The Econometrics of Multi-dimensional Panels

Chapter 2: Random Effects Models

version 2.2

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

Preliminary version, please do not quote

November 4, 2015

Introduction

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

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

This chapter follows in spirit the analysis of the two-way panels by Baltagi et al. (2008), that is, in Section 2.1 we introduce the most frequently used models in a three-dimensional (3D) panel data setup, Section 2.2 deals with the Feasible GLS estimation of these models, while Section 2.3 analyses the behaviour of this estimator for incomplete/unbalanced data. Section 2.4 generalizes the presented models to four and higher dimensional data sets, and extends the random effects approach toward a mixed effects framework. Finally, Section 2.5 deals with some testing issues.

2.1 Different Model Specifications

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

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

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

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

2.1.1 Various Heterogeneity Formulations

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

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

where $i = 1 \ldots N_i$, $j = 1 \ldots N_j$, 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 (2.1), 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,

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

The covariance matrix of such error components structure is simply

$$\Omega = E(uu') = \sigma^2_\mu(I_{N_iN_j}\otimes J_T)+\sigma^2_\upsilon(I_{N_i}\otimes J_{N_j}\otimes I_T)+\sigma^2_\zeta(J_{N_i}\otimes I_{N_jT})+\sigma^2_\varepsilon I_{N_iN_jT}\,, \tag{2.3}$$

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 (2.2). The model which only uses individual-time-varying effects reads as

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

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

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

A further restriction on the above model is

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

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

$$\Omega = \sigma^2_\zeta(J_{N_i} \otimes I_{N_jT}) + \sigma^2_\varepsilon I_{N_iN_jT}\,. \tag{2.7}$$

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

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

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

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

Now

$$\Omega = \sigma^2_\mu(I_{N_iN_j} \otimes J_T) + \sigma^2_\lambda(J_{N_iN_j} \otimes I_T) + \sigma^2_\varepsilon I_{N_iN_jT}\,. \tag{2.9}$$

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

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

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

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

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

Its covariance structure is

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

Lastly, the simplest model is

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

with

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

Note, that model (2.12) 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]This model has in fact been introduced in Matyas (1998), and before that, in Ghosh (1976).

2.1.2 Spectral Decomposition of the Covariance Matrices

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

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

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

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

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

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

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

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

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

$$\begin{aligned}\sigma_\varepsilon^2 \Omega^{-1} = \ & I_{N_iN_jT} - (1-\theta_1)(\bar{J}_{N_i} \otimes I_{N_jT}) - (1-\theta_2)(I_{N_i} \otimes \bar{J}_{N_j} \otimes I_T) \\ & - (1-\theta_3)(I_{N_iN_j} \otimes \bar{J}_T) + (1-\theta_1-\theta_2+\theta_4)(\bar{J}_{N_iN_j} \otimes I_T) \\ & + (1-\theta_1-\theta_3+\theta_5)(\bar{J}_{N_i} \otimes I_{N_j} \otimes \bar{J}_T) \\ & + (1-\theta_2-\theta_3+\theta_6)(I_{N_i} \otimes \bar{J}_{N_jT}) \\ & - (1-\theta_1-\theta_2-\theta_3+\theta_4+\theta_5+\theta_6-\theta_7)\bar{J}_{N_iN_jT} \, ,\end{aligned} \tag{2.15}$$

with

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

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

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

where, following the standard ANOVA notation, a bar over the variable means that mean of the variable was taken with respect to the missing indices. By using the OLS on these transformed variables, we get back the GLS estimator.

For other models, the job is essentially the same. For model (2.4),

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

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

$$\tilde{y}_{ijt} = y_{ijt} - (1-\sqrt{\theta_8})\bar{y}_{i.t} - (1-\sqrt{\theta_9})\bar{y}_{.jt} + (1-\sqrt{\theta_8}-\sqrt{\theta_9}+\sqrt{\theta_{10}})\bar{y}_{..t}\,,$$

with

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

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

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

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

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

For Model (2.8), it is

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

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

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

with

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

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

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

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

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

where

$$
\theta_{15} = \tfrac{\sigma_\varepsilon^2}{N_jT\sigma_\upsilon^2+\sigma_\varepsilon^2}\,, \quad \theta_{16} = \tfrac{\sigma_\varepsilon^2}{N_iT\sigma_\zeta^2+\sigma_\varepsilon^2}\,, \quad \theta_{17} = \tfrac{\sigma_\varepsilon^2}{N_iN_j\sigma_\lambda^2+\sigma_\varepsilon^2}\,, \quad \text{and}
$$

$$
\theta_{18} = \tfrac{\sigma_\varepsilon^2}{N_jT\sigma_\upsilon^2+N_iT\sigma_\zeta^2+N_iN_j\sigma_\lambda^2+\sigma_\varepsilon^2}\,.
$$

For model (2.12), the inversion gives

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

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

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

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

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

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

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

When the number of observations grow in one or more dimensions, it can be interesting to find the limits of the θ_k weights. It is easy to see, that if all N_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 (2.2), 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 2.2 collects the asymptotic conditions, when the models' (F)GLS estimator is converging to a Within estimator.

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

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

2.2 FGLS Estimation

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

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

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 (2.2) (see for details Balazsi et al. (2015)) is

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

Note, that this transformation corresponds to the projection matrix

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

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 $i = 1 \ldots N_i$, and $j = 1 \ldots N_j$,

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

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

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

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

$$\begin{aligned} E[(\tilde{u}_{ijt}^a)^2] = E[(\tilde{\mu}_{ij}^a + \tilde{\varepsilon}_{ijt}^a)^2] &= E[(\tilde{\mu}_{ij}^a)^2] + E[(\tilde{\varepsilon}_{ijt}^a)^2] \\ &= \tfrac{(N_i-1)(N_j-1)}{N_iN_j}(\sigma_\mu^2 + \sigma_\varepsilon^2) \,. \end{aligned}$$

Transforming according to

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

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

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

Finally, using

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

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

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

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

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

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

Note, however, that the FGLS estimator of model (2.2) is only consistent if the data grows in at least two dimensions, that is, any two of $N_i \to \infty$, $N_j \to \infty$, 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, or 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 2.3 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, for example, of an employee-firm data with an extensive number of workers, but with few hiring firms observed annually)? This would mean, that σ^2_μ and σ^2_υ is estimated consistently, unlike σ^2_ζ. In such cases, it makes more sense to assume ζ_{jt} to be fixed instead of random (while still assuming the randomness of μ_{ij} and υ_{it}), arriving to the so-called "mixed effects models", something explored in Section 2.4.

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

$$\begin{aligned} E[\tilde{u}^2_{ijt}] &= \tfrac{(N_i-1)(N_j-1)}{N_iN_j}\sigma^2_\varepsilon\,, \quad E[(\tilde{u}^a_{ijt})^2] = \tfrac{N_i-1}{N_i}(\sigma^2_\upsilon+\sigma^2_\varepsilon) \quad \text{and} \\ E[(\tilde{u}^b_{ijt})^2] &= \tfrac{N_j-1}{N_j}(\sigma^2_\zeta+\sigma^2_\varepsilon)\,, \end{aligned}$$

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

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

respectively. The estimators for the variance components then are

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

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

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

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

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

For model (2.8),

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

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

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

respectively. The estimators for the variance components are

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

For model (2.10), as

$$\begin{aligned} E[\tilde{u}_{ijt}^2] &= \tfrac{(N_iN_j-1)T-(N_i-1)-(N_j-1)}{N_iN_jT}\sigma_\varepsilon^2 \\ E[(\tilde{u}_{ijt}^a)^2] &= \tfrac{(N_iN_j-1)T-(N_j-1)}{N_iN_jT}(\sigma_\upsilon^2+\sigma_\varepsilon^2) \\ E[(\tilde{u}_{ijt}^b)^2] &= \tfrac{(N_iN_j-1)T-(N_i-1)}{N_iN_jT}(\sigma_\zeta^2+\sigma_\varepsilon^2) \\ E[(\tilde{u}_{ijt}^c)^2] &= \tfrac{N_iN_jT-N_i-N_j+1}{N_iN_jT}(\sigma_\mu^2+\sigma_\varepsilon^2) \end{aligned}$$

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

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

respectively. The estimators for the variance components are

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

Lastly, for model (2.12) we get

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

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

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

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 $Q^{-1}_{X\Omega X}$ variance, where $Q^{-1}_{X\Omega X} = \text{plim}_{N_i,T\to\infty} \frac{X'\hat{\Omega}^{-1}X}{N_iT}$ is assumed to be a finite, positive definite matrix. This holds model-wide.

We have no such luck, however, with the OLS estimator. The issue is best illustrated with model (2.12). It can be shown, just as with the usual 2D panel models, $\text{Var}(\hat{\beta}_{OLS}) = (X'X)^{-1}X'\hat{\Omega}X(X'X)^{-1}$ (with $\hat{\Omega}$ being model-specific, but let us assume for now, that it corresponds to (2.13)). In the asymptotic case, when $N_i, N_j \to \infty$, $\sqrt{N_iN_j}(\hat{\beta}_{OLS} - \beta)$ 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 2.4 collects

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

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

normalization factors needed for a finite $\mathrm{Var}(\hat{\beta}_{OLS})$ for the different models considered. As it is uncommon to normalize with 1, or with expression like $\frac{\sqrt{N_iN_j}}{\sqrt{A}}$, some insights into the normalizations are given in Appendix 1. Another interesting aspect is revealed by comparing Tables 2.2 and 2.3, that

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

Model	(2.2)	(2.4)	(2.6)	(2.8)	(2.10)	(2.12)
$N_i \to \infty$	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
$N_i, N_j \to \infty$	$\frac{\sqrt{N_iN_j}}{\sqrt{A}}$	$\frac{\sqrt{N_iN_j}}{\sqrt{A}}$	$\sqrt{N_j}$	1	1	$\sqrt{N_iN_j}$
$N_i, T \to \infty$	$\frac{\sqrt{N_iT}}{\sqrt{A}}$	$\sqrt{T}$	$\sqrt{T}$	$\frac{\sqrt{N_iT}}{\sqrt{A}}$	1	$\sqrt{N_i}$
$N_j, T \to \infty$	$\frac{\sqrt{N_jT}}{\sqrt{A}}$	$\sqrt{T}$	$\sqrt{N_jT}$	$\frac{\sqrt{N_jT}}{\sqrt{A}}$	1	$\sqrt{N_j}$
$N_i, N_j, T \to \infty$	$\frac{\sqrt{N_iN_jT}}{\sqrt{A}}$	$\frac{\sqrt{N_iN_j}}{\sqrt{A}}\sqrt{T}$	$\sqrt{N_jT}$	$\frac{\sqrt{N_iN_jT}}{\sqrt{A}}$	$\frac{\sqrt{N_iN_jT}}{\sqrt{A_1A_2}}$	$\sqrt{N_iN_j}$

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.

is the consistency requirements for the estimation of the variance components (Table 2.2) and the asymptotic results, when the FGLS converges to the Within estimator (Table 2.3).

Table 2.5: Asymptotic Results when the OLS should be used

Model	(2.2)	(2.4)	(2.6)	(2.8)	(2.10)	(2.12)
$N_i \to \infty$			$+$			$-$
$N_j \to \infty$			$-$			$-$
$T \to \infty$		$-$	$-$			$+$
$N_i, N_j \to \infty$	$-$	$+$	$+$		$+$	$-$
$N_i, T \to \infty$	$-$	$-$	$+$	$+$	$+$	$+$
$N_j, T \to \infty$	$-$	$-$	$-$	$+$	$+$	$+$
$N_i, N_j, T \to \infty$	$+$	$+$	$+$	$+$	$+$	$+$

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 2.5, 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 one. This is problematic, as some parameters, previously estimable, become suddenly unidentified. In such cases, we have to rely on the OLS estimates, rather than the FGLS. This is generally the case whenever a "$+$" sign is found in Table 2.5, most significant for models (2.8) and (2.10). For them, the FGLS is only consistent, when it is in fact the Within Estimator, leading to likely severe identification issues. The best case scenarios are indicated with a "$-$" sign, where the respective asymptotics are already enough for the consistency of the FGLS, but do not yet cause identification problems. Lastly, blank spaces are left in the table if, under the given asymptotic, the FGLS is not consistent. In such cases we can again rely on the consistency of the OLS, but its standard errors are inconsistent, just as with the FGLS.

2.3 Unbalanced Data

2.3.1 Structure of the Covariance Matrices

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 slope parameters are inconsistent, and further, the spectral decomposition of Ω is inapplicable. Next, we present the covariance matrices of the different models in an incomplete data framework, we show a feasible way to invert them, and then propose a method to estimate the variance components in this general setup.

In our modelling framework, just like in Chapter 1, incompleteness means, that for any (ij) pair of individuals, $t \in T_{ij}$, where T_{ij} index-set is a subset of the general $\{1, \dots, T\}$ index-set of the time periods spanned by the data. Further, let $|T_{ij}|$ denote the cardinality of T_{ij}, i.e., the number of its elements. Note, that for complete (balanced) data, $T_{ij} = \{1, \dots, T\}$, and $|T_{ij}| = T$ for all (ij). We also assume, that for each t there is at least one (ij) pair, for each i, there is at least one (jt) pair, and for each j, there is at least one (it) pair observed. This assumption is almost natural, as it simply requires individuals or time periods with no underlying observation to be dropped from the dataset. As the structure of the data now is quite complex, we need to introduce a few new notation and definitions along the way. Formally, let us call n_{it}, n_{jt}, n_i, n_j, and n_t the total number of observations for a given (it), (jt) pair, and for given individuals i, j, and time t, respectively. Further, let us call $\tilde{n}_{ij}$, $\tilde{n}_{it}$, $\tilde{n}_{jt}$ the total number of (ij), (it), and (jt) pairs present in the data. Remember, that in the balanced case, $\tilde{n}_{ij} = N_i N_j$, $\tilde{n}_{it} = N_i T$, 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 can 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 (2.2), u_{ijt} can be stacked into vector u. Remember, that in the complete case it is

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

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

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

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

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

These then can be used to construct the covariance matrix as

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

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

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

and

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

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

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

for model (2.4), and

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

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

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

---

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

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 (2.10) 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

$$\begin{aligned} D_5 &= \text{diag}\left\{(V_{11}'\iota_T, V_{12}'\iota_T, \ldots, V_{1N_j}'\iota_T)', \ldots, (V_{N_i1}'\iota_T, V_{N_i2}'\iota_T, \ldots, V_{N_iN_j}'\iota_T)'\right\} \\ D_6 &= \Big(\text{diag}\{V_{11}'\iota_T, V_{12}'\iota_T, \ldots, V_{1N_j}'\iota_T\}', \ldots \times \\ &\qquad \times \ldots, \text{diag}\{V_{N_i1}'\iota_T, V_{N_i2}'\iota_T, \ldots, V_{N_iN_j}'\iota_T\}'\Big)'\,. \end{aligned}$$

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

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

An important practical difficulty is that the spectral decomposition of the covariance matrices introduced in Section 2.2 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 matrices, which is simply done by leaving out the missing rows from the appropriate B. Specifically, let us call B^* the incompleteness-adjusted versions of any B, which we get by removing the rows corresponding to the missing observations. For example, the spectral decomposition (2.14) for the all-encompassing model reads as

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

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.[4] Namely, when the number of missing observations relative to the total number of observations is small, the inverse of Ω based on the quasi-spectral decomposition of it, Ω^{*-1}, approximate arbitrarily well Ω^{-1}. More precisely, if $[N_iN_jT - \sum_i \sum_j |T_{ij}|]/[N_iN_jT] \to 0$, then $(\Omega^{-1} - \Omega^{*-1}) \to 0$. This means that in large data sets, when the number of missing observation is small relative to the total number of observations, Ω^{*-1} can safely be used in the GLS estimator instead of Ω^{-1}. Let us give an example. Multi-dimensional panel data are often used to deal with trade (gravity) models. In these cases, however, when country i trade with country j, there are no (ii) (or (jj)) observations, there is no self-trade. Then the total number of observations is $N^2T - NT$ with NT being the number of missing observations due to no self-trade. Given that $[N^2T - (N^2T - NT)]/N^2T \to 0$ as the sample size increases, the quasi-spectral decomposition can be used in large data.

2.3.2 The Inverse of the Covariance Matrices

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

[4]This can be demonstrated by simulation.

(2002)). This leads us, for model (2.2), to

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

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

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

where D_1, D_2, D_3 are the incompleteness-adjusted dummy variable matrices, and are used to construct the P and R matrices sequentially: first, construct R^a to get P^a, then construct R^b to get P^b, and finally, construct R^c to get P^c. Proof of (2.17) can be found in Appendix 2. Note, that to get the inverse, we have to invert $\min\{N_iT;\ N_jT;\ N_iN_j\}$ matrices. The quasi-scalar form of (2.17) (which corresponds to the incomplete data version of transformation (2.15)) is

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

with

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

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

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

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

where now

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

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

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

defining the scalar transformation

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

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

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

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 (2.10), the inverse is again the result of a three-step procedure:

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

where

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

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

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

defining the scalar transformation

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

on a typical y_{ijt} variable.

2.3.3 Estimation of the Variance Components

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

$$E[u_{ijt}^2] = \sigma_\mu^2 + \sigma_\upsilon^2 + \sigma_\zeta^2 + \sigma_\varepsilon^2\,, \tag{2.23}$$

but now, let us define

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

It can be seen that

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

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

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

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

The estimation strategy of the variance components is exactly the same for all the other models. Let us keep for now the definitions of $\tilde{u}^b_{ijt}$, and $\tilde{u}^c_{ijt}$. For model (2.4), with $u_{ijt} = \upsilon_{it} + \zeta_{jt} + \varepsilon_{ijt}$, the estimators read as

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

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

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

Note, that these latter two estimators can be obtained from (2.25), by assuming $\hat{\sigma}^2_\mu = 0$ for model (2.4), and $\hat{\sigma}^2_\mu = \hat{\sigma}^2_\upsilon = 0$ for model (2.6).
For model (2.8), let us redefine the $\tilde{u}^k_{ijt}$-s, as

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

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

ε_{ijt},

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

From this set of identifying equations, the estimators are simply

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

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

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

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

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

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

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

in turn leading to

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

2.4 Extensions

So far we have seen how to formulate and estimate three-way error components models. However, it is more and more typical to have data sets which require an even higher dimensional approach. As the number of feasible model formulations grow exponentially along with the dimensions, there is no point to attempt to collect all of them. Rather, we will take the 4D representation of the all-encompassing model (2.2), and show how the extension to higher dimensions can be carried out.

2.4.1 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 $t = 1 \dots T$,

$$y_{ijst} = x'_{ijst}\beta + \mu_{ijs} + \upsilon_{ist} + \zeta_{jst} + \lambda_{ijt} + \varepsilon_{ijst} = x'_{ijst}\beta + u_{ijst}\,, \qquad (2.31)$$

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

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

The covariance matrix of such error components formulation is

$$\Omega = E(uu') = \sigma_\mu^2 (I_{N_iN_jN_s} \otimes J_T) + \sigma_\upsilon^2 (I_{N_i} \otimes J_{N_j} \otimes I_{N_sT}) \\ + \sigma_\zeta^2 (J_{N_i} \otimes I_{N_jN_sT}) + \sigma_\lambda^2 (I_{N_iN_j} \otimes J_{N_s} \otimes I_T) + \sigma_\varepsilon^2 I_{N_iN_jN_sT} . \tag{2.32}$$

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 2.3, 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 (2.31) as

$$\Omega = E(uu') = \sigma_\varepsilon^2 I + \sigma_\mu^2 D_1 D_1' + \sigma_\upsilon^2 D_2 D_2' + \sigma_\zeta^2 D_3 D_3' + \sigma_\lambda^2 D_4 D_4' , \tag{2.33}$$

where, in the complete case,

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

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 matrices respectively, but now we delete, from each D_k, the rows corresponding to the missing observations to reflect the unbalanced nature of the data. The inverse of such covariance formulation can be reached in steps, that is, one has to derive

$$\Omega^{-1}\sigma_\varepsilon^2 = P^c - P^c D_4 (R^d)^{-1} D_4' P^c \tag{2.34}$$

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

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

Even though the calculation above alleviates some of the "dimensionality curse",[5] to perform the inverse we still have to manipulate potentially large matrices. The last step in finishing the FGLS estimation of the incomplete 4D models is to estimate the variance components. Fortunately, this is not too difficult, however, due to the size of the formulas, the results are presented in Appendix 3.

2.4.2 Mixed FE-RE Models

As briefly mentioned in Section 2.3, when one of the indices is small, it makes more sense to treat the effects depending on that index as fixed. As an illustration, consider an *employee i–employer j–time t*-type dataset, where we usually have a very large set of i, but relatively low j and t. All this means, that the all-encompassing model (2.2) can now be rewritten as

$$y_{ijt} = x_{ijt}'\beta + \alpha_{jt} + \mu_{ij} + \upsilon_{it} + \varepsilon_{ijt}\,, \tag{2.35}$$

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} enters the model as a fixed dummy, and that $u_{ijt} =$

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

$\mu_{ij} + \upsilon_{it} + \varepsilon_{ijt}$ remains the random component. To estimate such model specification, keeping an eye on optimality, we have to follow a two step procedure. First, to get rid of the fixed effects, we define a projection orthogonal to α_{jt}. Then, on this transformed model, we perform FGLS. The resulting estimator is analytically not too complicated, and although as seen in Chapter 1, restricted x_{jt} regressors can not be estimated from (2.35), $\hat{\beta}_{Mixed}$ is identified and consistent for the rest of the variables. This is a substantial improvement over the FGLS estimation of model (2.2), when N_j and T are both small, as in such cases, as shown in Section 2.3, the inconsistency of the variance components estimators carries over to the model parameters. The projection needed to eliminate α_{jt} is

$$M_{D_1} = I - D_1(D_1'D_1)^{-1}D_1' \quad \text{or in scalar form,} \quad \tilde{y}_{ijt} = y_{ijt} - \bar{y}_{.jt}\,. \tag{2.36}$$

Notice, that the resulting transformed (2.35),

$$\tilde{y}_{ijt} = \tilde{x}_{ijt}\beta + \tilde{u}_{ijt}\,, \tag{2.37}$$

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

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

while its inverse can be derived using the trick introduced in Section 2.2 (using the substitution $I_{N_i} = Q_{N_i} + \bar{J}_{N_i}$), giving

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

with

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

After all, the mixed effects estimation of (2.35) is identical to the FGLS estimation of (2.37). The only step remaining is to estimate the variance components. In particular,

$$\begin{aligned}
\hat{\sigma}_\varepsilon^2 &= \tfrac{1}{(N_i-1)(N_j-1)(T-1)} \textstyle\sum_{ijt} (\tilde{\hat{u}}_{ijt}^a)^2 \\
\hat{\sigma}_\mu^2 &= \tfrac{1}{(N_i-1)(N_j-1)T} \textstyle\sum_{ijt} (\tilde{\hat{u}}_{ijt}^b)^2 - \hat{\sigma}_\varepsilon^2 \\
\hat{\sigma}_\upsilon^2 &= \tfrac{1}{(N_i-1)N_j(T-1)} \textstyle\sum_{ijt} (\tilde{\hat{u}}_{ijt}^c)^2 - \hat{\sigma}_\varepsilon^2\,,
\end{aligned} \tag{2.38}$$

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

$$\begin{aligned}
\tilde{u}_{ijt}^a &= u_{ijt} - \bar{u}_{.jt} - \bar{u}_{i.t} - \bar{u}_{ij.} + \bar{u}_{..t} + \bar{u}_{.j.} + \bar{u}_{i..} - \bar{u}_{...}\,, \\
\tilde{u}_{ijt}^b &= u_{ijt} - \bar{u}_{.jt} - \bar{u}_{i.t} + \bar{u}_{..t}\,, \quad \text{and} \quad \tilde{u}_{ijt}^c = u_{ijt} - \bar{u}_{.jt} - \bar{u}_{ij.} + \bar{u}_{.j.}\,.
\end{aligned} \tag{2.39}$$

The next question is to what extent the above algorithm has to be modified for unbalanced data. First, transformation (2.36) is successful in eliminating α_{jt} from model (2.35) in this case as well. Second, the resulting transformed covariance matrix now can not be represented by kronecker products, instead, to invert it, we have to rely on tricks derived in Section 2.3. The estimation of the variance components are done by first adjusting the transformations $\tilde{u}_{ijt}^a$, $\tilde{u}_{ijt}^b$ in (2.39) to incomplete data, that is, using their semi-scalar representatives (1.25), (1.24), and for $\tilde{u}_{ijt}^c$

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

where u contain the stacked disturbances (with elements u_{ijt}), $\tilde{u}^c$ is its transformed counterpart, $M^{(1)} = I - \tilde{D}_1(\tilde{D}_1'\tilde{D}_1)^-\tilde{D}_1'$, and $\tilde{D}_1$ and $\tilde{D}_3$ are obtained from $D_1 = (I_{N_iN_j} \otimes \iota_T)$ and $D_3 = (\iota_{N_i} \otimes I_{N_jT})$ by leaving out the rows corresponding to missing observations. Finally, we have to set the proper sample sizes in (2.38).

2.5 Testing

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

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

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

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

where both M_1 and M_2 are orthogonal projectors, and the degrees of freedom is calculated from p, r, and $\tilde{k}$. Table 2.6 captures each specific matrix and constant for all hypotheses listed above.

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

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

Table 2.6: Specific functional forms of the ANOVA F-test

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

where, as defined, $M_Z = I - Z(Z'Z)^- Z'$, $D_1 = (I_{N_iN_j} \otimes \iota_T)$, $D_2 = (I_{N_i} \otimes \iota_{N_j} \otimes I_T)$, *and* $D_3 = (\iota_{N_i} \otimes I_{N_jT})$.

where

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

and

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

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

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

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

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

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

Not much changes when the underlying data is incomplete. In principle, the orthogonal projections M_{Z_1} and M_{Z_2} now cannot be represented as linear transformations on the data, only in semi-scalar form, with the inclusion of some matrix operations, listed in Section 1.4.2. For example, (2.41) corresponds to (1.25), while (2.42) corresponds to (1.24) in case of incomplete data. Once we have the incomplete-robust $\tilde{y}$, $\tilde{\tilde{y}}$ (similarly for X) variables, the F statistic is obtained as in (2.40), with the properly computed degrees of freedom.

2.6 Conclusion

For large data sets, when observations can be considered as samples from an underlying population, random effects specifications seem to be more suited to deal with multi-dimensional data sets. FGLS estimators for three-way error components models are almost as easily obtained as for the traditional 2D panel models, however the resulting asymptotic requirements for their consistency are more peculiar. In fact, now the data can grow in three directions, and only some of the asymptotic cases are sufficient for consistency. Interestingly, for some error components specifications, consistency implies the convergence of the FGLS estimator to the Within one. This is utterly important, as under the Within estimation, the parameters of some fixed regressors are unidentified, which is in fact carried over to the FGLS estimation of those parameters as well. To solve this, we have shown that a simple OLS can be sufficient to get the full set of parameter estimates (of course, bearing the price of inefficiency), wherever this identification problem persists. The main results of the chapter are also extended to treat incomplete data and towards higher dimensions as well.

References

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

Balazsi, Laszlo, Laszlo Matyas, and Tom Wansbeek, "The Estimation of Multi-dimensional Fixed Effects Panel Data Models," *Forthcoming in Econometrics Reviews*, 2015.

Baldwin, Richard and Daria Taglioni, "Gravity for Dummies and Dummies for the Gravity Equations," *NBER Working Paper 12516*, 2006.

Baltagi, Badi H., "On estimating from a more general time-series cum cross-section data structure," *The American Economist*, 1987, *31*, 69–71.

_ **and Young-Jae Chang**, "Incomplete panels: A comparative study of alternative estimators for the unbalanced one-way error component regression model," *Journal of Econometrics*, 1994, *62*, 67–89.

_ **, Laszlo Matyas, and Patrick Sevestre**, *Error Components Models*, Springer Verlag, 2008.

_ **, Peter Egger, and Michael Pfaffermayr**, "A generalized Design for Bilateral Trade Flow Models," *Economic Letters*, 2003, *80*, 391–397.

— , **Seuck H. Song, and Byoung C. Jung**, "The unbalanced nested error component regression model," *Journal of Econometrics*, 2001, *101*, 357–381.

— , — , **and** — , "A comparative study of alternative estimators for the unbalanced two-way error component regression model," *Econometrics Journal*, 2002, *5*, 480–493.

— , **Young-Jae Chang, and Qi Li**, "Monte Carlo results on several new and existing tests for the error component model," *Journal of Econometrics*, 1992, *54*, 95–120.

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

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

Ebbes, Peter, Ulf Bockenholt, and Michel Wedel, "Regressor and Random-effects Dependencies in Multilevel Models," *Statistica Neerlandica*, 2004, *58(2)*, 161–178.

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

Fuller, Wayne A. and George E. Battese, "Transformations for Estimation of Linear Models with Nested-Error Structure," *Journal of the American Statistical Association*, 1973, *68*, 626–632.

Ghosh, Sukesh K., "Estimating from a More General Time-Series cum Cross-Section Data Structure," *The American Economist*, 1976, *20*, 15–21.

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

Hornok, Cecilia, "Gravity or Dummies? The Limits of Identification in Gravity Estimations," *CeFiG Working Papers*, 2011, *No. 15.*

Hubler, Olaf, "Multilevel and Nonlinear Panel Data Models," *Allgemeines Statistisches Archiv*, 2006, *90*, 121–136.

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

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

Luke, Douglas A., *Multilevel Modeling*, Thousand Oaks, CA: SAGE Publications, 2004.

Matyas, Laszlo, "The Gravity Model: Some Econometric Considerations," *The World Economy*, 1998, *21*, 397–401.

Rao, Radhakrishna and Jurgen Kleffe, *Handbook of Statistics 1*, North-Holland Publisher, 1980.

Scott, Marc A., Jeffrey S. Simonoff, and Brian D. Marx, *The SAGE handbook of multilevel modeling*, London: SAGE Publications Ltd., 2013.

Wansbeek, Tom and Arie Kapteyn, "A Simple Way to Obtain the Spectral Decomposition of Variance Components Models for Balanced Data," *Communications in Statistics - Theory and Methods*, 1982, *11(18)*, 2105–2112.

— **and** — , "Estimation of the Error-Components Model with Incomplete Panels," *Journal of Econometrics*, 1989, *41*, 341–361.

Appendix 1

Example for normalizing with 1: Model (2.12), $T \to \infty$.

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

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

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

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

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

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

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

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

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

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

Example for normalizing with $\frac{\sqrt{N_iN_j}}{A}$: Model (2.2), $N_i, N_j \to \infty$.

Using the standard $\sqrt{N_iN_j}$ normalization factor gives

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

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

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

Observe, that

$$\begin{aligned}\text{plim}_{N_i,N_j\to\infty} \quad \frac{X'\Omega X}{N_iN_j} &= \text{plim}_{N_i,N_j\to\infty} \frac{\sigma_\varepsilon^2 X'X}{N_iN_j} + \text{plim}_{N_i,N_j\to\infty} \frac{\sigma_\mu X'(I_{N_iN_j}\otimes J_T)X}{N_iN_j} \\ &+ \text{plim}_{N_i,N_j\to\infty} \frac{\sigma_\upsilon^2 X'(I_{N_i}\otimes J_{N_j}\otimes I_T)X}{N_iN_j} + \text{plim}_{N_i,N_j\to\infty} \frac{\sigma_\zeta^2 X'(J_{N_i}\otimes I_{N_jT})X}{N_iN_j}\end{aligned} \tag{2.43}$$

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 (2.43) are zero, but the fourth is finite:

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

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 (2.17).

Let us make the proof only for model (2.2) (so formula (2.17)), the rest is just direct application of the derivation below. The outline of the proof is based on Wansbeek and Kapteyn (1989).

First, notice, that using the Woodbury matrix identity,

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

Second, using that

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

gives

$$R^b - D_2'P^aD_2 = \frac{\sigma_\varepsilon^2}{\sigma_\upsilon^2}I \, .$$

Using the Woodbury matrix identity for the second time,

$$
\begin{aligned}
(P^b)^{-1} &= \left(P^a - P^a D_2 (R^b)^{-1} D_2' P^a\right)^{-1} \\
&= (P^a)^{-1} + (P^a)^{-1} P^a D_2 \left(R^b - D_2' P^a (P^a)^{-1} P^a D_2\right)^{-1} D_2' P^a (P^a)^{-1} \\
&= (P^a)^{-1} + D_2 \left(R^b - D_2' P^a D_2\right)^{-1} D_2' = (P^a)^{-1} + D_2 \left(\tfrac{\sigma_\varepsilon^2}{\sigma_\upsilon^2} I\right)^{-1} D_2' \\
&= I + \tfrac{\sigma_\mu^2}{\sigma_\varepsilon^2} D_1 D_1' + \tfrac{\sigma_\upsilon^2}{\sigma_\varepsilon^2} D_2 D_2' \,.
\end{aligned}
$$

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

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

again, and so

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

Appendix 3

Inverse of (2.32), and the estimation of the variance components

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

with

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

The estimation of the variance components in case of complete data are as follows:

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

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

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

Estimation of the variance components in case of incomplete data yields

$$\begin{aligned}
\hat{\sigma}^2_\mu &= \tfrac{1}{\sum_{ijs}|T_{ijs}|}\sum_{ijst}\hat{u}^2_{ijst} - \tfrac{1}{\tilde{n}_{ijs}}\sum_{ijs}\tfrac{1}{|T_{ijs}|-1}\sum_t(\tilde{\hat{u}}^a_{ijst})^2 \\
\hat{\sigma}^2_\upsilon &= \tfrac{1}{\sum_{ijs}|T_{ijs}|}\sum_{ijst}\hat{u}^2_{ijst} - \tfrac{1}{\tilde{n}_{ist}}\sum_{ist}\tfrac{1}{n_{ist}-1}\sum_j(\tilde{\hat{u}}^b_{ijst})^2 \\
\hat{\sigma}^2_\zeta &= \tfrac{1}{\sum_{ijs}|T_{ijs}|}\sum_{ijst}\hat{u}^2_{ijst} - \tfrac{1}{\tilde{n}_{jst}}\sum_{jt}\tfrac{1}{n_{jst}-1}\sum_i(\tilde{\hat{u}}^c_{ijst})^2 \\
\hat{\sigma}^2_\lambda &= \tfrac{1}{\sum_{ijs}|T_{ijs}|}\sum_{ijst}\hat{u}^2_{ijst} - \tfrac{1}{\tilde{n}_{ijt}}\sum_{ijt}\tfrac{1}{n_{ijt}-1}\sum_s(\tilde{\hat{u}}^d_{ijst})^2 \\
\hat{\sigma}^2_\varepsilon &= \tfrac{1}{\sum_{ijs}|T_{ijs}|}\sum_{ijst}\hat{u}^2_{ijst} - \hat{\sigma}^2_\mu - \hat{\sigma}^2_\upsilon - \hat{\sigma}^2_\zeta - \hat{\sigma}^2_\lambda\,,
\end{aligned} \tag{2.44}$$

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

$$\begin{aligned}
\tilde{u}^a_{ijst} &= u_{ijst} - \tfrac{1}{|T_{ijs}|}\sum_t u_{ijst}\,, \quad \tilde{u}^b_{ijst} = u_{ijst} - \tfrac{1}{n_{ist}}\sum_j u_{ijst}\,, \\
\tilde{u}^c_{ijst} &= u_{ijst} - \tfrac{1}{n_{jst}}\sum_i u_{ijst}\,, \quad \tilde{u}^d_{ijst} = u_{ijst} - \tfrac{1}{n_{ijt}}\sum_s u_{ijst}\,.
\end{aligned}$$

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