\Omega X(X'X)^{-1} \\ &= \operatorname*{plim}_{N_i,N_j\to\infty} \left(\frac{X'X}{N_i N_j}\right)^{-1} \frac{X'\Omega X}{N_i N_j} \left(\frac{X'X}{N_i N_j}\right)^{-1} \\ &= Q_{XX}^{-1} \cdot \operatorname*{plim}_{N_i,N_j\to\infty} \frac{X'\Omega X}{N_i N_j} \cdot Q_{XX}^{-1} \,,\end{aligned}$$

where we assumed, that $\operatorname*{plim}_{N_i,N_j\to\infty} X'X/N_i N_j = Q_{XX}$, is a positive definite, finite matrix. Further, we use, that

$$\Omega = \sigma_\varepsilon^2 I_{N_i N_j T} + \sigma_\mu (I_{N_i N_j} \otimes J_T) + \sigma_\upsilon^2 (I_{N_i} \otimes J_{N_j} \otimes I_T) + \sigma_\zeta^2 (J_{N_i} \otimes I_{N_j T}) \,.$$

Observe, that

$$\begin{aligned}\operatorname*{plim}_{N_i,N_j\to\infty} \frac{X'\Omega X}{N_i N_j} = \operatorname*{plim}_{N_i,N_j\to\infty} \frac{\sigma_\varepsilon^2 X'X}{N_i N_j} + \operatorname*{plim}_{N_i,N_j\to\infty} \frac{\sigma_\mu X'(I_{N_i N_j} \otimes J_T)X}{N_i N_j} \\ + \operatorname*{plim}_{N_i,N_j\to\infty} \frac{\sigma_\upsilon^2 X'(I_{N_i} \otimes J_{N_j} \otimes I_T)X}{N_i N_j} + \operatorname*{plim}_{N_i,N_j\to\infty} \frac{\sigma_\zeta^2 X'(J_{N_i} \otimes I_{N_j T})X}{N_i N_j}\end{aligned} \tag{38}$$

is an expression where the first two terms are finite, but the third grows with $O(N_j)$ (because of J_{N_j}), and the last with $O(N_i)$ (because of J_{N_i}), which in turns yields unbounded variance of $\hat{\beta}_{OLS}$. To obtain a finite variance, we have to normalize the variance additionally with either $1/\sqrt{N_i}$ or $1/\sqrt{N_j}$, depending on which grows faster. Let us assume, without loss of generality, that N_i grows at a higher rate ($A = N_i$). In this way, the effective normalization factor is $\frac{\sqrt{N_i N_j}}{\sqrt{A}} = \frac{\sqrt{N_i N_j}}{\sqrt{N_i}} = \sqrt{N_j}$, under which the first three plim terms in (38) are zero, but the forth is finite:

$$\operatorname*{plim}_{N_i,N_j\to\infty} \frac{\sigma_\zeta^2 X'(J_{N_i} \otimes I_{N_j T})X}{N_i^2 N_j} = Q_{XBX} \,,$$

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

Appendix 2

Proof of formula (17).

Let us make the proof only for model (2) (so formula (17)), the rest is just direct application of the derivation below. The outline of the proof is based on *Wansbeek and Kapteyn* [1989].

First, notice, that using the Woodbury matrix identity,

$$
\begin{aligned}
(P^a)^{-1} &= \left(I - D_1(D_1'D_1 + I\frac{\sigma_\varepsilon^2}{\sigma_\mu^2})^{-1}D_1\right)^{-1} = I + D_1\left(D_1'D_1 + I\frac{\sigma_\varepsilon^2}{\sigma_\mu^2} - D_1'D_1\right)^{-1} D_1' \\
&= I + \frac{\sigma_\mu^2}{\sigma_\varepsilon^2} D_1 D_1'
\end{aligned}
$$

Second, using that

$$
D_2'P^aD_2 = D_2'D_2 - D_2'D_1(R^a)^{-1}D_1'D_2 = R^b - \frac{\sigma_\varepsilon^2}{\sigma_\upsilon^2}I \quad \text{gives} \quad R^b - D_2'P^aD_2 = \frac{\sigma_\varepsilon^2}{\sigma_\upsilon^2}I\,.
$$

Using the Woodbury matrix identity for the second time,

$$
\begin{aligned}
(P^b)^{-1} &= \left(P^a - P^aD_2(R^b)^{-1}D_2'P^a\right)^{-1} \\
&= (P^a)^{-1} + (P^a)^{-1}P^aD_2\left(R^b - D_2'P^a(P^a)^{-1}P^aD_2\right)^{-1} D_2'P^a(P^a)^{-1} \\
&= (P^a)^{-1} + D_2\left(R^b - D_2'P^aD_2\right)^{-1} D_2' = (P^a)^{-1} + D_2\left(\frac{\sigma_\varepsilon^2}{\sigma_\upsilon^2}I\right)^{-1} D_2' \\
&= I + \frac{\sigma_\mu^2}{\sigma_\varepsilon^2}D_1D_1' + \frac{\sigma_\upsilon^2}{\sigma_\varepsilon^2}D_2D_2'\,.
\end{aligned}
$$

Now we are almost there, we only have to repeat the last step one more time. That is,

$$
D_3'P^bD_3 = D_3'D_3 - D_3'D_2(R^b)^{-1}D_2'D_3 = R^c - \frac{\sigma_\varepsilon^2}{\sigma_\zeta^2}I \quad \text{gives} \quad R^c - D_3'P^bD_3 = \frac{\sigma_\varepsilon^2}{\sigma_\zeta^2}I\,.
$$

again, and so

$$
\begin{aligned}
\left(\Omega^{-1}\sigma_\varepsilon^2\right)^{-1} &= \left(P^b - P^bD_3(R^c)^{-1}D_3'P^b\right)^{-1} \\
&= (P^b)^{-1} + (P^b)^{-1}P^bD_3\left(R^c - D_3'P^b(P^b)^{-1}P^bD_3\right)^{-1} D_3'P^b(P^b)^{-1} \\
&= (P^b)^{-1} + D_3\left(R^c - D_3'P^bD_3\right)^{-1} D_3' = (P^b)^{-1} + D_3\left(\frac{\sigma_\varepsilon^2}{\sigma_\zeta^2}I\right)^{-1} D_3' \\
&= I + \frac{\sigma_\mu^2}{\sigma_\varepsilon^2}D_1D_1' + \frac{\sigma_\upsilon^2}{\sigma_\varepsilon^2}D_2D_2' + \frac{\sigma_\zeta^2}{\sigma_\varepsilon^2}D_3D_3' = \Omega\sigma_\varepsilon^{-2}\,.
\end{aligned}
$$

Appendix 3

The inverse of (32), and the estimation of the variance components

$$
\begin{aligned}
\sigma_\varepsilon^2\Omega^{-1} =& I_{N_iN_jN_sT} - (1-\theta_{20})(J_{N_i}\otimes I_{N_jN_sT}) - (1-\theta_{21})(I_{N_i}\otimes J_{N_j}\otimes I_{N_sT}) \\
&- (1-\theta_{22})(I_{N_iN_j}\otimes J_{N_s}\otimes I_T) - (1-\theta_{23})(I_{N_iN_jN_s}\otimes J_T) \\
&+ (1-\theta_{24})(J_{N_iN_j}\otimes I_{N_sT}) + (1-\theta_{25})(J_{N_i}\otimes I_{N_j}\otimes J_{N_s}\otimes I_T) \\
&+ (1-\theta_{26})(J_{N_i}\otimes I_{N_jN_s}\otimes J_T) + (1-\theta_{27})(I_{N_i}\otimes J_{N_jN_s}\otimes I_T) \\
&+ (1-\theta_{28})(I_{N_i}\otimes J_{N_j}\otimes I_{N_s}\otimes J_T) + (1-\theta_{29})(I_{N_iN_j}\otimes J_{N_sT}) \\
&- (1-\theta_{30})(J_{N_iN_jN_s}\otimes I_T) - (1-\theta_{31})(J_{N_iN_j}\otimes I_{N_s}\otimes J_T) \\
&- (1-\theta_{32})(J_{N_i}\otimes I_{N_j}\otimes J_{N_sT}) - (1-\theta_{33})(I_{N_i}\otimes J_{N_jN_sT}) \\
&+ (1-\theta_{34})J_{N_iN_jN_sT}
\end{aligned}
$$

with

$$
\begin{aligned}
&\theta_{20} = \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + N_i\sigma_\zeta} \quad \theta_{21} = \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + N_j\sigma_\upsilon} \quad \theta_{22} = \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + N_s\sigma_\lambda} \quad \theta_{23} = \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T\sigma_\mu} \\
&\theta_{24} = \theta_{20} + \theta_{21} - \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + N_i\sigma_\zeta^2 + N_j\sigma_\upsilon^2} \quad \theta_{25} = \theta_{20} + \theta_{22} - \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + N_i\sigma_\zeta^2 + N_s\sigma_\lambda^2} \\
&\theta_{26} = \theta_{20} + \theta_{23} - \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + N_i\sigma_\zeta^2 + T\sigma_\mu^2} \quad \theta_{27} = \theta_{21} + \theta_{22} - \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + N_j\sigma_\upsilon^2 + N_s\sigma_\lambda^2} \\
&\theta_{28} = \theta_{21} + \theta_{23} - \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + N_j\sigma_\upsilon^2 + T\sigma_\mu^2} \quad \theta_{29} = \theta_{22} + \theta_{23} - \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + N_s\sigma_\lambda^2 + T\sigma_\mu^2} \\
&\theta_{30} = \theta_{24} + \theta_{25} + \theta_{27} - \theta_{20} - \theta_{21} - \theta_{22} + \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + N_i\sigma_\zeta^2 + N_j\sigma_\upsilon^2 + N_s\sigma_\lambda^2} \\
&\theta_{31} = \theta_{24} + \theta_{26} + \theta_{28} - \theta_{20} - \theta_{21} - \theta_{23} + \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + N_i\sigma_\zeta^2 + N_j\sigma_\upsilon^2 + T\sigma_\mu^2} \\
&\theta_{32} = \theta_{25} + \theta_{26} + \theta_{29} - \theta_{20} - \theta_{22} - \theta_{23} + \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + N_i\sigma_\zeta^2 + N_s\sigma_\lambda^2 + T\sigma_\mu^2} \\
&\theta_{33} = \theta_{27} + \theta_{28} + \theta_{29} - \theta_{21} - \theta_{22} - \theta_{23} + \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + N_j\sigma_\upsilon^2 + N_s\sigma_\lambda^2 + T\sigma_\mu^2} \\
&\theta_{34} = \theta_{20} + \theta_{21} + \theta_{22} + \theta_{23} - \theta_{24} - \theta_{25} - \theta_{26} - \theta_{27} - \theta_{28} - \theta_{29} \\
&\qquad + \theta_{30} + \theta_{31} + \theta_{32} + \theta_{33} - \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + N_i\sigma_\zeta^2 + N_j\sigma_\upsilon^2 + N_s\sigma_\lambda^2 + T\sigma_\mu^2} \,.
\end{aligned}
$$

For the estimation of the variance components in case of complete data we get:

$$\hat{\sigma}^2_\varepsilon = \frac{1}{(N_i-1)(N_j-1)(N_s-1)(T-1)} \sum_{ijst} \tilde{\hat{u}}^2_{ijst}$$

$$\hat{\sigma}^2_\mu = \frac{1}{(N_i-1)(N_j-1)(N_s-1)T} \sum_{ijst} (\tilde{\hat{u}}^a_{ijst})^2 - \hat{\sigma}^2_\varepsilon$$

$$\hat{\sigma}^2_\upsilon = \frac{1}{(N_i-1)N_j(N_s-1)(T-1)} \sum_{ijst} (\tilde{\hat{u}}^b_{ijst})^2 - \hat{\sigma}^2_\varepsilon$$

$$\hat{\sigma}^2_\zeta = \frac{1}{N_i(N_j-1)(N_s-1)(T-1)} \sum_{ijst} (\tilde{\hat{u}}^c_{ijst})^2 - \hat{\sigma}^2_\varepsilon$$

$$\hat{\sigma}^2_\lambda = \frac{1}{(N_i-1)(N_j-1)N_s(T-1)} \sum_{ijst} (\tilde{\hat{u}}^d_{ijst})^2 - \hat{\sigma}^2_\varepsilon \,,$$

where, as before, $\hat{u}_{ijst}$ are the OLS residuals, and

$$\begin{aligned}
\tilde{u}_{ijst} =& u_{ijst} - \bar{u}_{ijs.} - \bar{u}_{ij.t} - \bar{u}_{i.st} - \bar{u}_{.jst} + \bar{u}_{ij..} + \bar{u}_{i.s.} + \bar{u}_{.js.} + \bar{u}_{i..t} + \bar{u}_{.j.t} + \bar{u}_{..st} \\
& - \bar{u}_{i...} - \bar{u}_{.j..} - \bar{u}_{..s.} - \bar{u}_{...t} + \bar{u}_{....} \\
\tilde{u}^a_{ijst} =& u_{ijst} - \bar{u}_{ij.t} - \bar{u}_{i.st} - \bar{u}_{.jst} + \bar{u}_{i..t} + \bar{u}_{.j.t} + \bar{u}_{..st} - \bar{u}_{...t} \\
\tilde{u}^b_{ijst} =& u_{ijst} - \bar{u}_{ijs.} - \bar{u}_{ij.t} - \bar{u}_{.jst} + \bar{u}_{ij..} + \bar{u}_{.js.} + \bar{u}_{.j.t} - \bar{u}_{.j..} \\
\tilde{u}^c_{ijst} =& u_{ijst} - \bar{u}_{ijs.} - \bar{u}_{ij.t} - \bar{u}_{i.st} + \bar{u}_{ij..} + \bar{u}_{i.s.} + \bar{u}_{i..t} - \bar{u}_{i...} \\
\tilde{u}^d_{ijst} =& u_{ijst} - \bar{u}_{ijs.} - \bar{u}_{i.st} - \bar{u}_{.jst} + \bar{u}_{i.s.} + \bar{u}_{.js.} + \bar{u}_{..st} - \bar{u}_{..s.} \,.
\end{aligned}$$

For the estimation of the variance components in case of incomplete data we get

$$\begin{aligned}
\hat{\sigma}^2_\mu &= \frac{1}{\sum_{ijs}|T_{ijs}|} \sum_{ijst} \hat{u}^2_{ijst} - \frac{1}{\tilde{n}_{ijs}} \sum_{ijs} \frac{1}{|T_{ijs}|-1} \sum_t (\tilde{\hat{u}}^a_{ijst})^2 \\
\hat{\sigma}^2_\upsilon &= \frac{1}{\sum_{ijs}|T_{ijs}|} \sum_{ijst} \hat{u}^2_{ijst} - \frac{1}{\tilde{n}_{ist}} \sum_{ist} \frac{1}{n_{ist}-1} \sum_j (\tilde{\hat{u}}^b_{ijst})^2 \\
\hat{\sigma}^2_\zeta &= \frac{1}{\sum_{ijs}|T_{ijs}|} \sum_{ijst} \hat{u}^2_{ijst} - \frac{1}{\tilde{n}_{jst}} \sum_{jt} \frac{1}{n_{jst}-1} \sum_i (\tilde{\hat{u}}^c_{ijst})^2 \\
\hat{\sigma}^2_\lambda &= \frac{1}{\sum_{ijs}|T_{ijs}|} \sum_{ijst} \hat{u}^2_{ijst} - \frac{1}{\tilde{n}_{ijt}} \sum_{ijt} \frac{1}{n_{ijt}-1} \sum_s (\tilde{\hat{u}}^d_{ijst})^2 \\
\hat{\sigma}^2_\varepsilon &= \frac{1}{\sum_{ijs}|T_{ijs}|} \sum_{ijst} \hat{u}^2_{ijst} - \hat{\sigma}^2_\mu - \hat{\sigma}^2_\upsilon - \hat{\sigma}^2_\zeta - \hat{\sigma}^2_\lambda \,,
\end{aligned} \tag{39}$$

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

$$\tilde{u}^a_{ijst} = u_{ijst} - \frac{1}{|T_{ijs}|}\sum_t u_{ijst}\,, \quad \tilde{u}^b_{ijst} = u_{ijst} - \frac{1}{n_{ist}}\sum_j u_{ijst}\,,$$

$$\tilde{u}^c_{ijst} = u_{ijst} - \frac{1}{n_{jst}}\sum_i u_{ijst}\,, \quad \tilde{u}^d_{ijst} = u_{ijst} - \frac{1}{n_{ijt}}\sum_s u_{ijst}\,.$$

Further, $|T_{ijs}|$, $\tilde{n}_{ist}$, $\tilde{n}_{jst}$, and $\tilde{n}_{ijt}$ denote the total number of observations for a given (ijs), (ist), (jst), and (ijt) pair respectively, and finally, $\tilde{n}_{ijs}$, $\tilde{n}_{ist}$, $\tilde{n}_{jst}$, and $\tilde{n}_{ijt}$ are the total number of unique (ijs), (ist), (jst), and (ijt) observations in the data.

Table 6: Random Effects Estimates of Different Models

Variable	Model (2)			Model (4)			Model (6)			Model (6b)			Model (8)			Model (10)			Model (12)		
	$\hat{\beta}$	Se($\hat{\beta}$)	T-value	$\hat{\beta}$	Se($\hat{\beta}$)	T-value	$\hat{\beta}$	Se($\hat{\beta}$)	T-value	$\hat{\beta}$	Se($\hat{\beta}$)	T-value	$\hat{\beta}$	Se($\hat{\beta}$)	T-value	$\hat{\beta}$	Se($\hat{\beta}$)	T-value	$\hat{\beta}$	Se($\hat{\beta}$)	T-value
C	-54.166	0.512	105.877	-54.434	0.319	170.750	-54.463	0.183	297.086	-52.749	0.283	186.417	-49.146	0.395	124.419	-40.040	0.804	49.773	-59.473	0.351	169.456
BOTHIN	0.510	0.062	8.263	0.581	0.031	18.526	0.927	0.018	50.845	0.456	0.018	25.143	0.907	0.049	18.367	0.783	0.050	15.762	0.553	0.049	11.325
LNRGDP1	1.987	0.022	91.738	2.043	0.018	116.053	1.996	0.004	453.868	1.993	0.017	114.042	1.808	0.011	158.191	1.391	0.021	64.875	2.136	0.010	224.823
LNRGDP2	1.623	0.017	95.470	1.577	0.010	155.777	1.538	0.010	152.741	1.562	0.004	368.386	1.498	0.012	128.997	1.513	0.018	84.500	1.846	0.010	191.176
LNRGDPPOP1	0.045	0.027	1.685	0.059	0.022	2.624	0.096	0.005	17.811	0.070	0.022	3.130	-0.098	0.008	11.833	-0.106	0.011	9.854	-0.191	0.008	23.806
LNRGDPPOP2	-0.145	0.019	7.538	-0.022	0.013	1.675	-0.021	0.013	1.592	-0.019	0.005	3.562	-0.160	0.009	17.894	-0.174	0.011	15.318	-0.264	0.009	30.648
LNDIST	-2.084	0.036	57.903	-2.242	0.010	229.589	-2.027	0.010	209.880	-2.255	0.009	243.552	-1.958	0.034	57.157	-2.252	0.011	213.058	-1.915	0.034	56.006
LNLAND1	-0.287	0.019	14.810	-0.262	0.015	17.595	-0.239	0.004	65.323	-0.259	0.015	17.416	-0.152	0.012	12.703	0.063	0.050	1.271	-0.310	0.012	26.707
LNLAND2	-0.156	0.015	10.191	-0.140	0.009	16.134	-0.171	0.009	19.656	-0.138	0.004	38.872	-0.128	0.012	10.601	-0.094	0.023	4.004	-0.301	0.012	25.712
CLANG	0.817	0.058	14.002	0.878	0.016	54.323	0.644	0.015	41.988	0.796	0.015	53.271	0.706	0.055	12.930	0.892	0.018	50.483	0.749	0.055	13.741
CBORD	1.180	0.173	6.821	1.139	0.042	27.328	1.387	0.044	31.385	1.248	0.042	29.534	1.224	0.173	7.065	1.164	0.045	26.122	1.324	0.173	7.655
LLOCK1	0.682	0.091	7.457	0.144	0.070	2.057	0.201	0.018	11.340	0.152	0.070	2.184	0.001	0.063	0.012	-0.690	0.277	2.493	0.350	0.062	5.627
LLOCK2	-0.705	0.072	9.755	-0.752	0.039	19.221	-0.666	0.039	16.879	-0.729	0.017	43.894	-0.694	0.062	11.179	-0.734	0.122	6.006	-0.308	0.061	5.017
ISLAND1	0.330	0.092	3.581	0.276	0.067	4.104	0.178	0.017	10.555	0.265	0.067	3.929	0.358	0.066	5.467	0.162	0.299	0.544	0.370	0.065	5.656
ISLAND2	0.434	0.075	5.762	0.250	0.039	6.350	0.243	0.040	6.113	0.256	0.016	15.710	0.529	0.066	8.012	0.556	0.134	4.136	0.577	0.066	8.749
EVCOL	0.861	0.277	3.106	1.157	0.070	16.498	1.498	0.070	21.486	1.063	0.071	14.892	1.496	0.269	5.565	1.335	0.075	17.803	1.010	0.268	3.763
COMCOL	0.954	0.081	11.800	1.238	0.021	59.226	1.151	0.021	55.223	1.076	0.020	53.653	0.552	0.078	7.090	1.270	0.023	56.090	0.892	0.077	11.528
MUNI	1.406	0.093	15.137	0.684	0.024	29.053	1.352	0.023	58.072	0.878	0.023	38.894	1.798	0.090	20.064	0.586	0.026	22.962	1.688	0.089	18.865
TA	1.226	0.068	18.123	1.047	0.017	59.945	1.015	0.017	58.097	0.846	0.017	50.396	1.321	0.065	20.204	1.078	0.019	57.030	1.142	0.065	17.496
R^2	0.765			0.795			0.798			0.788			0.712			0.661			0.797		

Estimation is done on the GATT/WTO World dataset, with LNREXPORT as dependent variable. Model (2): $\mu_{ij} + \upsilon_{it} + \zeta_{jt}$; Model (4): $\upsilon_{it} + \zeta_{jt}$; Model (6): ζ_{jt}; Model (6b): υ_{it}; Model (8): $\gamma_{ij} + \lambda_t$; Model (10): $\upsilon_i + \zeta_j + \lambda_t$; Model (12): γ_{ij}.

Table 7: OLS Estimates of Different Models

Variable		Model (2)		Model (4)		Model (6)		Model (6b)		Model (8)		Model (10)		Model (12)	
	$\hat{\beta}$	Se($\hat{\beta}$)	T-value	Se($\hat{\beta}$)	T-value	Se($\hat{\beta}$)	T-value	Se($\hat{\beta}$)	T-value	Se($\hat{\beta}$)	T-value	Se($\hat{\beta}$)	T-value	Se($\hat{\beta}$)	T-value
C	-54.365	0.584	93.143	0.353	154.193	0.197	275.613	0.316	172.000	0.220	246.718	1.334	40.746	0.480	113.147
BOTHIN	0.601	0.069	8.756	0.045	13.482	0.025	24.038	0.040	15.161	0.021	28.987	0.159	3.785	0.054	11.097
LNRGDP1	2.025	0.024	83.206	0.018	111.704	0.005	436.696	0.018	112.063	0.010	207.197	0.073	27.584	0.017	120.360
LNRGDP2	1.549	0.020	78.613	0.011	143.531	0.010	149.493	0.005	289.675	0.009	179.637	0.032	49.151	0.017	90.762
LNRGDPPOP1	0.076	0.030	2.562	0.023	3.307	0.005	13.883	0.023	3.309	0.016	4.852	0.088	0.869	0.020	3.890
LNRGDPPOP2	-0.031	0.024	1.321	0.013	2.364	0.013	2.369	0.006	5.571	0.017	1.805	0.040	0.796	0.021	1.532
LNDIST	-2.058	0.040	50.885	0.017	118.363	0.012	175.407	0.016	130.115	0.037	54.922	0.065	31.531	0.038	54.634
LNLAND1	-0.259	0.021	12.406	0.015	16.916	0.004	68.201	0.015	16.946	0.011	24.385	0.065	3.952	0.015	17.664
LNLAND2	-0.166	0.017	9.760	0.009	18.331	0.009	18.683	0.004	40.185	0.011	15.314	0.028	5.854	0.015	11.160
CLANG	0.588	0.066	8.914	0.031	18.719	0.019	30.400	0.029	20.510	0.059	10.021	0.126	4.653	0.060	9.835
CBORD	1.450	0.188	7.720	0.049	29.351	0.046	31.666	0.048	29.939	0.187	7.766	0.124	11.739	0.187	7.768
LLOCK1	0.233	0.099	2.350	0.072	3.226	0.018	12.828	0.072	3.228	0.061	3.831	0.305	0.763	0.070	3.315
LLOCK2	-0.642	0.079	8.140	0.041	15.705	0.041	15.778	0.018	35.521	0.063	10.264	0.132	4.848	0.070	9.209
ISLAND1	0.170	0.098	1.726	0.070	2.440	0.017	9.932	0.070	2.440	0.070	2.410	0.320	0.531	0.072	2.370
ISLAND2	0.238	0.081	2.935	0.041	5.816	0.041	5.805	0.017	13.850	0.071	3.362	0.144	1.648	0.072	3.300
EVCOL	1.266	0.315	4.011	0.129	9.825	0.071	17.890	0.129	9.819	0.295	4.293	0.518	2.443	0.297	4.267
COMCOL	1.013	0.090	11.301	0.037	27.727	0.025	40.790	0.034	30.219	0.083	12.230	0.142	7.145	0.084	12.017
MUNI	1.434	0.102	14.120	0.042	34.370	0.028	51.826	0.038	37.264	0.094	15.239	0.166	8.665	0.095	15.050
TA	0.818	0.075	10.908	0.030	27.201	0.021	39.431	0.028	29.679	0.070	11.703	0.115	7.109	0.071	11.563
R^2	0.804														

Estimation is done on the GATT/WTO World dataset, with LNREXPORT as dependent variable. Model (2): $\mu_{ij} + \upsilon_{it} + \zeta_{jt}$; Model (4): $\upsilon_{it} + \zeta_{jt}$; Model (6): ζ_{jt}; Model (6b): υ_{it}; Model (8): $\gamma_{ij} + \lambda_t$; Model (10): $\upsilon_i + \zeta_j + \lambda_t$; Model (12): γ_{ij}.

Table 8: Fixed Effects Estimates of Different Models

Variable	Model (2)			Model (4)			Model (6)			Model (6b)			Model (8)			Model (10)			Model (12)		
	$\hat{\beta}$	Se($\hat{\beta}$)	T-value	$\hat{\beta}$	Se($\hat{\beta}$)	T-value	$\hat{\beta}$	Se($\hat{\beta}$)	T-value	$\hat{\beta}$	Se($\hat{\beta}$)	T-value	$\hat{\beta}$	Se($\hat{\beta}$)	T-value	$\hat{\beta}$	Se($\hat{\beta}$)	T-value	$\hat{\beta}$	Se($\hat{\beta}$)	T-value
C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
BOTHIN	-	-	-	0.615	0.048	12.804	1.111	0.020	56.748	0.443	0.018	24.360	-	-	-	0.758	0.052	14.654	-	-	-
LNRGDP1	-	-	-	-	-	-	1.983	0.004	456.998	-	-	-	1.265	0.018	68.506	1.301	0.022	60.266	1.948	0.014	134.772
LNRGDP2	-	-	-	-	-	-	-	-	-	1.563	0.004	376.917	1.449	0.019	77.347	1.486	0.022	68.515	2.170	0.015	148.304
LNRGDPPOP1	-	-	-	-	-	-	0.100	0.005	19.043	-	-	-	-0.083	0.009	9.569	-0.100	0.010	9.739	-0.219	0.009	25.474
LNRGDPPOP2	-	-	-	-	-	-	-	-	-	-0.018	0.005	3.549	-0.180	0.009	19.007	-0.181	0.011	16.198	-0.338	0.009	36.303
LNDIST	-	-	-	-2.287	0.009	242.715	-2.019	0.010	209.944	-2.269	0.009	249.987	-	-	-	-2.256	0.010	224.459	-	-	-
LNLAND1	-	-	-	-	-	-	-0.231	0.004	64.209	-	-	-	-	-	-	-	-	-	-	-	-
LNLAND2	-	-	-	-	-	-	-	-	-	-0.137	0.003	39.389	-	-	-	-	-	-	-	-	-
CLANG	-	-	-	0.923	0.016	58.122	0.668	0.015	43.611	0.810	0.015	55.239	-	-	-	0.896	0.017	53.301	-	-	-
CBORD	-	-	-	0.992	0.040	25.208	1.368	0.043	31.516	1.234	0.041	29.916	-	-	-	1.161	0.042	27.471	-	-	-
LLOCK1	-	-	-	-	-	-	0.185	0.017	10.643	-	-	-	-	-	-	-	-	-	-	-	-
LLOCK2	-	-	-	-	-	-	-	-	-	-0.734	0.016	45.274	-	-	-	-	-	-	-	-	-
ISLAND1	-	-	-	-	-	-	0.184	0.017	11.139	-	-	-	-	-	-	-	-	-	-	-	-
ISLAND2	-	-	-	-	-	-	-	-	-	0.257	0.016	16.124	-	-	-	-	-	-	-	-	-
EVCOL	-	-	-	0.876	0.064	13.651	1.561	0.068	22.816	1.036	0.070	14.814	-	-	-	1.331	0.071	18.698	-	-	-
COMCOL	-	-	-	1.253	0.020	61.837	1.186	0.021	57.237	1.081	0.020	54.999	-	-	-	1.277	0.022	59.345	-	-	-
MUNI	-	-	-	0.575	0.023	25.120	1.331	0.023	57.556	0.837	0.022	37.846	-	-	-	0.576	0.024	23.759	-	-	-
TA	-	-	-	1.062	0.017	62.718	1.058	0.017	60.967	0.846	0.016	51.487	-	-	-	1.076	0.018	59.919	-	-	-
R^2	0			0.165			0.456			0.393			0.014			0.159			0.170		

Estimation is done on the GATT/WTO World dataset, with LNREXPORT as dependent variable. Model (2): $\mu_{ij} + \alpha_{it} + \alpha^*_{jt}$; Model (4): $\alpha_{it} + \alpha^*_{jt}$; Model (6): α^*_{jt}; Model (6b): α_{it}; Model (8): $\gamma_{ij} + \lambda_t$; Model (10): $\alpha_i + \gamma_j + \lambda_t$; Model (12): γ_{ij}.