DEPARTMENT OF ECONOMICS
WORKING PAPERS

economics.ceu.hu

The Estimation of Multi-dimensional Fixed Effects Panel Data Models[1]

by

László Balázsi[2], László Mátyás[3] and Tom Wansbeek[4]

2014/1

[1] *This is an expanded, updated and re-written version of CEU-Economics Working Paper 2012/2.*

[2] Department of Economics, Central European University
[3] Department of Economics, Central European University, matyas@ceu.hu
[4] University of Groningen

Abstract

The paper introduces for the most frequently used three-dimensional fixed effects panel data models the appropriate Within estimators. It analyzes the behavior of these estimators in the case of no-self-flow data, unbalanced data and dynamic autoregressive models. Then the main results are generalized for higher dimensional panel data sets as well.

KEY WORDS: panel data, unbalanced panel, dynamic panel data model, multidimensional panel data, fixed effects, trade models, gravity models, FDI

Key words: panel data, multidimensional panel data, random effects, error components model, trade model, gravity model, firm and product level data, micro trade data.

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

Acknowledgements

Support by Australian Research Council grant DP110103824 is kindly acknowledged. Comments received from the participants of the 18th and 19th Panel Data Conferences, Paris 2012 and London 2013 respectively, are also much appreciated.

1. Introduction

Multidimensional panel data sets are becoming more readily available, and used to study phenomena like international trade and/or capital flow between countries or regions, the trading volume across several products and stores over time (three panel dimensions), the air passenger numbers between multiple hubs deserved by different airlines (four panel dimensions) and so on. Over the years several, mostly fixed effects, specifications have been worked out to take into account the specific three (or higher) dimensional nature and heterogeneity of these kinds of data sets. In this paper in Section 2 we present the different fixed effects formulations introduced in the literature to deal with three-dimensional panels and derive the proper Within[2] transformations for each model. In Section 3 we first have a closer look at a problem typical for such data sets, that is the lack of self-flow observations. Then we also analyze the properties of the Within estimators in an unbalanced data setting. In Section 4 we investigate how the different Within estimators behave in the case of a dynamic specification, generalizing the seminal results of *Nickell* [1981], in Section 5 we extend our results for higher dimensional data sets, in Section 6 we allow a simple form of cross-correlation for the disturbances and finally, we draw some conclusions in Section 7.

2. Models with Different Types of Heterogeneity and the Within Transformation

In three-dimensional panel data sets the dependent variable of a model is observed along three indices such as y_{ijt}, $i = 1, \ldots, N_1$, $j = 1, \ldots, N_2$, and $t = 1, \ldots, T$. As in economic flows such as trade, capital (FDI), etc., there is some kind of reciprocity, we assume to start with, that $N_1 = N_2 = N$. Implicitly we also assume that the set of individuals in the observation sets i and j are the same, then we relax this assumption later on. The main question is how to formalize the individual and time heterogeneity, in our case the fixed effects. Different forms of heterogeneity yield naturally different models. In theory any fixed effects three-dimensional panel data model can directly be estimated, say for example, by least squares (LS). This involves the explicit incorporation in the model of the fixed effects through dummy variables (see for example formulation (14) later on). The resulting estimator is usually called

[2] We must notice here, for those familiar with the usual panel data terminology, that in a higher dimensional setup the within and between groups variation of the data is somewhat arbitrary, and so the distinction between Within and Between estimators would make our narrative unnecessarily complex. Therefore in this paper all estimators using a kind of projection are called Within estimators.

Least Squares Dummy Variable (LSDV) estimator. However, it is well known that the first moment of the LS estimators is invariant to linear transformations, as long as the transformed explanatory variables and disturbance terms remain uncorrelated. So if we could transform the model, that is all variables of the model, in such a way that the transformation wipes out the fixed effects, and then estimate this transformed model by least squares, we would get parameter estimates with similar first moment properties (unbiasedness) as those from the estimation of the original untransformed model. This would be simpler as the fixed effects then need not to be estimated or explicitly incorporated into the model.[3] We must emphasize, however, that these transformations are usually not unique in our context. The resulting different Within estimators (for the same model), although have the same bias/unbiasedness, may not give numerically the same parameter estimates. This comes from the fact that the different Within transformations represent different projection in the (i, j, t) space, so the corresponding Within estimators may in fact use different subsets of the three-dimensional data space. Due to the Gauss-Markov and the Frisch-Waugh theorems (see, for example, *Gourieroux and Monfort* [1989]), there is always an optimal Within estimator, exactly the one which is based on the transformations generated by the appropriate LSDV estimator. Why to bother then, and not always use the LSDV estimator directly? First, because when the data becomes larger, the estimation of a model with the fixed effects explicitly incorporated into it is quite difficult, or even practically impossible, so the use of Within estimators can be quite useful. Then, we may also exploit the different projections and the resulting various Within estimators to deal with some data generated problems.

The first attempt to properly extend the standard fixed effects panel data model (see for example *Baltagi* [1995] or *Balestra and Krishnakumar* [2008]) to a multidimensional setup was proposed by *Matyas* [1997]. The specification of this model is

$$y_{ijt} = \beta' x_{ijt} + \alpha_i + \gamma_j + \lambda_t + \varepsilon_{ijt} \qquad i = 1, \dots, N \quad j = 1, \dots, N, \quad t = 1, \dots, T, \tag{1}$$

where the α, γ and λ parameters are time and country specific fixed effects, the x variables are the usual covariates, β $(K \times 1)$ the focus structural parameters and ε is the idiosyncratic disturbance term, for which (unless otherwise stated)

$$E(\varepsilon_{ijt}) = 0, \qquad E(\varepsilon_{ijt}\varepsilon_{i'j't'}) = \begin{cases} \sigma^2_\varepsilon & \text{if } i = i',\ j = j' \text{ and } t = t' \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

[3] An early partial overview of these transformations can be found in *Matyas, Harris and Konya* [2011].

and we also assume that the covariates and the disturbance terms are uncorrelated.

The simplest Within transformation for this model is

$$(y_{ijt} - \bar{y}_{ij.} - \bar{y}_{..t} + \bar{y}_{...}) \tag{3}$$

where

$$\bar{y}_{ij.} = 1/T \sum_t y_{ijt}$$
$$\bar{y}_{..t} = 1/N^2 \sum_i \sum_j y_{ijt}$$
$$\bar{y}_{...} = 1/N^2 T \sum_i \sum_j \sum_t y_{ijt}$$

However, the optimal Within transformation (which actually gives numerically the same parameter estimates as the direct LS estimation of model (1), that is the LSDV estimator) is in fact

$$(y_{ijt} - \bar{y}_{i..} - \bar{y}_{.j.} - \bar{y}_{..t} + 2\bar{y}_{...}) \tag{4}$$

where

$$\bar{y}_{i..} = 1/(NT) \sum_j \sum_t y_{ijt}$$
$$\bar{y}_{.j.} = 1/(NT) \sum_i \sum_t y_{ijt}$$

Let us note here that this model is suited to deal with purely cross sectional data as well (that is when $T = 1$). In this case, there are only the α_i and γ_j fixed effects and the the appropriate Within transformation is $(y_{ij} - \bar{y}_{.j} - \bar{y}_{i.} + \bar{y}_{..})$ with $\bar{y}_{..} = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} y_{ij}$.

Another model has been proposed by *Egger and Pfanffermayr* [2003] which takes into account bilateral interaction effects. The model specification is

$$y_{ijt} = \beta' x_{ijt} + \gamma_{ij} + \varepsilon_{ijt} \tag{5}$$

where the γ_{ij} are the bilateral specific fixed effects (this approach can easily be extended to account for multilateral effects as well). The simplest (and optimal) Within transformation which clears the fixed effects now is

$$(y_{ijt} - \bar{y}_{ij.}) \quad \text{where} \quad \bar{y}_{ij.} = 1/T \sum_t y_{ijt} \tag{6}$$

It can be seen that the use of the Within estimator here, and even more so for the models discussed later, is highly recommended as direct estimation of the model by

LS would involve the estimation of $(N \times N)$ parameters which is no very practical for larger N. For model (14) this would even be practically impossible.

A variant of model (5) often used in empirical studies is

$$y_{ijt} = \beta' x_{ijt} + \gamma_{ij} + \lambda_t + \varepsilon_{ijt} \tag{7}$$

As model (1) is in fact a special case of this model (7), transformation (3) can be used to clear the fixed effects. While transformation (3) leads to the optimal Within estimator for model (7), its is clear why it is not optimal for model (1): it "over-clears" the fixed effects, as it does not take into account the parameter restrictions $\gamma_{ij} = \alpha_i + \gamma_i$. It is worth noticing that models (5) and (7) are in fact straight panel data models where the individuals are now the (ij) pairs.

Baltagi et al. [2003], *Baldwin and Taglioni* [2006] and *Baier and Bergstrand* [2007] suggested several other forms of fixed effects. A simpler model is

$$y_{ijt} = \beta' x_{ijt} + \alpha_{jt} + \varepsilon_{ijt} \tag{8}$$

The Within transformation which clears the fixed effects is

$$(y_{ijt} - \bar{y}_{.jt}) \quad \text{where} \quad \bar{y}_{.jt} = 1/N \sum_i y_{ijt} \tag{9}$$

It is reasonable to present the symmetric version of this model (with α_{it} fixed effects), however, as it has the exact same properties, we take the two models together.

The most frequently used variation of this model is

$$y_{ijt} = \beta' x_{ijt} + \alpha_{it} + \alpha^*_{jt} + \varepsilon_{ijt} \tag{10}$$

The required Within transformation here is

$$(y_{ijt} - 1/N \sum_i y_{ijt} - 1/N \sum_j y_{ijt} + 1/N^2 \sum_i \sum_j y_{ijt})$$

or in short

$$(y_{ijt} - \bar{y}_{.jt} - \bar{y}_{i.t} + \bar{y}_{..t}) \tag{11}$$

Let us notice here that transformation (11) clears the fixed effects for model (1) as well, but of course the resulting Within estimator is not optimal. The model which encompasses all above effects is

$$y_{ijt} = \beta' x_{ijt} + \gamma_{ij} + \alpha_{it} + \alpha^*_{jt} + \varepsilon_{ijt} \tag{12}$$

By applying suitable restrictions to model (12) we can obtain the models discussed above. The Within transformation for this model is

$$(y_{ijt} - 1/T\sum_t y_{ijt} - 1/N\sum_i y_{ijt} - 1/N\sum_j y_{ijt} + 1/N^2\sum_i\sum_j y_{ijt} \\ + 1/(NT)\sum_i\sum_t y_{ijt} + 1/(NT)\sum_j\sum_t y_{ijt} - 1/(N^2T)\sum_i\sum_j\sum_t y_{ijt})$$

or in a shorter form

$$(y_{ijt} - \bar{y}_{ij.} - \bar{y}_{.jt} - \bar{y}_{i.t} + \bar{y}_{..t} + \bar{y}_{.j.} + \bar{y}_{i..} - \bar{y}_{...}) \tag{13}$$

We can write up these Within transformations in a more compact matrix form using *Davis'* [2002] and *Hornok's* [2011] approach. Model (12) in matrix form is

$$y = X\beta + \tilde{D}_1\gamma + \tilde{D}_2\alpha + \tilde{D}_3\alpha^* + \varepsilon \tag{14}$$

where $y, (N^2T \times 1)$ is the vector of the dependent variable, $X, (N^2T \times K)$ is the matrix of explanatory variables, γ, α and α^* are the vectors of fixed effects with size $(N^2T \times N^2)$, $(N^2T \times NT)$ and $(N^2T \times NT)$ respectively,

$$\tilde{D}_1 = I_{N^2} \otimes l_T, \quad \tilde{D}_2 = I_N \otimes l_N \otimes I_T \quad \text{and} \quad \tilde{D}_3 = l_N \otimes I_{NT}$$

l is the vector of ones and I is the identity matrix with the appropriate size in the index. Let $D = (\tilde{D}_1, \tilde{D}_2, \tilde{D}_3)$, $Q_D = D(D'D)^- D'$ and $P_D = I - Q_D$, where $(D'D)^-$ denotes the Moore-Penrose generalized inverse, as in general $D'D$ has no full rank. Using *Davis'* [2002] method it can be shown that $P_D = P_1 - Q_2 - Q_3$ where

$$\begin{aligned} P_1 &= (I_N - \bar{J}_N) \otimes I_{NT} \\ Q_2 &= (I_N - \bar{J}_N) \otimes \bar{J}_N \otimes I_T \\ Q_3 &= (I_N - \bar{J}_N) \otimes (I_N - \bar{J}_N) \otimes \bar{J}_T \\ \bar{J}_N &= \frac{1}{N} J_N, \quad \bar{J}_T = \frac{1}{T} J_T \end{aligned}$$

and J is the matrix of ones with its size in the index. Collecting all these terms we get

$$\begin{aligned} P_D &= \left[(I_N - \bar{J}_N) \otimes (I_N - \bar{J}_N) \otimes (I_T - \bar{J}_T)\right] \\ &= I_{N^2T} - (\bar{J}_N \otimes I_{NT}) - (I_N \otimes \bar{J}_N \otimes I_T) - (I_{N^2} \otimes \bar{J}_T) \\ &\quad + (I_N \otimes \bar{J}_{NT}) + (\bar{J}_N \otimes I_N \otimes \bar{J}_T) + (\bar{J}_{N^2} \otimes I_T) - \bar{J}_{N^2T} \end{aligned}$$

The typical element of P_D gives the transformation (13). By appropriate restrictions on the parameters of (14) we get back the previously analysed Within transformations. Now transforming model (14) with transformation (13) leads to

$$\underbrace{P_D y}_{y_p} = \underbrace{P_D X}_{X_p} \beta + \underbrace{P_D \tilde{D}_1}_{=0} \gamma + \underbrace{P_D \tilde{D}_2}_{=0} \alpha + \underbrace{P_D \tilde{D}_3}_{=0} \alpha^* + \underbrace{P_D \varepsilon}_{\varepsilon_p}$$

and the corresponding Within estimator is

$$\widehat{\beta}_W = (X_p' X_p)^{-1} X_p y_p$$

This in fact is the optimal estimator as P_D is the Frisch-Waugh projection matrix, implying the optimality of $\widehat{\beta}_W$.

3. Some Data Problems

3.1 No Self Flow Data

As these multidimensional panel data models are frequently used to deal with flow types of data like trade, capital movements (FDI), etc., it is important to have a closer look at the case when, by nature, we do not observe self flow. This means that from the (ijt) indexes we do not have observations for the dependent variable of the model when $i = j$ for any t. This is the first step to relax our initial assumption that $N_1 = N_2 = N$ and that the observation sets i and j are equivalent.

For most of the previously introduced models this is not a problem, the Within transformations work as they are meant to eliminate the fixed effects. However, this is not true in general. The fixed effects are not fully eliminated in models (1) (transformation (4)), (10) and (12). For example in the case of model (1) and transformation (4), instead of canceled out fixed effects, we end up with the following remaining fixed effects

$$\tilde{\alpha}_i = \frac{1}{N}\alpha_j - \frac{1}{N(N-1)} \sum_{i=1;\ i\neq j}^{N} \alpha_i$$

$$\tilde{\gamma}_j = \frac{1}{N}\gamma_i - \frac{1}{N(N-1)} \sum_{j=1;\ j\neq i}^{N} \gamma_j$$

and for the time effects

$$\tilde{\lambda}_t = 0$$

Detailed calculations can be found in Appendix 1. So clearly this Within estimator now is biased. The bias is of course eliminated if we add the (ii) observations back to the above bias formulae, and also, quite intuitively, when $N \to \infty$.

In order to derive the optimal unbiased Within transformation, let us consider first the purely cross sectional version of model (1)

$$y_{ij} = \beta' x_{ij} + \alpha_i + \gamma_j + \varepsilon_{ij}, \tag{15}$$

or in matrix form, with l_N an N-vector of ones,

$$\begin{aligned} y &= X\beta + (I_N \otimes l_N)\alpha + (l_N \otimes I_N)\gamma + \varepsilon \\ &\equiv X\beta + D_\alpha \alpha + D_\gamma \gamma + \varepsilon. \end{aligned}$$

There are no data with $i = j$, so we want to eliminate these from the model. The matrix L of order $N^2 \times N(N-1)$ does that. Then the model for the observed data can be written as

$$L'y = L'X\beta + L'D_\alpha \alpha + L'D_\gamma \gamma + L'\varepsilon.$$

Let $D_{(1)} \equiv (D_\alpha, D_\gamma)$ and notice that $D_{(1)}$ and $L'D_{(1)}$ have incomplete column rank. We are interested in estimating β. In order to avoid possible dimensionality problems, we use the Frisch-Waugh theorem and project the α_i and γ_j out of the model, to perform OLS on the transformed model. The projection is achieved by the projection matrix orthogonal to $L'D_{(1)}$,

$$P_{(1)} \equiv I_{N(N-1)} - L'D_{(1)} W_{(1)}^{-} D_{(1)}' L,$$

with $W_{(1)} \equiv D_{(1)}' L L' D_{(1)}$. We are interested in a simple expression for the elements of the transformed variable $P_{(1)} L' y$, now indicated by a tilde. When the data with $i = j$ are observed, remember this expression was

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

but now it is more complicated since, for $i \neq j$ so $(e_i \otimes e_j)' D_{(1)} = (e_i \otimes e_j)' L L' D_{(1)} = (e_i', e_j')$,

$$\begin{aligned} \tilde{y}_{ij} &= (e_i \otimes e_j)' L P_{(1)} L' y \\ &= y_{ij} - (e_i', e_j') W_{(1)}^{-} D_{(1)}' L L' y, \end{aligned} \tag{16}$$

with e_i the ith unit vector.

We first have to further elaborate on $W_{(1)}$. For detailed calculations please see Appendix 2. Let us return to (16) with the updated form of $W_{(1)}$

$$(e_i', e_j')W_{(1)}^{-} = \frac{1}{N(N-2)}\left((N-1)e_i' + e_j', e_i' + (N-1)e_j'\right) - \frac{1}{2(N-1)(N-2)}\left(l_N', l_N'\right).$$

So at the end, after multiplying $(e_i', e_j')W_{(1)}^{-}$ and $D_{(1)}'LL'y$, we obtain

$$\begin{aligned}\tilde{y}_{ij} = y_{ij} &- \frac{N-1}{N(N-2)}\left(y_{i+} + y_{+j}\right) - \frac{1}{N(N-2)}\left(y_{+i} + y_{j+}\right) \\ &+ \frac{1}{(N-1)(N-2)}y_{++}.\end{aligned} \tag{17}$$

with

$$\begin{aligned} y_{i+} &= \sum_j y_{ij}, \quad y_{+j} = \sum_i y_{ij}, \quad y_{+i} = \sum_i y_{ij,(j=i)}, \\ y_{j+} &= \sum_j y_{ij,(i=j)}, \quad y_{++} = \sum_{ij} y_{ij} \end{aligned}$$

The introduction of this "+" notation is needed in order to avoid confusions with the indexing. For example, now we take sums with respect to the first index, fixing the second j, but we also take the same kind of sum, only fixing i now as the second index. This can not be represented with our previous $\bar{y}_{.j}$-type notations. Let us demonstrate these new sums with a small example. Let $N = 3$ and let $y_{ij} = y_{12}$. Then $y_{+i} = y_{+1} = y_{11} + y_{21} + y_{31}$. Similarly, $y_{j+} = y_{2+} = y_{21} + y_{22} + y_{23}$.

Now, let us turn our attention back to the three-dimensional models. To derive the optimal Within transformation for model (1), we start from its matrix form

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

To adjust the model for the no self-flow type data, the model takes the form

$$\tilde{L}'y = \tilde{L}'X\beta + \tilde{L}'D_{\alpha_*}\alpha + \tilde{L}'D_{\gamma_*}\gamma + \tilde{L}'D_\lambda\lambda + \tilde{L}'\varepsilon.$$

where $\tilde{L}$ (as above) eliminates the $i = j$ observations from the data. Now we use the Frisch-Waugh theorem to obtain the projection orthogonal to $\tilde{L}'D_{(2)}$ with $D_{(2)} \equiv (D_{\alpha_*}, D_{\gamma_*}, D_\lambda)$,

$$P_{(2)} \equiv I_{N(N-1)T} - \tilde{L}'D_{(2)}W_{(2)}^{-}D_{(2)}'\tilde{L}$$

with $W_{(2)} = D_{(2)}'\tilde{L}\tilde{L}'D_{(2)}$. From now on we have three things to do. First, get the $\tilde{L}'D_{(2)}$ for a particular (ijt) element. Second, elaborate $W_{(2)}^{-}$ (this is the hardest part), and third, get $D'_{(2)}\tilde{L}\tilde{L}'y$. Detailed calculations of $W^-_{(2)}$ is found in Appendix 2. Notice that $(e_i \otimes e_j \otimes e_t)\tilde{L}\tilde{L}'D_{(2)} = (e'_i, e'_j, e'_t)$, so

$$\tilde{y}_{ijt} = y_{ijt} - (e'_i, e'_j, e'_t)W^-_{(2)}D'_{(2)}\tilde{L}\tilde{L}'y. \tag{18}$$

Now returning to (18),

$$\begin{aligned}
&(e'_i, e'_j, e'_t)W^-_{(2)} = \\
&\quad = \frac{1}{TN(N-2)}\left[(N-1)(e'_i, e'_j, 0) + (e'_j, e'_i, 0)\right] + \left[r + s - \frac{1}{TN(N-1)}\right](l'_N, l'_N, 0) \\
&\quad + \frac{1}{TN(N-1)}(0, 0, Te'_t - l'_T) \\
&\quad = \frac{1}{TN(N-2)}\left[(N-1)(e'_i, e'_j, 0) + (e'_j, e'_i, 0)\right] - \frac{1}{2(N-1)(N-2)T}(l'_N, l'_N, 0) \\
&\quad + \frac{1}{TN(N-1)}(0, 0, Te'_t - l'_T).
\end{aligned}$$

Multiplying this with $D'_{(2)}\tilde{L}\tilde{L}'y$ leads to

$$\begin{aligned}
\tilde{y}_{ijt} = y_{ijt} &- \frac{N-1}{N(N-2)T}(y_{i++} + y_{+j+}) - \frac{1}{N(N-2)T}(y_{j++} + y_{+i+}) \\
&- \frac{1}{N(N-1)}y_{++t} + \frac{2}{N(N-2)T}y_{+++}
\end{aligned} \tag{19}$$

with

$$\begin{aligned}
y_{i++} &= \sum_{jt} y_{ijt}\,, \qquad y_{+j+} = \sum_{it} y_{ijt}\,, \qquad y_{j++} = \sum_{jt} y_{ijt,(i=j)}\,, \\
y_{+i+} &= \sum_{it} y_{ijt,(j=i)}\,, \qquad \sum_{++t} = \sum_{ij} y_{ijt}\,, \qquad y_{+++} = \sum_{ijt} y_{ijt}\,.
\end{aligned}$$

Now, let us continue with model (10). After the Within transformation (11), instead of canceled out fixed effects we end up with the following remaining fixed effects

$$\tilde{\alpha}_{it} = -\frac{1}{N(N-1)}\sum_{k=1;k\neq j}^{N} \alpha_{kt} + \frac{1}{N}\alpha_{jt}$$

and

$$\tilde{\alpha}^*_{jt} = -\frac{1}{N(N-1)}\sum_{l=1;l\neq i}^{N} \alpha^*_{lt} + \frac{1}{N}\alpha^*_{it}$$

As long as the $\tilde{\alpha}$ and $\tilde{\alpha}^*$ parameters are not zero, the Within estimators will be biased. The optimal Within transformation, following the method above is obtained in the form

$$\begin{aligned}\tilde{y}_{ijt} = y_{ijt} &- \frac{N-1}{N(N-2)}\left(y_{i+t} + y_{+jt}\right) - \frac{1}{N(N-2)}\left(y_{+it} + y_{j+t}\right) \\ &+ \frac{1}{(N-1)(N-2)} y_{++t}.\end{aligned} \tag{20}$$

with

$$y_{i+t} = \sum_j y_{ijt} \quad y_{+jt} = \sum_i y_{ijt} \quad y_{+it} = \sum_i y_{ijt,(j=i)} \quad y_{j+t} = \sum_j y_{ijt,(i=j)}$$

Lastly for model (12), the remaining fixed effects are

$$\tilde{\gamma}_{ij} = 0$$

but

$$\tilde{\alpha}_{it} = \frac{1}{N(N-1)T} \sum_{i=1;i\neq j}^{N} \sum_{t=1}^{T} \alpha_{it} + \frac{1}{NT} \sum_{t=1}^{T} \alpha_{jt} - \frac{1}{N(N-1)} \sum_{i=1;i\neq j}^{N} \alpha_{it} + \frac{1}{N} \alpha_{jt}$$

and, finally

$$\tilde{\alpha}^*_{jt} = \frac{1}{N(N-1)T} \sum_{j=1;j\neq i}^{N} \sum_{t=1}^{T} \alpha^*_{jt} + \frac{1}{NT} \sum_{t=1}^{T} \alpha^*_{it} - \frac{1}{N(N-1)} \sum_{j=1;j\neq i}^{N} \alpha^*_{jt} + \frac{1}{N} \alpha^*_{it}$$

It can be seen, as expected, these remaining fixed effects are indeed wiped out when ii type observations are present in the data. When $N \to \infty$ the remaining effects also go to zero, which implies that the bias of the Within estimators go to zero as well. Unfortunately, we have no such luck in applying the method elaborated above to find the optimal Within transformation. Instead, we can go along and use a second-best transformation which takes the form

$$\begin{aligned} y_{ijt} &- \frac{1}{N-1} y_{+jt} - \frac{1}{N-1} y_{i+t} - \frac{1}{T} y_{ij+} + \frac{1}{(N-1)^2} y_{++t} + \frac{1}{(N-1)T} y_{+j+} \\ &+ \frac{1}{(N-1)T} y_{i++} - \frac{1}{(N-1)^2 T} y_{+++} + \frac{1}{(N-1)T} y_{ji+} - \frac{1}{N-1} y_{jit} \end{aligned} \tag{21}$$

with

$$y_{ji+} = \sum_t y_{ijt,(i=j,j=i)} \qquad y_{jit} = y_{ijt,(i=j,j=i)}$$

The matrix form of the above expression is included to Appendix 3.

So overall, the self flow data problem can be overcome by using an appropriate Within transformation leading to an unbiased estimator.

Next, we can go further along the above lines and see what going is to happen if the observation sets i and j are different. If the two sets are completely disjunct, say for example if we are modeling export activity between the EU and APEC countries, intuitively enough, for all the models considered the Within estimators are unbiased, even in finite samples, as the no-self-trade problem do not arise. If the two sets are not completely disjunct, on the other hand, say for example in the case of trade between the EU and OECD countries, as the no-self-trade do arise, we are face with the same biases outlined above. Unfortunately, however, transformations (19), (20) and (21) do not work in this case, and there are no obvious transformations that could be worked out for this scenario.

3.2 Unbalanced Data

Like in the case of the usual panel data sets (see *Wansbeek and Kapteyn* [1989] or *Baltagi* [1995], for example), just more frequently, one may be faced with the situation when the data at hand is unbalanced. In our framework of analysis this means that for all the previously studied models, in general $t = 1, \ldots, T_{ij}$, $\sum_i \sum_j T_{ij} = T$ and T_{ij} is often not equal to $T_{i'j'}$. For models (5), (8) and (10) the unbalanced nature of the data does not cause any problems, the Within transformations can be used, and have exactly the same properties, as in the balanced case. However, for models (1) and (12) we are facing trouble.

In the case of model (1) and transformation (3) we get for the fixed effects the following terms (let us remember: this in fact is the optimal transformation for model (7))

$$\tilde{\alpha}_i = \frac{1}{NT} \sum_{i=1}^{N} \alpha_i \cdot (N \sum_{j=1}^{N} T_{ij} - T)$$

$$\tilde{\gamma}_j = \frac{1}{NT} \sum_{j=1}^{N} \gamma_j \cdot (N \sum_{i=1}^{N} T_{ij} - T)$$

and

$$\tilde{\lambda}_t = -\frac{1}{T_{ij}} \sum_{t=1}^{T_{ij}} \lambda_t + \frac{1}{T} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=1}^{T_{ij}} \lambda_t$$

These terms clearly do not add up to zero in general, so the Within transformation does not clear the fixed effects, as a result this Within estimator will be biased. (It can easily checked that the above $\tilde{\alpha}_i$, $\tilde{\gamma}_j$ and $\tilde{\lambda}_t$ terms add up to zero when $\forall i,j$ $T_{ij}=T$.) As (3) is the optimal Within estimator for model (7), this is bad news for the estimation of that model as well. We, unfortunately, get very similar results for transformation (4) too. The good news is, on the other hand, as seen earlier, that for model (1) transformation (11) clears the fixed effects, and although not optimal in this case, it does not depend on time, so in fact the corresponding Within estimator is still unbiased in this case.

Unfortunately, no such luck in the case of model (12) and transformation (13). The remaining fixed effects are now

$$\tilde{\gamma}_{ij} = -\frac{1}{N}\sum_{i=1}^{N}\gamma_{ij} - \frac{1}{N}\sum_{j=1}^{N}\gamma_{ij} + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\gamma_{ij} + \frac{1}{\sum_{i=1}^{N}T_{ij}}\sum_{i=1}^{N}\gamma_{ij}T_{ij} + \\ + \frac{1}{\sum_{j=1}^{N}T_{ij}}\sum_{j=1}^{N}\gamma_{ij}T_{ij} - \frac{1}{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\gamma_{ij}T_{ij}$$

$$\tilde{\alpha}_{it} = -\frac{1}{T_{ij}}\sum_{t=1}^{T_{ij}}\alpha_{it} + \frac{1}{\sum_{i=1}^{N}T_{ij}}\sum_{i=1}^{N}\sum_{t=1}^{T_{ij}}\alpha_{it} + \frac{1}{\sum_{j=1}^{N}T_{ij}}\sum_{j=1}^{N}\sum_{t=1}^{T_{ij}}\alpha_{it} - \frac{1}{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T_{ij}}\alpha_{it}$$

and

$$\tilde{\alpha}^*_{jt} = -\frac{1}{T_{ij}}\sum_{t=1}^{T_{ij}}\alpha^*_{jt} + \frac{1}{\sum_{i=1}^{N}T_{ij}}\sum_{i=1}^{N}\sum_{t=1}^{T_{ij}}\alpha^*_{jt} + \frac{1}{\sum_{j=1}^{N}T_{ij}}\sum_{j=1}^{N}\sum_{t=1}^{T_{ij}}\alpha^*_{jt} - \frac{1}{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T_{ij}}\alpha^*_{jt}$$

These terms clearly do not cancel out in general, as a result the corresponding Within estimator is biased. Unfortunately, the increase of N does not deal with the problem, so the bias remains even when $N\to\infty$. Expressions of the bias for the Within estimator of β are included in Appendix 4. It can easily be checked, however, that in the balanced case, i.e., when each $T_{ij}=T/N^2$ the fixed effects drop out indeed from the above formulations. Therefore, from a practical point of view, the estimation of both models (7) and (12) is quite problematic. However, luckily, the *Wansbeek and Kapteyn* [1989] approach can be extended to these cases. In the case of model (7), picking up the notation used in (14), $\tilde{D}_1$ and $\tilde{D}_2$ have to be modified to reflect the unbalanced nature of the data. Recall that t goes from 1 to some T_{ij}, and we assume $\sum_{ij}T_{ij}\equiv T$ and let $max\{T_{ij}\}\equiv T^*$ Then let the V_t-s be the sequence of I_{N^2} matrixes, $(t=1\ldots T^*)$ in which the following adjustments were made: for each ij observation, we leave the row (representing ij) in the first T_{ij} matrixes untouched, but delete them from the remaining $T^* - T_{ij}$ matrixes. In this way we end up with the following dummy variable setup

$$D_1 = [V_1', V_2' \dots V_{T^*}']', \; (T \times N^2)\,;$$
$$D_2^a = diag\left\{V_1 \cdot l_{N^2}, V_2 \cdot l_{N^2} \dots, V_{T^*} \cdot l_{N^2}\right\}, \; (T \times T^*);$$

So the complete dummy variable structure now is $D^a = (D_1, D_2^a)$. Let us note here, that in this case, just as in *Wansbeek and Kapteyn* [1989], index t goes "slowly" and ij "fast".

Let now

$$\Delta_{N^2} \equiv D_1' D_1\,, \qquad \Delta_{T^*} \equiv D_2^{a'} D_2^a\,, \qquad A^a \equiv D_2^{a'} D_1^a\,,$$

and

$$\bar{D}^a \equiv D_2^a - D_1 \Delta_{N^2}^{-1} A^{a'} = \left(I_T - D_1 \left(D_1' D_1\right)^{-1} D_1'\right) D_2^a$$
$$Q^a \equiv \Delta_{T^*} - A^a \Delta_{N^2}^{-1} A^{a'} = D_2^{a'} \bar{D}^a$$

Note that in the original balanced case $\Delta_{N^2} = T \cdot I_{N^2}$, $\Delta_{T^*} = N^2 \cdot I_T$ and $A^a = l_T \otimes l_{N^2}'$. So finally, the appropriate transformation for model (7) is

$$P^a = \left(I_T - D_1 \Delta_{N^2}^{-1} D_1'\right) - \bar{D}^a Q^{a-} \bar{D}^{a'} \tag{22}$$

where Q^{a-} denotes the Moore-Penrose generalized inverse, as, like in the case of *Wansbeek and Kapteyn* [1989], the Q^a matrix has no full rank. We can re-write transformation (22) using scalar notation for the ease of computation. For y let $\bar{\phi}^a \equiv Q^{a-} \bar{D}^{a'} y$. In that way, a particular element (ijt) of $P^a y$ can be written up as

$$[P^a y]_{ijt} = y_{ijt} - \frac{1}{T_{ij}} \sum_{t=1}^{T_{ij}} y_{ijt} - \bar{\phi}_t^a + \frac{1}{T_{ij}} a_{ij}^{a'} \bar{\phi}^a,$$

where a_{ij}^a is the ij-th column of matrix A^a (A^a has N^2 columns), and $\bar{\phi}_t^a$ is the t-th element of the $(T^* \times 1)$ column vector $\bar{\phi}^a$. (Note that we only have to calculate the inverse of a $(T^* \times T^*)$ matrix, which is easily doable.)

Let us continue with model (12) and let now the matrix of dummy variables for the fixed effects be $D^b = (D_1, D_2^b, D_3)$ where D_1 is defined as above,

$$D_2^b = diag\left\{U_1, \dots U_{T^*}\right\}$$

with the U_t-s being the $I_N \otimes l_N$ matrixes at time t but modified in the following way: we leave untouched the rows corresponding to observation ij in the first T_{ij} matrix, but delete them from the other $T^* - T_{ij}$ matrixes, and

$$D_3 = diag\{W_1, \ldots W_{T^*}\}$$

with the W_t-s being the $l_N \otimes I_N$ matrixes at time t, with the same modifications as above.

Defining the partial projector matrixes B and C as

$$\begin{aligned} &B \equiv I_T - D_1(D_1'D_1)^{-1}D_1' \text{ and} \\ &C \equiv B - (BD_2^b)[(BD_2^b)'(BD_2^b)]^-(BD_2^b)' \end{aligned}$$

the appropriate transformation for model (12) now is

$$P^b \equiv C - (CD_3)[(CD_3)'(CD_3)]^-(CD_3)' \tag{23}$$

It can easily be verified that P^b is idempotent and $P^bD^b = 0$, so all the fixed effects are indeed eliminated.[4]

It is worth noting that both transformations (22) and (23) are dealing in a natural way with the no-self-flow problem, as only the rows corresponding to the $i = j$ observations need to be deleted from the corresponding dummy variables matrixes (in the unbalanced case, in fact from the D_1, D_2^a and D_1, D_2^b, D_3 matrixes).

Transformation (23) can also be re-written in scalar form. First, let

$$\bar{\phi}^b \equiv \left(Q^b\right)^- \left(\bar{D}^b\right)' y \quad \text{where} \quad Q^b \equiv \left(D_2^b\right)' \bar{D}^b \quad \text{and} \quad \bar{D}^b \equiv \left(I_T - D_1(D_1'D_1)^{-1}D_1'\right) D_2^b,$$

$$\bar{\omega} \equiv \tilde{Q}^-(CD_3)'y \quad \text{where} \quad \tilde{Q} \equiv (CD_3)'(CD_3)$$

and lastly

$$\bar{\xi} \equiv \left(Q^b\right)^- \left(\bar{D}^b\right)' D_3\bar{\omega}$$

Now the scalar representation of transformation (23) is

$$\left[P^b y\right]_{ijt} = y_{ijt} - \frac{1}{T_{ij}}\sum_{t=1}^{T_{ij}} y_{ijt} + \frac{1}{T_{ij}}\left(a_{ij}^b\right)'\bar{\phi}^b - \bar{\phi}_{it}^b - \bar{\omega}_{jt} + \frac{1}{T_{ij}}\tilde{a}_{ij}'\bar{\omega} + \bar{\xi}_{it} - \frac{1}{T_{ij}}\left(a_{ij}^b\right)'\bar{\xi}$$

[4] A STATA program code for transformation (23) with a user friendly detailed explanations is available at www.personal.ceu.hu/staff/repec/pdf/stata-program_document-dofile.pdf. Estimation of model (12) is then easily doable for any kind of incompleteness. Moreover, as model (12) is the most general, the estimation of subsequent models is just a special case and requires only small modifications in the code.

where a^b_{ij} and $\tilde{a}_{ij}$ are the column vectors corresponding to observations ij from matrixes $A^b \equiv \left(D^b_2\right)' D_1$ and $\tilde{A} \equiv D_3' D_1$ respectively. $\bar{\phi}^b_{it}$ is the it-th element of the $(NT^* \times 1)$ column vector, $\bar{\phi}^b$. $\bar{\omega}_{jt}$ is the jt-th element of the $(NT^* \times 1)$ column vector, $\bar{\omega}$, and finally, $\bar{\xi}_{it}$ is the element corresponding to the it-th observation from the $(NT^* \times 1)$ column vector, $\bar{\xi}$.[5]

4. Dynamic Models

In the case of dynamic autoregressive models, the use of which is unavoidable if the data generating process has partial adjustment or some kind of memory, the Within estimators in a usual panel data framework are biased. In this section we generalize these well known results to this higher dimensional setup. We derive the semi-asymptotic bias for each of the models introduced in Section 2.

In order to show the problem, let us start with the simple linear dynamic model with bilateral interaction effects, that is model (5)

$$y_{ijt} = \rho y_{ijt-1} + \gamma_{ij} + \varepsilon_{ijt} \tag{24}$$

With backward substitution we get

$$y_{ijt} = \rho^t y_{ij0} + \frac{1-\rho^t}{1-\rho}\gamma_{ij} + \sum_{k=0}^{t} \rho^k \varepsilon_{ijt-k} \tag{25}$$

and

$$y_{ijt-1} = \rho^{t-1} y_{ij0} + \frac{1-\rho^{t-1}}{1-\rho}\gamma_{ij} + \sum_{k=0}^{t-1} \rho^k \varepsilon_{ijt-1-k}$$

What needs to be checked is the correlation between the right hand side variables of model (24) after applying the appropriate Within transformation, that is the correlation between $(y_{ijt-1} - \bar{y}_{ij.-1})$ where $\bar{y}_{ij.-1} = 1/T \sum_t y_{ijt-1}$ and $(\varepsilon_{ijt} - \bar{\varepsilon}_{ij.})$ where $\bar{\varepsilon}_{ij.} = 1/T \sum_t \varepsilon_{ijt}$. This amounts to check the correlations $(y_{ijt-1}\bar{\varepsilon}_{ij.})$, $(\bar{y}_{ij.-1}\varepsilon_{ijt})$ and $(\bar{y}_{ij.-1}\bar{\varepsilon}_{ij.})$ because $(y_{ijt-1}\varepsilon_{ijt})$ are uncorrelated. These correlations are obviously not zero in the semi-asymptotic case when $N \to \infty$, as we are facing the so called Nickell-type bias (*Nickell* [1981]). This may be the case for all other Within transformations as well.

[5] Let use make a remark here. From a computational point of view the calculation of matrix B, more precisely $D_1(D_1'D_1)^{-1}D-1'$ is by far the most resource requiring. Simplifications related to this can reduce dramatically CPU and Storage requirements.This topic, however, is beyond the limits of this paper, and the expertise of the authors.

Model (24) can of course be expanded to have exogenous explanatory variables as well

$$y_{ijt} = \rho y_{ijt-1} + x'_{ijt}\beta + \gamma_{ij} + \varepsilon_{ijt} \tag{26}$$

Let us turn now to the derivation of the semi-asymptotic bias and denote in general any of the above Within transformations by $\tilde{y}_{ijt}$. Using this notation we can derive the general form of the bias using *Nickell–type* calculations. Starting from the simple first order autoregressive model (24) introduced above we get

$$\tilde{y}_{ijt} = \rho \cdot \tilde{y}_{ijt-1} + \tilde{\varepsilon}_{ijt} \tag{27}$$

Using OLS to estimate ρ for each t period, we get

$$\widehat{\rho}_t = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} \tilde{y}_{ijt-1}\tilde{y}_{ijt}}{\sum_{i=1}^{N}\sum_{j=1}^{N} \tilde{y}^2_{ijt-1}} \tag{28}$$

So in the limit we have

$$\operatorname*{plim}_{N\to\infty} (\widehat{\rho}_t - \rho_t) = \frac{\operatorname*{plim}_{N\to\infty} \dfrac{1}{N^2}\displaystyle\sum_{i=1}^{N}\sum_{j=1}^{N} \tilde{y}_{ijt-1}\tilde{\varepsilon}_{ijt}}{\operatorname*{plim}_{N\to\infty} \dfrac{1}{N^2}\displaystyle\sum_{i=1}^{N}\sum_{j=1}^{N} \tilde{y}^2_{ijt-1}} \tag{29}$$

Continuing with model (24) and using now the appropriate (6) Within transformation we get

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

For the numerator of the bias in (29) from above we get

$$E[y_{ijt-1}\varepsilon_{ijt}] = 0$$

$$E[y_{ijt-1}\bar{\varepsilon}_{ij.}] = E\left[\left(\sum_{k=0}^{t-1}\rho^k \varepsilon_{ijt-1-k}\right)\cdot\left(\frac{1}{T}\cdot\sum_{t=1}^{T}\varepsilon_{ijt}\right)\right] = \frac{\sigma^2_\varepsilon}{T}\cdot\frac{1-\rho^{t-1}}{1-\rho}$$

$$E[\bar{y}_{ij.-1}\varepsilon_{ijt}] = E\left[\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{k=0}^{t-1}\rho^k \varepsilon_{ijt-1-k}\right)\cdot(\varepsilon_{ijt})\right] = \frac{\sigma^2_\varepsilon}{T}\cdot\frac{1-\rho^{T-t}}{1-\rho}$$

$$E[\bar{y}_{ij.-1}\bar{\varepsilon}_{ij.}] = E\left[\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{k=0}^{t-1}\rho^k \varepsilon_{ijt-1-k}\right)\cdot\left(\frac{1}{T}\cdot\sum_{t=1}^{T}\varepsilon_{ijt}\right)\right] = \frac{\sigma^2_\varepsilon}{T}\cdot\left(\frac{1}{1-\rho} - \frac{1}{T}\cdot\frac{1-\rho^T}{(1-\rho)^2}\right)$$

And for the denominator of the bias in (29)

$$E[y_{ijt-1}^2] = E\left[\left(\sum_{k=0}^{t-1}\rho^k\varepsilon_{ijt-1-k}\right)^2\right] = \sigma_\varepsilon^2\cdot\frac{1-\rho^{2t}}{1-\rho^2}$$

$$E[y_{ijt-1}\bar{y}_{ij.-1}] = E\left[\left(\sum_{k=0}^{t-1}\rho^k\varepsilon_{ijt-1-k}\right)\cdot\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{k=0}^{t-1}\rho^k\varepsilon_{ijt-1-k}\right)\right] =$$
$$= \frac{\sigma_\varepsilon^2}{T(1-\rho^2)}\left((1+\rho)\cdot\frac{1-\rho^t}{1-\rho}+\rho\cdot\frac{\rho^{T+t}-\rho^{T-t}}{1-\rho}\right)$$

$$E[\bar{y}_{ij.-1}^2] = E\left[\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{k=0}^{t-1}\rho^k\varepsilon_{ijt-1-k}\right)^2\right] =$$
$$= \frac{\sigma_\varepsilon^2}{T(1-\rho)^2}\left(1-\frac{\rho\cdot(1-\rho^T)\cdot(2+\rho-\rho^{T+1})}{T(1-\rho^2)}\right)$$

Note that in this case the probability limits define (or "estimate") the expected values (however, of course, this is not true in general). So the semi-asymptotic bias for $\widehat{\rho}_t$ is

$$\operatorname*{plim}_{N\to\infty}(\hat{\rho}_t-\rho_t) = \frac{-\frac{\sigma_\varepsilon^2}{T}\cdot\left(\frac{1-\rho^{t-1}}{1-\rho}\right)-\frac{\sigma_\varepsilon^2}{T}\cdot\left(\frac{1-\rho^{T-t}}{1-\rho}\right)+\frac{\sigma_\varepsilon^2}{T}\cdot\left(\frac{1}{1-\rho}-\frac{1}{T}\cdot\frac{1-\rho^T}{(1-\rho)^2}\right)}{\sigma_\varepsilon^2\cdot\left(\frac{1-\rho^{2t}}{1-\rho^2}\right)-A_t^1+B_t^1}$$

where

$$A_t^1 = \frac{2\sigma_\varepsilon^2}{T(1-\rho^2)}\left((1+\rho)\cdot\frac{1-\rho^t}{1-\rho}+\rho\cdot\frac{\rho^{T+t}-\rho^{T-t}}{1-\rho}\right)$$

and

$$B_t^1 = \frac{\sigma_\varepsilon^2}{T(1-\rho)^2}\left(1-\frac{\rho\cdot(1-\rho^T)\cdot(2+\rho-\rho^{T+1})}{T(1-\rho^2)}\right).$$

The bias of $\widehat{\rho}$ is simply the sum of the numerator and the denominator of the above bias formulae over time. After basic simplification it takes the form

$$\operatorname*{plim}_{N\to\infty}(\hat{\rho}-\rho) = \frac{-1+\frac{1-\rho^T}{T(1-\rho)}}{T+\frac{\rho^{2(T+1)}-1-\rho-\rho^2}{1-\rho^2}+\frac{\rho\cdot(1-\rho^T)\cdot(2+\rho-\rho^{T+1})}{T(1-\rho^2)(1-\rho)}} \tag{30}$$

It can be seen that these results are very similar to the original Nickell results.

Let us turn now our attention to model (1). In this case the Within transformation (3) leads to

$$(y_{ijt}-\bar{y}_{ij.}-\bar{y}_{..t}+\bar{y}_{...}) = \rho\cdot(y_{ijt-1}-\bar{y}_{ij.-1}-\bar{y}_{..t-1}+\bar{y}_{...-1})+(\varepsilon_{ijt}-\bar{\varepsilon}_{ij.}-\bar{\varepsilon}_{..t}+\bar{\varepsilon}_{...})$$

After lengthy derivations (see Appendix 5) we get for the semi-asymptotic bias for each t

$$\operatorname*{plim}_{N\to\infty}(\hat{\rho}_t-\rho_t)=\frac{\operatorname*{plim}_{N\to\infty}\left(\frac{1-N^2}{N^2}\right)\frac{\sigma_\varepsilon^2}{T}\frac{1-\rho^{t-1}}{1-\rho}+\operatorname*{plim}_{N\to\infty}\left(\frac{1-N^2}{N^2}\right)\frac{\sigma_\varepsilon^2}{T}\frac{1-\rho^{T-t}}{1-\rho}+A_t^2}{\operatorname*{plim}_{N\to\infty}\left(\frac{N^2-1}{N^2}\right)\cdot\sigma_\varepsilon^2\frac{1-\rho^{2t}}{1-\rho^2}-B_t^2+C_t^2}$$

where

$$A_t^2=\operatorname*{plim}_{N\to\infty}\left(\frac{N^2-1}{N^2}\right)\frac{\sigma_\varepsilon^2}{T}\left(\frac{1}{1-\rho}-\frac{1}{T}\frac{1-\rho^T}{(1-\rho)^2}\right)$$

$$B_t^2=\operatorname*{plim}_{N\to\infty}2\left(\frac{N^2-1}{N^2}\right)\cdot\frac{\sigma_\varepsilon^2}{T(1-\rho^2)}\left((1+\rho)\cdot\frac{1-\rho^t}{1-\rho}+\rho\cdot\frac{\rho^{T+t}-\rho^{T-t}}{1-\rho}\right)$$

and

$$C_t^2=\operatorname*{plim}_{N\to\infty}\left(\frac{N^2-1}{N^2}\right)\cdot\frac{\sigma_\varepsilon^2}{T(1-\rho)^2}\left(1-\frac{\rho\cdot(1-\rho^T)\cdot(2+\rho-\rho^{T+1})}{T(1-\rho^2)}\right)$$

It is easy to see then that

$$\operatorname*{plim}_{N\to\infty}(\hat{\rho}_t-\rho_t)=\frac{-\frac{\sigma_\varepsilon^2}{T}\cdot\left(\frac{1-\rho^{t-1}}{1-\rho}\right)-\frac{\sigma_\varepsilon^2}{T}\cdot\left(\frac{1-\rho^{T-t}}{1-\rho}\right)+\frac{\sigma_\varepsilon^2}{T}\cdot\left(\frac{1}{1-\rho}-\frac{1}{T}\cdot\frac{1-\rho^T}{(1-\rho)^2}\right)}{\sigma_\varepsilon^2\cdot\left(\frac{1-\rho^{2t}}{1-\rho^2}\right)-A_t^1+B_t^1},$$

the same as for model (24). It has to be the case then that the semi-asymptotic bias of $\widehat{\rho}$ is actually (30).

As seen earlier, the optimal Within transformation for model (1) is in fact (4)

$$(y_{ijt}-\bar{y}_{i..}-\bar{y}_{.j.}-\bar{y}_{..t}+2\bar{y}_{...})$$

For this Within estimator the bias is (see the derivation in Appendix 5)

$$\operatorname*{plim}_{N\to\infty}(\hat{\rho}_t-\rho_t)=\frac{\operatorname*{plim}_{N\to\infty}\left(\frac{2-2N}{N^2}\right)\cdot\frac{\sigma_\varepsilon^2}{T}\cdot\frac{1-\rho^{t-1}}{1-\rho}+\operatorname*{plim}_{N\to\infty}\left(\frac{2-2N}{N^2}\right)\cdot\frac{\sigma_\varepsilon^2}{T}\cdot\frac{1-\rho^{T-t}}{1-\rho}+A_t^3}{\operatorname*{plim}_{N\to\infty}\left(\frac{N^2-1}{N^2}\right)\cdot\sigma_\varepsilon^2\frac{1-\rho^{2t}}{1-\rho^2}-B_t^3+C_t^3}$$

where

$$A_t^3=\operatorname*{plim}_{N\to\infty}\left(\frac{2N-2}{N^2}\right)\cdot\frac{\sigma_\varepsilon^2}{T}\cdot\left(\frac{1}{1-\rho}-\frac{1}{T}\cdot\frac{1-\rho^T}{(1-\rho)^2}\right)$$

$$B_t^3 = \operatorname*{plim}_{N\to\infty} \left(\frac{4N-4}{N^2}\right) \cdot \frac{\sigma_\varepsilon^2}{T(1-\rho^2)} \left((1+\rho)\cdot\frac{1-\rho^t}{1-\rho} + \rho\cdot\frac{\rho^{T+t}-\rho^{T-t}}{1-\rho}\right)$$

and

$$C_t^3 = \operatorname*{plim}_{N\to\infty} \left(\frac{2N-4}{N^2}\right) \cdot \frac{\sigma_\varepsilon^2}{T(1-\rho)^2} \left(1 - \frac{\rho\cdot(1-\rho^T)\cdot(2+\rho-\rho^{T+1})}{T(1-\rho^2)}\right)$$

It is trivial to see that the bias goes to zero, so this estimator is semi-asymptotically unbiased (unlike the previous one).

As the optimal Within transformation for model (7) is in fact (3), we get the same bias in this case as for model (1).

Let us now continue with models (8) and (10) which can be considered as the same models from this point of view. Writing up model (8)

$$y_{ijt} = \rho y_{ijt-1} + \alpha_{jt} + \varepsilon_{ijt}$$

and applying the within transformation to it we get

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

The semi-asymptotic bias then can be expressed as

$$\operatorname*{plim}_{N\to\infty}(\hat{\rho}_t - \rho_t) = \frac{\operatorname*{plim}_{N\to\infty} \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}(y_{ijt-1}-\bar{y}_{.jt-1})(\varepsilon_{ijt}-\bar{\varepsilon}_{.jt})}{\operatorname*{plim}_{N\to\infty} \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}(y_{ijt-1}-\bar{y}_{.jt-1})^2}$$

It can easily be seen that the expressions in the numerator are zero, as both y_{ijt-1} and $\bar{y}_{jt-1}$ depend on the ε-s only up to time $t-1$, and are necessarily uncorrelated with the t-th period disturbance, ε_{ijt}. So as the denominator is finite, the semi-asymptotic bias is in fact nil. The same arguments are valid for model (10) as well.

Finally, let us turn to model (12)

$$y_{ijt} = \rho y_{ijt-1} + \gamma_{ij} + \alpha_{it} + \alpha_{jt}^* + \varepsilon_{ijt}$$

The Within transformation is

$$(y_{ijt} - \bar{y}_{ij.} - \bar{y}_{.jt} - \bar{y}_{i.t} + \bar{y}_{..t} + \bar{y}_{.j.} + \bar{y}_{i..} - \bar{y}_{...})$$

so we get

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

And for the semi-asymptotic bias of this model we get

$$
\operatorname*{plim}_{N\to\infty} (\hat{\rho}_t - \rho_t) = \frac{\operatorname*{plim}_{N\to\infty} \left(\frac{-(N-1)^2}{N^2} \right) \cdot \frac{\sigma_\varepsilon^2}{T} \cdot \frac{1-\rho^{t-1}}{1-\rho} + \operatorname*{plim}_{N\to\infty} \left(\frac{-(N-1)^2}{N^2} \right) \cdot \frac{\sigma_\varepsilon^2}{T} \cdot \frac{1-\rho^{T-t}}{1-\rho} + A_t^4}{\operatorname*{plim}_{N\to\infty} \left(\frac{(N-1)^2}{N^2} \right) \sigma_\varepsilon^2 \frac{1-\rho^{2t}}{1-\rho^2} - B_t^4 + C_t^4}
$$

where

$$
A_t^4 = \operatorname*{plim}_{N\to\infty} \left(\frac{(N-1)^2}{N^2} \right) \cdot \frac{\sigma_\varepsilon^2}{T} \cdot \left(\frac{1}{1-\rho} - \frac{1}{T} \cdot \frac{1-\rho^T}{(1-\rho)^2} \right)
$$

$$
B_t^4 = \operatorname*{plim}_{N\to\infty} \left(\frac{2(N-1)^2}{N^2} \right) \cdot \frac{\sigma_\varepsilon^2}{T(1-\rho^2)} \left((1+\rho) \cdot \frac{1-\rho^t}{1-\rho} + \rho \cdot \frac{\rho^{T+t} - \rho^{T-t}}{1-\rho} \right)
$$

and

$$
C_t^4 = \operatorname*{plim}_{N\to\infty} \left(\frac{(N-1)^2}{N^2} \right) \cdot \frac{\sigma_\varepsilon^2}{T(1-\rho)^2} \left(1 - \frac{\rho \cdot (1-\rho^T) \cdot (2 + \rho - \rho^{T+1})}{T(1-\rho^2)} \right)
$$

It is clear that this is in fact the semi-asymptotic bias of (24) for each t

$$
\operatorname*{plim}_{N\to\infty} (\hat{\rho}_t - \rho_t) = \frac{-\frac{\sigma_\varepsilon^2}{T} \cdot \left(\frac{1-\rho^{t-1}}{1-\rho} \right) - \frac{\sigma_\varepsilon^2}{T} \cdot \left(\frac{1-\rho^{T-t}}{1-\rho} \right) + \frac{\sigma_\varepsilon^2}{T} \cdot \left(\frac{1}{1-\rho} - \frac{1}{T} \cdot \frac{1-\rho^T}{(1-\rho)^2} \right)}{\sigma_\varepsilon^2 \cdot \left(\frac{1-\rho^{2t}}{1-\rho^2} \right) - A_t^1 + B_t^1}
$$

necessarily leading to the same bias formulae (30) for $\widehat{\rho}$.

As seen above, we have problems with the estimation and N inconsistency of models (5), (7) and (12) in the dynamic case (see Table 2). Luckily, many of the well known instrumental variables (IV) estimators developed to deal with dynamic panel data models can be generalized to these higher dimensions as well, as the number of available orthogonality conditions increases together with the dimensions. Let us take the example of one of the most frequently used, the Arellano and Bond IV estimator (see *Arellano and Bond* [1991] and *Harris, Matyas and Sevetre* [2005] p. 260) for the estimation of model (5).

The model is written up in first differences, such as

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

or

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

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

$$z_{ij} = \begin{pmatrix} y_{ij1} & 0 & \cdots & \cdots & 0 & \cdots & 0 \\ 0 & y_{ij1} & y_{ij2} & 0 & 0 & \cdots & 0 \\ \vdots & \cdots & & \vdots & \vdots & \cdots & \vdots \\ 0 & \cdots & 0 & 0 & y_{ij1} & \cdots & y_{ijT-2} \end{pmatrix}_{\left((T-2)\times \frac{(T-1)(T-2)}{2}\right)}$$

The resulting IV estimator of ρ is

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

where ΔY and ΔY_{-1} are the panel first differences, $Z_{AB} = [z_{11}', z_{12}', \ldots, z_{NN}']'$ and $\Omega = I_{N^2} \otimes \Sigma$ is the covariance matrix, with known form

$$\Sigma = \begin{pmatrix} 2 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & -1 & 2 & -1 \\ 0 & \cdots & 0 & -1 & 2 \end{pmatrix}_{((T-2)\times(T-2))}$$

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

In the case of models (7) and (12) to derive an Arellano-Bond type estimator we need to insert one further step. After taking the first differences, we implement a simple transformation in order to get to a model with only (ij) pairwise interaction effects, exactly as in model (5). Then we proceed as above as the Z_{AB} instruments are going to be valid for these transformed models as well. Let us start with model (7) and take the first differences

$$(y_{ijt} - y_{ijt-1}) = \rho\,(y_{ijt-1} - y_{ijt-2})) + (\lambda_t - \lambda_{t-1}) + (\varepsilon_{ijt} - \varepsilon_{ijt-1})$$

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

$$(y_{ijt} - y_{ijt-1}) - \frac{1}{N}\sum_{i=1}^{N}(y_{ijt} - y_{ijt-1}) = \rho\left[(y_{ijt-1} - y_{ijt-2}) - \frac{1}{N}\sum_{i=1}^{N}(y_{ijt-1} - y_{ijt-2})\right] +$$
$$+\left[(\lambda_t - \lambda_{t-1}) - \frac{1}{N}\sum_{i=1}^{N}(\lambda_t - \lambda_{t-1})\right] + \left[(\varepsilon_{ijt} - \varepsilon_{ijt-1}) - \frac{1}{N}\sum_{i=1}^{N}(\varepsilon_{ijt} - \varepsilon_{ijt-1})\right]$$

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

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

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

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

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

Continuing now with model (12), the transformation needed in this case is

$$\Delta y_{ijt} - \frac{1}{N}\sum_{i=1}^{N}\Delta y_{ijt} - \frac{1}{N}\sum_{j=1}^{N}\Delta y_{ijt} + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\Delta y_{ijt} =$$
$$= \rho\left[\Delta y_{ijt-1} - \frac{1}{N}\sum_{i=1}^{N}\Delta y_{ijt-1} - \frac{1}{N}\sum_{j=1}^{N}\Delta y_{ijt-1} + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\Delta y_{ijt-1}\right] +$$
$$+\left[\Delta\alpha_{it} - \frac{1}{N}\sum_{i=1}^{N}\Delta\alpha_{it} - \frac{1}{N}\sum_{j=1}^{N}\Delta\alpha_{it} + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\Delta\alpha_{it}\right] +$$
$$+\left[\Delta\alpha_{jt} - \frac{1}{N}\sum_{i=1}^{N}\Delta\alpha_{jt} - \frac{1}{N}\sum_{j=1}^{N}\Delta\alpha_{jt} + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\Delta\alpha_{jt}\right] +$$
$$+\left[\Delta\varepsilon_{ijt} - \frac{1}{N}\sum_{i=1}^{N}\Delta\varepsilon_{ijt} - \frac{1}{N}\sum_{j=1}^{N}\Delta\varepsilon_{ijt} + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\Delta\varepsilon_{ijt}\right]$$

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

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

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

5. Extensions to Higher Dimensions

Let us assume that we would like to study the volume of exports y from a given country to countries i, for some products j by firms s at time t. This would result in four dimensional observations for our variable of interest y_{ijst}, $i = 1, \ldots, N_i$, $j = 1, \ldots, N_j$, $s = 1, \ldots, N_s$ and, in the balanced case $t = 1, \ldots, T$. If the data is about trade between two countries, for example, at product/sector level, then $N_i = N_j = N$. If the data at hand is not only for a given country, but for several, with product and firm observations, then we would end up with a five dimensional panel data, and so on. In order to analyse the higher dimensional setup, let us use the all encompassing model (12), (14) with pair-wise interaction effects:

$$y_{ijst} = x'_{ijst}\beta + \gamma^0_{ijs} + \gamma^1_{ijt} + \gamma^2_{jst} + \gamma^3_{ist} + \varepsilon_{ijst} \tag{31}$$

The fixed effects of this model in a more compact and general form are

$$\gamma^0_{IS} + \sum_{k=1}^{M} \gamma^M_{i_k,t} \tag{32}$$

where i_k is any pair-wise, combination of the individual index-set IS, in the above case $IS = (i, j, s)$, and M is the number of such pair-wise combinations (in (31) $M = 3$). In the case of unbalanced panel data $t = 1, \ldots, T_{IS}$.

The Within transformation for model (31) is

$$\begin{aligned}(y_{ijst} - \bar{y}_{.jst} - \bar{y}_{i.st} - \bar{y}_{ij.t} - \bar{y}_{ijs.} + \bar{y}_{..st} + \bar{y}_{.j.t} + \bar{y}_{.js.} \\ + \bar{y}_{i..t} + \bar{y}_{i.s.} + \bar{y}_{ij..} - \bar{y}_{...t} - \bar{y}_{..s.} - \bar{y}_{.j..} - \bar{y}_{i...} + \bar{y}_{....})\end{aligned}$$

where

$$\bar{y}_{....} = \frac{1}{N_i N_j N_s T} \sum_{i=1}^{N_i} \sum_{j=1}^{N_j} \sum_{s=1}^{N_s} \sum_{t=1}^{T} y_{ijst}$$

$$\bar{y}_{i...} = \frac{1}{N_j N_s T} \sum_{j=1}^{N_j} \sum_{s=1}^{N_s} \sum_{t=1}^{T} y_{ijst}$$

$$\bar{y}_{ij..} = \frac{1}{N_s T} \sum_{s=1}^{N_s} \sum_{t=1}^{T} y_{ijst}$$

$$\bar{y}_{ijs.} = \frac{1}{T} \sum_{t=1}^{T} y_{ijst}$$

All other terms behave analogously, i.e., we take averages with respect to those indexes from (i, j, s, t), which do not appear in the subscript of $\bar{y}$. In matrix form

$$\begin{aligned}
P_D = & \left(I_{N_i} - \bar{J}_{N_i}\right) \otimes \left(I_{N_j} - \bar{J}_{N_j}\right) \otimes \left(I_{N_s} - \bar{J}_{N_s}\right) \otimes \left(I_T - \bar{J}_T\right) = \\
& \left(I_{N_i N_j N_s T} - \left(\bar{J}_{N_i} \otimes I_{N_j N_s T}\right) - \left(I_{N_i} \otimes \bar{J}_{N_j} \otimes I_{N_s T}\right) - \left(I_{N_i N_j} \otimes \bar{J}_{N_s} \otimes I_T\right)\right. \\
& - \left(I_{N_i N_j N_s} \otimes \bar{J}_T\right) + \left(\bar{J}_{N_i N_j} \otimes I_{N_s T}\right) + \left(\bar{J}_{N_i} \otimes I_{N_j} \otimes \bar{J}_{N_s} \otimes I_T\right) \\
& + \left(\bar{J}_{N_i} \otimes I_{N_j N_s} \otimes \bar{J}_T\right) + \left(I_{N_i} \otimes \bar{J}_{N_j N_s} \otimes I_T\right) + \left(I_{N_i} \otimes \bar{J}_{N_j} \otimes I_{N_s} \otimes \bar{J}_T\right) \\
& + \left(I_{N_i N_j} \otimes \bar{J}_{N_s T}\right) - \left(\bar{J}_{N_i N_j N_s} \otimes I_T\right) - \left(\bar{J}_{N_i N_j} \otimes I_{N_s} \otimes \bar{J}_T\right) - \left(\bar{J}_{N_i} \otimes I_{N_j} \otimes \bar{J}_{N_s T}\right) \\
& \left. - \left(I_{N_j} \otimes \bar{J}_{N_j N_s T}\right) + \bar{J}_{N_i N_j N_s T}\right)
\end{aligned} \tag{33}$$

The generalization of the Within transformation for model (31) for any higher dimensions can be done using the general form (32). There are basically two types of fixed effect, γ_{IS}^0, depending on all indices except t, and the rest, which are symmetric in a sense, since all consist two indices from IS and t. Let us see the method for γ_{IS}^0, and then for a representative fixed effect, from the other group, let it be γ_{ijt}.

Let denote $IS \cup \{t\}$ by IS', and its elements by $s_1, \ldots s_M$ (in the three-dimensional case $s_1 = i$, $s_2 = j$ and $s_3 = t$). The Within transformation then is

$$\left(y_{IS'} - \sum_{i=1}^{M} \tilde{y}_{s_i} + \sum_{i=1}^{M} \sum_{j=1}^{M} \tilde{y}_{s_i s_j} - \sum_{i=1}^{M} \sum_{j=1; i \neq j}^{M} \sum_{k=1; k \neq i,j}^{M} \tilde{y}_{s_i s_j s_k} + \ldots \ldots \pm \tilde{y}_{IS'}\right)$$

where

$$\tilde{y}_{s_{i_1} s_{i_2} \ldots s_{i_m}} = \frac{1}{N_{s_{i_1}} \ldots N_{s_{i_m}}} \sum_{s_{i_1}=1, \ldots s_{i_m}=1}^{N_{s_{i_1}}, \ldots, N_{s_{i_m}}} y_{IS'}$$

The method in fact is the following. First, we subtract the first order sums with respect to each variables from the original untransformed variable $y_{IS'}$. Then we add up the second order sums in every possible pair-wise combination, then subtract the third order sums, and so on. The sum with respect to t equals to γ^0_{IS}, clearing it out. All other first order sums still remain. In the next step we add the second order sums. All the previously remaining terms appear additionally summed with respect to t, but with an opposite sign, canceling out all the remaining terms from period 1. Continuing the process, all the remaining terms in period i appear in the next one, also summed with respect to t, and with an opposite sign, again clearing out all the terms from period i. The induction should now be clear. In the last but one period, the only remaining term is going to be the sum with respect to all indices but t, with a sign determined by the parity of the indices. In the last period, we are summing up γ^0_{IS} with respect to all indices including t, but with an opposite sign, which therefore cancels out the only previously remaining term.

It can be shown easily, that the properties of the Within estimator based on transformation (33) in the case of no-self-flow, unbalanced data and dynamic models are exactly the same as seen earlier for the three dimensional model.

In the case of no-self-flow transformation (21) can also be generalized to any higher dimensions. In the four dimensional case (with $N_i = N_j = N$) we get

$$\begin{aligned}
&(y_{ijst} - \frac{1}{N-1}y_{+jst} - \frac{1}{N-1}y_{i+st} - \frac{1}{N_s}y_{ij+t} - \frac{1}{T}y_{ijs+} + \frac{1}{(N-1)^2}y_{++st} \\
&\quad + \frac{1}{(N-1)N_s}y_{+j+t} + \frac{1}{(N-1)T}y_{+js+} + \frac{1}{(N-1)N_s}y_{i++t} + \frac{1}{(N-1)T}y_{i+s+} \\
&\quad + \frac{1}{N_sT}y_{ij++} - \frac{1}{(N-1)^2N_s}y_{+++t} - \frac{1}{(N-1)^2T}y_{++s+} - \frac{1}{(N-1)N_sT}y_{+j++} \\
&\quad - \frac{1}{(N-1)N_sT}y_{i+++} + \frac{1}{(N-1)^2N_sT}y_{++++} - \frac{1}{(N-1)N_sT}y_{ji++} \\
&\quad + \frac{1}{(N-1)T}y_{jis+} + \frac{1}{(N-1)N_s}y_{ji+t} - \frac{1}{N-1}y_{jist})
\end{aligned} \tag{34}$$

with

$$y_{ji++} = \sum_{st} y_{ijst,(i=j,j=i)} \quad y_{jis+} = \sum_{t} y_{ijst,(i=j,j=i)}$$
$$y_{ji+t} = \sum_{s} y_{ijst,(i=j,j=i)} \quad y_{jist} = y_{ijst,(i=j,j=i)}$$

The matrix form of the above expression is included in Appendix 6.

We can go further along the above lines, and generalize transformation (23) into any higher dimensional setup. Using the dummy variable structure (32), let the dummy variables matrixes for the $M+1$ fixed effects be denoted by $D_1^c, D_2^c, \ldots D_{M+1}^c$ respectively, and let $P^{(k)}$ be the transformation which clears out the first k fixed effects, namely $P^{(k)} \cdot (D_1^c, D_2^c, \ldots, D_k^c) = (0, 0, \ldots, 0)$. The appropriate within transformation to clear out the first $k+1$ fixed effects then is

$$P^{(k+1)} = P^{(k)} - \left(P^{(k)} D_{k+1}^c\right) \left[\left(P^{(k)} D_{k+1}^c\right)' \left(P^{(k)} D_{k+1}^c\right)\right]^{-} \left(P^{(k)} D_{k+1}^c\right)'$$

where

$$P^{(1)} = I_T - D_1^c \left((D_1^c)' D_1^c\right)^{-1} (D_1^c)'$$

Now, it should be clear from the induction, that the appropriate within transformation for model (31)–(32) is

$$P^c = P^{(M+1)} = P^{(M)} - \left(P^{(M)} D_{M+1}^c\right) \left[\left(P^{(M)} D_{M+1}^c\right)' \left(P^{(M)} D_{M+1}^c\right)\right]^{-} \left(P^{(M)} D_{M+1}^c\right)' \tag{35}$$

6. Further Extensions

We assumed so far throughout the paper that the idiosyncratic disturbance term ε is in fact a well behaved white noise, that is, all heterogeneity is introduced into the model through the fixed effects. In some applications this may be an unrealistic assumption, so next we relax it in two ways. We introduce heteroscedasticity and a simple form of cross correlation into the disturbance terms, and see how this impacts on the transformations introduced earlier. In fact, in our context, there are two different ways to deal with this issue. The standard procedure would be to first transform the model in order to get a model with a scalar covariance matrix and then, in a second transformation, eliminate the fixed effects. Unfortunately, this would lead to quite complicated formulae, very difficult to implement in practice. We therefore propose a second option by inverting the order of the transformations. First, eliminate the fixed effects and then estimate the model by Feasible GLS (FGLS). This is leading to formulae much easier to handle.

The first step is to derive the covariance matrix of the model and analyze how the different transformations introduced earlier impact on it. Then, we derive estimators for the variance components of the transformed model, in order to be able to use FGLS instead of OLS for the estimation.

6.1 Covariance Matrixes and the Within Transformations

The initial Assumption (2) about the disturbance terms now is replaced by

$$E\left(\varepsilon_{ij}\varepsilon_{kl}\right)=\begin{cases}\sigma^2_{\varepsilon i} & \text{if } i=k,\ j=l,\ \forall t\\ \rho_{(1)} & \text{if } i=k,\ j\neq l,\ \forall t\\ \rho_{(2)} & \text{if } i\neq k,\ j=l,\ \forall t\\ 0 & \text{otherwise}\end{cases}$$

Then the covariance matrix of all models introduced in Section 2 takes the form

$$\Upsilon\equiv L_N\otimes I_{NT}-\left(\rho_{(1)}+\rho_{(2)}\right)\cdot I_{N^2T}+\rho_{(1)}\cdot I_N\otimes J_N\otimes I_T+\rho_{(2)}\cdot J_N\otimes I_N\otimes I_T,$$

where

$$L_N\equiv\begin{pmatrix}\sigma^2_{\varepsilon 1} & 0 & \cdots & 0\\ 0 & \sigma^2_{\varepsilon 2} & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & \sigma^2_{\varepsilon N}\end{pmatrix}_{N\times N}$$

This covariance matrix is altered, depending on the Within transformation used to get rid of the fixed effects.

In the case of transformation (3) the P_D projection matrix is

$$P_D=I_{N^2T}-\frac{1}{T}I_{N^2}\otimes J_T-\frac{1}{N^2}J_{N^2}\otimes I_T+\frac{1}{N^2T}J_{N^2T}$$

and we get

$$\begin{aligned}P_D\Upsilon P_D=\Upsilon&-\frac{1}{T}L_N\otimes I_N\otimes J_T+\frac{1}{T}(\rho_{(1)}+\rho_{(2)})\cdot I_{N^2}\otimes J_T\\&-\frac{1}{N^2}L_NJ_N\otimes J_N\otimes I_T-\frac{1}{N^2}\left((N-1)\rho_{(1)}+(N-1)\rho_{(2)}\right)\cdot J_{N^2}\otimes I_T\\&+\frac{1}{N^2T}L_NJ_N\otimes J_{NT}+\frac{1}{N^2T}\left((N-1)\rho_{(1)}+(N-1)\rho_{(2)}\right)\cdot J_{N^2T}\\&-\frac{\rho_{(1)}}{T}\cdot I_N\otimes J_{NT}-\frac{\rho_{(2)}}{T}\cdot J_N\otimes I_N\otimes J_T\end{aligned}$$

For transformation (4) we get

$$P_D=I_{N^2T}-\frac{1}{NT}I_N\otimes J_{NT}-\frac{1}{NT}J_N\otimes I_N\otimes J_T-\frac{1}{N^2}J_{N^2}\otimes I_T+\frac{2}{N^2T}J_{N^2T}$$

and

$$P_D \Upsilon P_D = \Upsilon - \frac{1}{NT} L_N \otimes J_{NT} - \frac{1}{NT}\left((N-1)\rho_{(1)} - \rho_{(2)}\right) \cdot I_N \otimes J_{NT}$$
$$- \frac{1}{NT} L_N J_N \otimes I_N \otimes J_T - \frac{1}{NT}\left(-\rho_{(1)} + (N-1)\rho_{(2)}\right) \cdot J_N \otimes I_N \otimes J_T$$
$$- \frac{1}{N^2} L_N J_N \otimes J_N \otimes I_T - \frac{1}{N^2}\left((N-1)\rho_{(1)} + (N-1)\rho_{(2)}\right) \cdot J_{N^2} \otimes I_T$$
$$+ \frac{2}{N^2 T} L_N J_N \otimes J_T + \frac{1}{N^2 T}\left((N-2)\rho_{(1)} + (N-2)\rho_{(2)}\right) \cdot J_{N^2 T}$$

For transformation (6) we have

$$P_D = I_{N^2T} - \frac{1}{T} I_{N^2} \otimes J_T$$

and

$$P_D \Upsilon P_D = \Upsilon - \frac{1}{T}\left(L_N \otimes I_N \otimes J_T\right) + \frac{1}{T}(\rho_{(1)} + \rho_{(2)}) \cdot I_{N^2} \otimes J_T - \frac{\rho_{(1)}}{T} \cdot I_N \otimes J_{NT}$$
$$- \frac{\rho_{(2)}}{T} \cdot J_N \otimes I_N \otimes J_T$$

For transformation (9) we get

$$P_D = I_{N^2T} - \frac{1}{N} J_N \otimes I_{NT}$$

and

$$P_D \Upsilon P_D = \Upsilon - \frac{1}{N}\left(L_N J_N \otimes I_{NT}\right) + \frac{1}{N}(\rho_{(1)} + (1-N)\rho_{(2)}) \cdot J_N \otimes I_{NT}$$
$$- \frac{\rho_{(1)}}{N} \cdot J_{N^2} \otimes I_T$$

For transformation (11) we get

$$P_D = I_{N^2T} - \frac{1}{N} J_N \otimes I_{NT} - \frac{1}{N} I_N \otimes J_N \otimes I_T + \frac{1}{N^2} J_{N^2} \otimes I_T$$

and

$$P_D \Upsilon P_D = \Upsilon - \frac{1}{N} L_N J_N \otimes I_{NT} - \frac{1}{N}\left(-\rho_{(1)} + (N-1)\rho_{(2)}\right) \cdot J_N \otimes I_{NT}$$
$$- \frac{1}{N} L_N \otimes J_N \otimes I_T - \frac{1}{N}\left((N-1)\rho_{(1)} - \rho_{(2)}\right) \cdot I_N \otimes J_N \otimes I_T$$
$$+ \frac{1}{N^2 T} L_N J_N \otimes J_N \otimes I_T + \frac{1}{N^2}\left(-\rho_{(1)} - \rho_{(2)}\right) \cdot J_{N^2} \otimes I_T$$

And finally, for transformation (13) we get

$$\begin{aligned}P_D = I_{N^2T} - \frac{1}{N}J_N \otimes I_{NT} - \frac{1}{N}I_N \otimes J_N \otimes I_T - \frac{1}{T}I_{N^2} \otimes J_T \\ + \frac{1}{NT}J_N \otimes I_N \otimes J_T + \frac{1}{NT}I_N \otimes J_{NT} + \frac{1}{N^2}J_{N^2} \otimes I_T - \frac{1}{N^2T}J_{N^2T}\end{aligned}$$

and

$$\begin{aligned}P_D \Upsilon P_D = \Upsilon &- \frac{1}{N}L_N J_N \otimes I_{NT} - \frac{1}{N}\left(-\rho_{(1)} + (N-1)\rho_{(2)}\right) \cdot J_N \otimes I_{NT} \\ &- \frac{1}{N}L_N \otimes J_N \otimes I_T - \frac{1}{N}\left((N-1)\rho_{(1)} - \rho_{(2)}\right) \cdot I_N \otimes J_N \otimes I_T \\ &+ \frac{1}{N^2}L_N J_N \otimes J_N \otimes I_T + \frac{1}{N^2}\left(-\rho_{(1)} - \rho_{(2)}\right) \cdot J_{N^2} \otimes I_T \\ &- \frac{1}{T}L_N \otimes I_N \otimes J_T - \frac{1}{T}\left(-\rho_{(1)} - \rho_{(2)}\right) \cdot I_{N^2} \otimes J_T \\ &+ \frac{1}{NT}L_N \otimes J_{NT} + \frac{1}{NT}\left(-\rho_{(1)} - \rho_{(2)}\right) \cdot I_N \otimes J_{NT} \\ &+ \frac{1}{NT}L_N J_N \otimes I_N \otimes J_T + \frac{1}{NT}\left(-\rho_{(1)} - \rho_{(2)}\right) \cdot J_N \otimes I_N \otimes J_T \\ &- \frac{1}{N^2T}L_N J_N \otimes J_{NT} - \frac{1}{N^2T}\left(-\rho_{(1)} - \rho_{(2)}\right) \cdot J_{N^2T}\end{aligned}$$

6.2 Estimation of the Variance Components and the Cross Correlations

What now remains to be done is to estimate to variance components in order to make the GLS feasible. However, as we are going to see, some difficulties lay ahead. Let us start with the simplest case, model (5). Applying transformation (6) leads to the following model to be estimated

$$(y_{ijt} - \bar{y}_{ij}) = (x_{ijt} - \bar{x}_{ij.})'\beta + (\varepsilon_{ijt} - \bar{\varepsilon}_{ij.})$$

Let us denote the transformed disturbance terms by u_{ijt}. In this way now

$$E\left[u_{ijt}^2\right] = E\left[(\varepsilon_{ijt} - \bar{\varepsilon}_{ij.})^2\right] = \frac{T-1}{T}\sigma_{\varepsilon i}^2$$

These are in fact N equations for N unknown parameters, so the system can be solved:

$$\frac{1}{NT}\sum_{j=1}^{N}\sum_{t=1}^{T}\hat{u}_{ijt}^2 = \frac{T-1}{T}\hat{\sigma}_{\varepsilon i}^2$$

$$\hat{\sigma}_{\varepsilon i}^2 = \frac{1}{N(T-1)}\sum_{j=1}^{N}\sum_{t=1}^{T}\hat{u}_{ijt}^2$$

where $\hat{u}$ is the OLS residual from the estimation of the transformed model. We also have to estimate the two cross correlations, $\rho_{(1)}$ and $\rho_{(2)}$. This is done by taking the averages of the residuals with respect to i and j. Let us start with $\rho_{(1)}$

$$E\left[\left(\frac{1}{N}\sum_{j=1}^{N}u_{ijt}\right)^2\right] = E\left[\bar{u}_{i.t}^2\right] = \frac{T-1}{NT}\sigma_{\varepsilon i}^2 + \frac{(N-1)(T-1)}{NT}\rho_{(1)}$$

As we already have an estimator for $\sigma_{\varepsilon i}^2$,

$$\begin{aligned}\hat{\rho}_{(1)} &= \frac{NT}{(N-1)(T-1)}\left(E\left[\bar{u}_{i.t}^2\right] - \frac{T-1}{NT}\sum_{i=1}^{N}\hat{\sigma}_{\varepsilon i}^2\right) = \\ &= \frac{1}{N^2(N-1)(T-1)}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\sum_{j=1}^{N}\hat{u}_{ijt}\right)^2 - \frac{1}{N(N-1)}\sum_{i=1}^{N}\hat{\sigma}_{\varepsilon i}^2\end{aligned}$$

Now for $\rho_{(2)}$,

$$E\left[\left(\frac{1}{N}\sum_{i=1}^{N}u_{ijt}\right)^2\right] = E\left[\bar{u}_{.jt}\right] = \frac{T-1}{N^2T}\sum_{i=1}^{N}\sigma_{\varepsilon i}^2 + \frac{(N-1)(T-1)}{NT}\rho_{(2)},$$

and so

$$\begin{aligned}\hat{\rho}_{(2)} &= \frac{NT}{(N-1)(T-1)}\left(E\left[\hat{\bar{u}}_{jt}^2\right] - \frac{T-1}{N^2T}\sum_{i=1}^{N}\hat{\sigma}_{\varepsilon i}^2\right) = \\ &= \frac{1}{N^2(N-1)(T-1)}\sum_{j=1}^{N}\sum_{t=1}^{T}\left(\sum_{i=1}^{N}\hat{u}_{ijt}\right)^2 - \frac{1}{N(N-1)}\sum_{i=1}^{N}\hat{\sigma}_{\varepsilon i}^2\end{aligned}$$

For the other models the above exercise is slightly more complicated. Let us continue with model (1). For this model there were three transformations put forward in this paper, here we are using two of them (3) and (11):

$$\begin{aligned}E\left[u_{ijt}^2\right] &= E\left[\left(\varepsilon_{ijt} - \bar{\varepsilon}_{ij.} - \bar{\varepsilon}_{..t} + \bar{\varepsilon}_{...}\right)^2\right] = \\ &= \frac{(N^2-2)(T-1)}{N^2T}\sigma_{\varepsilon i}^2 + \frac{T-1}{N^3T}\sum_{i=1}^{N}\sigma_{\varepsilon i}^2 - \frac{(N-1)(T-1)}{N^2T}\left(\rho_{(1)} + \rho_{(2)}\right) \\ E\left[u_{ijt}^{*2}\right] &= E\left[\left(\varepsilon_{ijt} - \bar{\varepsilon}_{i.t} - \bar{\varepsilon}_{.jt} + \bar{\varepsilon}_{..t}\right)^2\right] = \\ &= \frac{(N-2)(N-1)}{N^2}\sigma_{\varepsilon i}^2 + \frac{N-1}{N^3}\sum_{i=1}^{N}\sigma_{\varepsilon i}^2 - \frac{(N-1)^2}{N^2}\left(\rho_{(1)} + \rho_{(2)}\right)\end{aligned}$$

Let us notice notice that if we subtract $\frac{T-1}{T(N-1)}$ times the second equation from the first, we get

$$E\left[u_{ijt}^2\right] - \frac{T-1}{T(N-1)}E\left[u_{ijt}^{*2}\right] = -\frac{(N-1)^2}{N}\sigma_{\varepsilon i}^2$$

As a result

$$\hat{\sigma}_{\varepsilon i}^2 = -\frac{1}{(N-1)^2 T}\sum_{j=1}^{N}\sum_{t=1}^{T}\hat{u}_{ijt}^2 + \frac{N(T-1)}{(N-1)^3 T}\sum_{j=1}^{N}\sum_{t=1}^{T}\hat{u}_{ijt}^{*2}$$

Just as with the previous model, we can estimate $\rho_{(1)}$ and $\rho_{(2)}$ by taking the averages of the residuals. For $\rho_{(2)}$

$$\begin{aligned} E\left[\bar{u}_{.jt}^2\right] &= E\left[\left(\bar{\varepsilon}_{.jt} - \bar{\varepsilon}_{.j.} - \bar{\varepsilon}_{..t} + \bar{\varepsilon}_{...}\right)^2\right] = \\ &= \frac{(N-1)(T-1)}{N^3 T}\sum_{i=1}^{N}\sigma_{\varepsilon i}^2 - \frac{(N-1)(T-1)}{N^2 T}\rho_{(1)} + \frac{(N-1)^2(T-1)}{N^2 T}\rho_{(2)} \end{aligned}$$

Now we are ready to express $\hat{\rho}_{(2)}$

$$\frac{(N-1)(T-1)}{NT}\rho_{(2)} = \left[E\left[\bar{u}_{.jt}^2\right] - \frac{T-1}{(N-1)T}\frac{1}{N}\sum_{i=1}^{N}E\left[u_{ijt}^{*2}\right]\right]$$

This leads to

$$\begin{aligned} \hat{\rho}_{(2)} &= \frac{NT}{(N-1)(T-1)}\left[\frac{1}{NT}\sum_{j=1}^{N}\sum_{t=1}^{T}\left(\frac{1}{N}\sum_{i=1}^{N}\hat{u}_{ijt}\right)^2 - \frac{T-1}{N^2(N-1)T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\hat{u}_{ijt}^{*2}\right] = \\ &= \frac{1}{N^2(N-1)(T-1)}\sum_{j=1}^{N}\sum_{t=1}^{T}\left(\sum_{i=1}^{N}\hat{u}_{ijt}\right)^2 - \frac{1}{N(N-1)^2 T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\hat{u}_{ijt}^{*2} \end{aligned}$$

Doing the same for $\rho_{(1)}$ gives

$$\begin{aligned} E\left[\bar{u}_{i.t}^2\right] = E\left[\left(\bar{\varepsilon}_{i.t} - \bar{\varepsilon}_{i..} - \bar{\varepsilon}_{..t} + \bar{\varepsilon}_{...}\right)^2\right] = \frac{(N-2)(T-1)}{N^2 T}\sigma_{\varepsilon i}^2 + \frac{T-1}{N^3 T}\sum_{i=1}^{N}\sigma_{\varepsilon i}^2 - \\ - \frac{(N-1)(T-1)}{N^2 T}\rho_{(2)} + \frac{(N-1)^2(T-1)}{N^2 T}\rho_{(1)} \end{aligned}$$

And so

$$\frac{(N-1)(T-1)}{NT}\rho_{(1)} = \left[E\left[\frac{1}{N}\sum_{i=1}^{N}\bar{u}_{i.t}^2\right] - \frac{T-1}{(N-1)T}\frac{1}{N}\sum_{i=1}^{N}E\left[u_{ijt}^{*2}\right]\right]$$

which leads to

$$\hat{\rho}_{(1)} = \frac{NT}{(N-1)(T-1)}\left[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\frac{1}{N}\sum_{j=1}^{N}\hat{u}_{ijt}\right)^2 - \frac{T-1}{N^2(N-1)T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\hat{u}_{ijt}^{*2}\right] =$$

$$= \frac{1}{N^2(N-1)(T-1)}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\sum_{j=1}^{N}\hat{u}_{ijt}\right)^2 - \frac{1}{N(N-1)^2T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\hat{u}_{ijt}^{*2}$$

Let us continue with model (7). In this case we need to use two new Within transformations, and calculate the variances of the resulting transformed disturbance terms (denoted by u^a and u^b)

$$E\left[(u^a_{ijt})^2\right] = E\left[(\varepsilon_{ijt} - \bar{\varepsilon}_{ij.} - \bar{\varepsilon}_{i.t} + \bar{\varepsilon}_{i..})^2\right] = \frac{(N-1)(T-1)}{NT}\sigma^2_{\varepsilon i} - \frac{(N-1)(T-1)}{NT}\rho_{(1)}$$

$$E\left[(u^b_{ijt})^2\right] = E\left[(\varepsilon_{ijt} - \bar{\varepsilon}_{ij.} - \bar{\varepsilon}_{.jt} + \bar{\varepsilon}_{.j.})^2\right] =$$

$$= \frac{(N-2)(T-1)}{NT}\sigma^2_{\varepsilon i} + \frac{T-1}{N^2T}\sum_{i=1}^{N}\sigma^2_{\varepsilon i} - \frac{(N-1)(T-1)}{NT}\rho_{(2)}$$

Now, in order to express $\rho_{(1)}$ from the equations one has to transform further u^b_{ijt} by taking the averages with respect to j, and then take the average of the obtained variances with respect to i

$$\frac{1}{N}\sum_{i=1}^{N}E\left[(\bar{u}^b_{i.t})^2\right] = \frac{1}{N}\sum_{i=1}^{N}E\left[(\bar{\varepsilon}_{i.t} - \bar{\varepsilon}_{i..} - \bar{\varepsilon}_{..t} + \bar{\varepsilon}_{...})^2\right] =$$

$$= \frac{(N-1)(T-2)}{N^3T}\sum_{i=1}^{N}\sigma^2_{\varepsilon i} + \frac{(N-1)^2(T-2)}{N^2T}\rho_{(1)} - \frac{(N-1)(T-2)}{N^2T}\rho_{(2)}$$

It can be noticed that

$$\hat{\rho}_{(1)} = \frac{N^2T}{(N-1)^2(T-2)}\left\{\frac{1}{N}\sum_{i=1}^{N}E\left[(\bar{\hat{u}}^b_{i.t})^2\right] - \frac{(T-2)}{N^2(T-1)}\sum_{i=1}^{N}E\left[(\hat{u}^b_{ijt})^2\right]\right\} =$$

$$= \frac{1}{N(N-1)^2(T-2)}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\sum_{j=1}^{N}\hat{u}^b_{ijt}\right)^2 - \frac{1}{N(N-1)^2(T-1)}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}(\hat{u}^b_{ijt})^2$$

The other components can easily be derived

$$\hat{\sigma}^2_{\varepsilon i} = \hat{\rho}_{(1)} + \frac{NT}{(N-1)(T-1)}E\left[(\hat{u}^a_{ijt})^2\right] = \hat{\rho}_{(1)} + \frac{1}{(N-1)(T-1)}\sum_{j=1}^{N}\sum_{t=1}^{T}(\hat{u}^a_{ijt})^2$$

$$\hat{\rho}_{(2)} = \frac{1}{N}\sum_{i=1}^{N}\hat{\sigma}_{\varepsilon i}^2 - \frac{NT}{(N-1)(T-1)}E\left[(u_{ijt}^b)^2\right] =$$

$$= \frac{1}{N}\sum_{i=1}^{N}\hat{\sigma}_{\varepsilon i}^2 - \frac{1}{N(N-1)(T-1)}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}(\hat{u}_{ijt}^b)^2$$

Unfortunately, we are not that lucky with the other models. The difficulty is that we are not able to transform the residuals or come forward with other transformations, which produce new, linearly independent equations to estimate the variance components. Instead we need to impose further restrictions on the models. Let us assume from now on that $\rho_{(1)} = \rho_{(2)} = \rho$.

For model (8) we have

$$E\left[u_{ijt}^2\right] = E\left[(\varepsilon_{ijt} - \bar{\varepsilon}_{.jt})^2\right] = \frac{N-2}{N}\sigma_{\varepsilon i}^2 + \frac{1}{N^2}\sum_{i=1}^{N}\sigma_{\varepsilon i}^2 - \frac{N-1}{N}\rho$$

Just like before, taking averages of u_{ijt} with respect to j leads to

$$E\left[\bar{u}_{i.t}^2\right] = E\left[(\bar{\varepsilon}_{i.t} - \bar{\varepsilon}_{..t})^2\right] = \frac{N-2}{N^2}\sigma_{\varepsilon i}^2 + \frac{1}{N^3}\sum_{i=1}^{N}\sigma_{\varepsilon i}^2 + \frac{(N-1)(N-2)}{N^2}\rho$$

In this way we can estimate ρ

$$\hat{\rho} = \frac{N^2}{(N-1)^2}\left[E\left[\bar{\hat{u}}_{i.t}^2\right] - \frac{1}{N}E\left[\hat{u}_{ijt}^2\right]\right] =$$

$$= \frac{N}{(N-1)^2 T}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\frac{1}{N}\sum_{j=1}^{N}\hat{u}_{ijt}\right)^2 - \frac{1}{N(N-1)^2 T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\hat{u}_{ijt}^2 =$$

$$= \frac{1}{N(N-1)^2 T}\left\{\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\sum_{j=1}^{N}\hat{u}_{ijt}\right)^2 - \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\hat{u}_{ijt}^2\right\}$$

and now we can move to estimate $\sigma^2_{\varepsilon i}$

$$\hat{\sigma}^2_{\varepsilon i} = \frac{N^2}{N-2}\left\{E\left[\bar{\hat{u}}^2_{i.t}\right] - \frac{1}{N^2(N-1)}\sum_{i=1}^{N} E\left[\hat{u}^2_{ijt}\right] - \frac{(N-1)(N-2)+1}{N^2}\hat{\rho}\right\} =$$

$$= \frac{N^2}{(N-2)T}\sum_{t=1}^{T}\left(\frac{1}{N}\sum_{j=1}^{N}\hat{u}_{ijt}\right)^2 - \frac{1}{N(N-1)(N-2)T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\hat{u}^2_{ijt} -$$
$$- \frac{(N-1)(N-2)+1}{N-2}\hat{\rho} =$$
$$= \frac{1}{(N-2)T}\sum_{t=1}^{T}\left(\sum_{j=1}^{N}\hat{u}_{ijt}\right)^2 - \frac{1}{N(N-1)(N-2)T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\hat{u}^2_{ijt} -$$
$$- \frac{(N-1)(N-2)+1}{N-2}\hat{\rho}$$

We still have two models, namely (10) and (12), to deal with. For these, however, unfortunately it is not possible to estimate any cross correlation at all. So we have to assume zero cross correlation and focus only on the heteroscedasticity and the estimation of the $\sigma^2_{\varepsilon i}$ variances.

For model (10) we have

$$E\left[u^2_{ijt}\right] = E\left[\left(\varepsilon_{ijt} - \bar{\varepsilon}_{i.t} - \bar{\varepsilon}_{.jt} + \bar{\varepsilon}_{..t}\right)^2\right] = \frac{(N-1)(N-2)}{N^2}\sigma^2_{\varepsilon i} + \frac{N-1}{N^3}\sum_{i=1}^{N}\sigma^2_{\varepsilon i}$$

Taking the averages with respect to i

$$\frac{1}{N}\sum_{i=1}^{N} E\left[u^2_{ijt}\right] = \frac{(N-1)^2}{N^3}\sum_{i=1}^{N}\sigma^2_{\varepsilon i}$$

As a result,

$$\hat{\sigma}^2_{\varepsilon i} = \frac{N^2}{(N-1)^2(N-2)}\left\{(N-1)E\left[\hat{u}^2_{ijt}\right] - \frac{1}{N}\sum_{i=1}^{N}E\left[\hat{u}^2_{ijt}\right]\right\} =$$
$$= \frac{N}{(N-1)(N-2)T}\sum_{j=1}^{N}\sum_{t=1}^{T}\hat{u}^2_{ijt} - \frac{1}{(N-1)^2(N-2)T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\hat{u}^2_{ijt}$$

We can proceed likewise for model (12)

$$E\left[u_{ijt}^2\right] = E\left[\left(\varepsilon_{ijt} - \bar{\varepsilon}_{i.t} - \bar{\varepsilon}_{.jt} - \bar{\varepsilon}_{ij.} + \bar{\varepsilon}_{i..} + \bar{\varepsilon}_{.j.} + \bar{\varepsilon}_{..t} - \bar{\varepsilon}_{...}\right)^2\right] =$$

$$= \frac{(N-1)(N-2)(T-1)}{N^2T}\sigma_{\varepsilon i}^2 + \frac{(N-1)(T-1)}{N^3T}\sum_{i=1}^{N}\sigma_{\varepsilon i}^2$$

Again,

$$\frac{1}{N}\sum_{i=1}^{N}E\left[u_{ijt}^2\right] = \frac{(N-1)^2(T-1)}{N^3T}\sum_{i=1}^{N}\sigma_{\varepsilon i}^2$$

and as a result,

$$\hat{\sigma}_{\varepsilon i}^2 = \frac{N^2T}{(N-1)^2(N-2)(T-1)}\left\{(N-1)E\left[\hat{u}_{ijt}^2\right] - \frac{1}{N}\sum_{i=1}^{N}E\left[\hat{u}_{ijt}^2\right]\right\} =$$

$$= \frac{N}{(N-1)(N-2)(T-1)}\sum_{j=1}^{N}\sum_{t=1}^{T}\hat{u}_{ijt}^2 - \frac{1}{(N-1)^2(N-2)(T-1)}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\hat{u}_{ijt}^2$$

7. Conclusion

In the case of three and higher dimensional fixed effects panel data models, due to the many interaction effects, the number of dummy variables in the model increases dramatically. As a consequence, even when the number of individuals is not too large, the LSDV estimator becomes, unfortunately, practically unfeasible. The obvious answer to this challenge is to use appropriate Within estimators, which do not require the explicit incorporation of the fixed effects into the model. Although these Within estimators are more complex than for the usual two dimensional panel data models, they are quite useful in these higher dimensional setups. However, unlike in the two dimensional case, they are biased and inconsistent in the case of some very relevant data problems like the lack of self-trade, or unbalanced observations. These properties must be taken into account by all researchers relying on these methods. The summary of the most important findings of the paper about the behaviour of the many transformations available in higher dimensional panel data sets can be found in Table 1.

Appendix 1

Finite sample bias derivations for no self flow data

For model (1) with transformation (4)

$$\tilde{\alpha}_i = \alpha_i - \frac{1}{(N-1)T} \cdot (N-1)T \cdot \alpha_i - \frac{1}{(N-1)T} \sum_{i=1;\ i\neq j}^{N} T \cdot \alpha_i$$

$$- \frac{1}{N(N-1)} \sum_{i=1}^{N} (N-1) \cdot \alpha_i + \frac{2}{N(N-1)T} \sum_{i=1}^{N} (N-1)T \cdot \alpha_i$$

$$= \alpha_i - \alpha_i - \frac{1}{N-1} \sum_{i=1;\ i\neq j}^{N} \alpha_i + \frac{1}{N} \sum_{i=1}^{N} \alpha_i = \frac{1}{N}\alpha_j - \frac{1}{N(N-1)} \sum_{i=1;\ i\neq j}^{N} \alpha_i$$

$$\tilde{\gamma}_j = \gamma_j - \frac{1}{(N-1)T} \sum_{j=1;\ j\neq i}^{N} T \cdot \gamma_j - \frac{1}{(N-1)T} \cdot (N-1)T \cdot \gamma_j$$

$$- \frac{1}{N(N-1)} \sum_{j=1}^{N} (N-1) \cdot \gamma_j + \frac{2}{N(N-1)T} \sum_{j=1}^{N} (N-1)T \cdot \gamma_j$$

$$= \gamma_j - \frac{1}{N-1} \sum_{j=1;\ j\neq i}^{N} \gamma_j - \gamma_j + \frac{1}{N} \sum_{j=1}^{N} \gamma_j = \frac{1}{N}\gamma_i - \frac{1}{N(N-1)} \sum_{j=1;\ j\neq i}^{N} \gamma_j$$

and for the time effects

$$\tilde{\lambda}_t = \lambda_t - \frac{1}{(N-1)T} \sum_{t=1}^{T} (N-1) \cdot \lambda_t - \frac{1}{(N-1)T} \sum_{t=1}^{T} (N-1) \cdot \lambda_t$$

$$- \frac{1}{N(N-1)} \cdot N(N-1)\lambda_t + \frac{2}{N(N-1)T} \sum_{t=1}^{T} N(N-1) \cdot \lambda_t =$$

$$= \lambda_t - \frac{1}{T} \sum_{t=1}^{T} \lambda_t - \frac{1}{T} \sum_{t=1}^{T} \lambda_t - \lambda_t + \frac{2}{T} \sum_{t=1}^{T} \lambda_t = 0$$

For model (10) with Within transformation (11)

$$\tilde{\alpha}_{it} = \alpha_{it} - \frac{1}{N-1} \sum_{i=1;i\neq j}^{N} \alpha_{it} - \frac{1}{N-1}(N-1)\alpha_{it} + \frac{1}{N(N-1)} \sum_{i=1}^{N} (N-1)\alpha_{it}$$

$$= -\frac{1}{N(N-1)} \sum_{k=1;k\neq j}^{N} \alpha_{kt} + \frac{1}{N}\alpha_{jt}$$

and

$$\tilde{\alpha}_{jt}^{*} = \alpha_{jt}^{*} - \frac{1}{N-1}(N-1)\alpha_{jt}^{*} - \frac{1}{N-1}\sum_{j=1;j\neq i}^{N}\alpha_{jt}^{*} + \frac{1}{N(N-1)}\sum_{j=1}^{N}(N-1)\alpha_{jt}^{*}$$

$$= -\frac{1}{N(N-1)}\sum_{l=1;l\neq i}^{N}\alpha_{lt}^{*} + \frac{1}{N}\alpha_{it}^{*}$$

For model (12) and transformation (13)

$$\tilde{\gamma}_{ij} = \gamma_{ij} - \frac{1}{T}T\cdot\gamma_{ij} - \frac{1}{N-1}\sum_{i=1;i\neq j}^{N}\gamma_{ij} - \frac{1}{N-1}\sum_{j=1;j\neq i}^{N}\gamma_{ij}$$

$$+ \frac{1}{N(N-1)}\sum_{i=1}^{N}\sum_{j=1;j\neq i}^{N}\gamma_{ij} + \frac{1}{(N-1)T}\sum_{i=1;i\neq j}^{N}T\gamma_{ij}$$

$$+ \frac{1}{(N-1)T}\sum_{j=1;j\neq i}^{N}T\gamma_{ij} - \frac{1}{N(N-1)T}\sum_{i=1}^{N}\sum_{j=1;j\neq i}^{N}T\gamma_{ij} = 0$$

but

$$\tilde{\alpha}_{it} = \alpha_{it} - \frac{1}{T}\sum_{t=1}^{T}\alpha_{it} - \frac{1}{N-1}\sum_{i=1;i\neq j}^{N}\alpha_{it} - \frac{1}{N-1}(N-1)\alpha_{it}$$

$$+ \frac{1}{N(N-1)}\sum_{i=1}^{N}(N-1)\alpha_{it} + \frac{1}{(N-1)T}\sum_{i=1;i\neq j}^{N}\sum_{t=1}^{T}\alpha_{it}$$

$$+ \frac{1}{(N-1)T}\sum_{t=1}^{T}(N-1)\alpha_{it} - \frac{1}{N(N-1)T}\sum_{i=1}^{N}\sum_{t=1}^{T}(N-1)\alpha_{it}$$

$$= \frac{1}{N(N-1)T}\sum_{i=1;i\neq j}^{N}\sum_{t=1}^{T}\alpha_{it} + \frac{1}{NT}\sum_{t=1}^{T}\alpha_{jt} - \frac{1}{N(N-1)}\sum_{i=1;i\neq j}^{N}\alpha_{it} + \frac{1}{N}\alpha_{jt}$$

and, finally

$$\tilde{\alpha}_{jt}^{*} = \alpha_{jt}^{*} - \frac{1}{T}\sum_{t=1}^{T}\alpha_{jt}^{*} - \frac{1}{N-1}(N-1)\alpha_{jt}^{*} - \frac{1}{N-1}\sum_{j=1;j\neq i}^{N}\alpha_{jt}^{*}$$

$$+ \frac{1}{N(N-1)}\sum_{j=1}^{N}(N-1)\alpha_{jt}^{*} + \frac{1}{(N-1)T}\sum_{t=1}^{T}(N-1)\alpha_{jt}^{*}$$

$$+ \frac{1}{(N-1)T}\sum_{j=1;j\neq i}^{N}\sum_{t=1}^{T}\alpha_{jt}^{*} - \frac{1}{N(N-1)T}\sum_{j=1}^{N}\sum_{t=1}^{T}(N-1)\alpha_{jt}^{*}$$

$$= \frac{1}{N(N-1)T}\sum_{j=1;j\neq i}^{N}\sum_{t=1}^{T}\alpha_{jt}^{*} + \frac{1}{NT}\sum_{t=1}^{T}\alpha_{it}^{*} - \frac{1}{N(N-1)}\sum_{j=1;j\neq i}^{N}\alpha_{jt}^{*} + \frac{1}{N}\alpha_{it}^{*}$$

Finite sample bias derivations for unbalanced data

For model (1) and transformation (3)

$$\begin{aligned}
\tilde{\alpha}_i &= \alpha_i - \frac{1}{T_{ij}}\sum_{t=1}^{T_{ij}}\alpha_i - \frac{1}{N^2}\sum_{i=1}^{N} N\alpha_i + \frac{1}{\sum_{i=1}^{N}\sum_{j=1}^{N} T_{ij}}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T_{ij}}\alpha_i \\
&= -\frac{1}{N}\sum_{i=1}^{N}\alpha_i + \frac{1}{T}\sum_{i=1}^{N}\left(\alpha_i \cdot \sum_{j=1}^{N} T_{ij}\right) \\
&= \frac{1}{NT}\sum_{i=1}^{N}\alpha_i \cdot (N\sum_{j=1}^{N} T_{ij} - T)
\end{aligned}$$

$$\begin{aligned}
\tilde{\gamma}_j &= \gamma_j - \frac{1}{T_{ij}}\sum_{t=1}^{T_{ij}}\gamma_j - \frac{1}{N^2}\sum_{j=1}^{N} N\gamma_j + \frac{1}{\sum_{i=1}^{N}\sum_{j=1}^{N} T_{ij}}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T_{ij}}\gamma_j \\
&= -\frac{1}{N}\sum_{j=1}^{N}\gamma_j + \frac{1}{T}\sum_{j=1}^{N}\left(\gamma_j \cdot \sum_{i=1}^{N} T_{ij}\right) \\
&= \frac{1}{NT}\sum_{j=1}^{N}\gamma_j \cdot (N\sum_{i=1}^{N} T_{ij} - T)
\end{aligned}$$

and

$$\begin{aligned}
\tilde{\lambda}_t &= \lambda_t - \frac{1}{T_{ij}}\sum_{t=1}^{T_{ij}}\lambda_t - \frac{1}{N^2}N^2\lambda_t + \frac{1}{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T_{ij}}\lambda_t \\
&= \lambda_t - \frac{1}{T_{ij}}\sum_{t=1}^{T_{ij}}\lambda_t - \lambda_t + \frac{1}{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T_{ij}}\lambda_t \\
&= -\frac{1}{T_{ij}}\sum_{t=1}^{T_{ij}}\lambda_t + \frac{1}{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T_{ij}}\lambda_t
\end{aligned}$$

For model (12) and transformation (13)

$$
\begin{aligned}
\tilde{\gamma}_{ij} &= \gamma_{ij} - \frac{1}{T_{ij}}\sum_{t=1}^{T_{ij}}\gamma_{ij} - \frac{1}{N}\sum_{i=1}^{N}\gamma_{ij} - \frac{1}{N}\sum_{j=1}^{N}\gamma_{ij} + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\gamma_{ij}+\\
&\quad + \frac{1}{\sum_{i=1}^{N}T_{ij}}\sum_{i=1}^{N}\sum_{t=1}^{T_{ij}}\gamma_{ij} + \frac{1}{\sum_{j=1}^{N}T_{ij}}\sum_{j=1}^{N}\sum_{t=1}^{T_{ij}}\gamma_{ij} - \frac{1}{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T_{ij}}\gamma_{ij}\\
&= \gamma_{ij} - \gamma_{ij} - \frac{1}{N}\sum_{i=1}^{N}\gamma_{ij} - \frac{1}{N}\sum_{j=1}^{N}\gamma_{ij} + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\gamma_{ij} + \frac{1}{\sum_{i=1}^{N}T_{ij}}\sum_{i=1}^{N}\gamma_{ij}T_{ij}+\\
&\quad + \frac{1}{\sum_{j=1}^{N}T_{ij}}\sum_{j=1}^{N}\gamma_{ij}T_{ij} - \frac{1}{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\gamma_{ij}T_{ij}\\
&= -\frac{1}{N}\sum_{i=1}^{N}\gamma_{ij} - \frac{1}{N}\sum_{j=1}^{N}\gamma_{ij} + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\gamma_{ij} + \frac{1}{\sum_{i=1}^{N}T_{ij}}\sum_{i=1}^{N}\gamma_{ij}T_{ij}+\\
&\quad + \frac{1}{\sum_{j=1}^{N}T_{ij}}\sum_{j=1}^{N}\gamma_{ij}T_{ij} - \frac{1}{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\gamma_{ij}T_{ij}
\end{aligned}
$$

$$
\begin{aligned}
\tilde{\alpha}_{it} &= \alpha_{it} - \frac{1}{T_{ij}}\sum_{t=1}^{T_{ij}}\alpha_{it} - \frac{1}{N}\sum_{i=1}^{N}\alpha_{it} - \frac{1}{N}\sum_{j=1}^{N}\alpha_{it} + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_{it}+\\
&\quad + \frac{1}{\sum_{i=1}^{N}T_{ij}}\sum_{i=1}^{N}\sum_{t=1}^{T_{ij}}\alpha_{it} + \frac{1}{\sum_{j=1}^{N}T_{ij}}\sum_{j=1}^{N}\sum_{t=1}^{T_{ij}}\alpha_{it} - \frac{1}{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T_{ij}}\alpha_{it}\\
&= \alpha_{it} - \frac{1}{T_{ij}}\sum_{t=1}^{T_{ij}}\alpha_{it} - \frac{1}{N}\sum_{i=1}^{N}\alpha_{it} - \alpha_{it} + \frac{1}{N}\sum_{i=1}^{N}\alpha_{it}+\\
&\quad + \frac{1}{\sum_{i=1}^{N}T_{ij}}\sum_{i=1}^{N}\sum_{t=1}^{T_{ij}}\alpha_{it} + \frac{1}{\sum_{j=1}^{N}T_{ij}}\sum_{j=1}^{N}\sum_{t=1}^{T_{ij}}\alpha_{it} - \frac{1}{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T_{ij}}\alpha_{it}\\
&= -\frac{1}{T_{ij}}\sum_{t=1}^{T_{ij}}\alpha_{it} + \frac{1}{\sum_{i=1}^{N}T_{ij}}\sum_{i=1}^{N}\sum_{t=1}^{T_{ij}}\alpha_{it} + \frac{1}{\sum_{j=1}^{N}T_{ij}}\sum_{j=1}^{N}\sum_{t=1}^{T_{ij}}\alpha_{it} - \frac{1}{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T_{ij}}\alpha_{it}
\end{aligned}
$$

and

$$\begin{aligned}
\tilde{\alpha}^*_{jt} &= \alpha^*_{jt} - \frac{1}{T_{ij}}\sum_{t=1}^{T_{ij}}\alpha^*_{jt} - \frac{1}{N}\sum_{i=1}^{N}\alpha^*_{jt} - \frac{1}{N}\sum_{j=1}^{N}\alpha^*_{jt} + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha^*_{jt} + \\
&\quad + \frac{1}{\sum_{i=1}^{N} T_{ij}}\sum_{i=1}^{N}\sum_{t=1}^{T_{ij}}\alpha^*_{jt} + \frac{1}{\sum_{j=1}^{N} T_{ij}}\sum_{j=1}^{N}\sum_{t=1}^{T_{ij}}\alpha^*_{jt} - \frac{1}{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T_{ij}}\alpha^*_{jt} \\
&= \alpha^*_{jt} - \frac{1}{T_{ij}}\sum_{t=1}^{T_{ij}}\alpha^*_{jt} - \alpha^*_{jt} - \frac{1}{N}\sum_{i=1}^{N}\alpha^*_{jt} + \frac{1}{N}\sum_{i=1}^{N}\alpha^*_{jt} + \frac{1}{\sum_{i=1}^{N} T_{ij}}\sum_{i=1}^{N}\sum_{t=1}^{T_{ij}}\alpha^*_{jt} + \\
&\quad + \frac{1}{\sum_{j=1}^{N} T_{ij}}\sum_{j=1}^{N}\sum_{t=1}^{T_{ij}}\alpha^*_{jt} - \frac{1}{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T_{ij}}\alpha^*_{jt} \\
&= -\frac{1}{T_{ij}}\sum_{t=1}^{T_{ij}}\alpha^*_{jt} + \frac{1}{\sum_{i=1}^{N} T_{ij}}\sum_{i=1}^{N}\sum_{t=1}^{T_{ij}}\alpha^*_{jt} + \frac{1}{\sum_{j=1}^{N} T_{ij}}\sum_{j=1}^{N}\sum_{t=1}^{T_{ij}}\alpha^*_{jt} - \frac{1}{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T_{ij}}\alpha^*_{jt}
\end{aligned}$$

Appendix 2

Elaboration on $W_{(1)}$

Since $LL' = (I_N \otimes I_N - \sum_i e_i e_i' \otimes e_i e_i')$, with J_N an $N \times N$ matrix of ones, then

$$\begin{aligned}
D_\alpha' LL' D_\alpha &= (l_N \otimes I_N)' \left\{ I_N \otimes I_N - \sum_i e_i e_i' \otimes e_i e_i' \right\} (l_N \otimes I_N) \\
&= (l_N' l_N) I_N - (l_N' e_i)^2 \sum_i e_i e_i' \\
&= (N-1) I_N \\
D_\alpha' LL' D_\gamma &= (l_N \otimes I_N)' \left\{ I_N \otimes I_N - \sum_i e_i e_i' \otimes e_i e_i' \right\} (I_N \otimes l_N) \\
&= J_N - \sum_i (l_N' e_i) e_i' \otimes e_i (e_i' l_N) \\
&= J_N - I_N \quad \text{and} \\
D_\gamma' LL' D_\gamma &= (N-1) I_N.
\end{aligned}$$

So, with $\bar{J}_N \equiv J_N / N$ and $A_N \equiv I_N - \bar{J}_N$,

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

Since A_N and $\bar{J}_N$ are idempotent and mutually orthogonal,

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

with

$$\begin{aligned}
p &\equiv \frac{1}{4(N-1)} - \frac{N-1}{N(N-2)} \\
q &\equiv \frac{1}{4(N-1)} - \frac{1}{N(N-2)} \\
\frac{p+q}{N} &= \frac{1}{N}\left\{\frac{1}{2(N-1)} - \frac{1}{N-2}\right\} \\
&= -\frac{1}{2(N-1)(N-2)}.
\end{aligned}$$

Elaboration on $W_{(2)}$

Notice again, that $\tilde{L}\tilde{L}' = (I_N \otimes I_N \otimes I_T - \sum_i e_i e_i' \otimes e_i e_i' \otimes I_T)$, and

$$\begin{aligned}
D'_{\alpha_*}\tilde{L}\tilde{L}'D_{\alpha_*} &= (I_N \otimes l_N \otimes l_T)'\left\{I_N \otimes I_N \otimes I_T - \sum_i e_i e_i' \otimes e_i e_i' \otimes I_T\right\}(I_N \otimes l_N \otimes l_T) \\
&= I_N \otimes l_N' l_N \otimes l_T' l_T - (l_N' e_i)^2 l_T' l_T \sum_i e_i e_i' \\
&= (N-1)T I_N \\
D'_{\gamma_*}\tilde{L}\tilde{L}'D_{\gamma_*} &= (N-1)T I_N \\
D'_{\lambda}\tilde{L}\tilde{L}'D_{\lambda} &= (l_N \otimes l_N \otimes I_T)'\left\{I_N \otimes I_N \otimes I_T - \sum_i e_i e_i' \otimes e_i e_i' \otimes I_T\right\}(l_N \otimes l_N \otimes I_T) \\
&= l_N' l_N \otimes l_N' l_N \otimes I_T - \sum_i (l_N' e_i e_i' l_N)^2 \otimes I_T \\
&= N(N-1) I_T \\
D'_{\alpha_*}\tilde{L}\tilde{L}'D_{\gamma_*} &= (I_N \otimes l_N \otimes l_T)'\left\{I_N \otimes I_N \otimes I_T - \sum_i e_i e_i' \otimes e_i e_i' \otimes I_T\right\}(l_N \otimes I_N \otimes l_T) \\
&= J_N \otimes l_T' l_T - (l_N' e_i)^2 l_T' l_T \sum_i e_i e_i' \\
&= T(J_N - I_N)
\end{aligned}$$

$$
\begin{aligned}
D'_{\gamma_*}\tilde{L}\tilde{L}'D_{\alpha_*} &= T(J_N - I_N) \\
D'_{\alpha_*}\tilde{L}\tilde{L}'D_{\lambda} &= (I_N \otimes l_N \otimes l_T)' \left\{ I_N \otimes I_N \otimes I_T - \sum_i e_i e_i' \otimes e_i e_i' \otimes I_T \right\} (l_N \otimes l_N \otimes I_T) \\
&= l_N \otimes l_N' l_N \otimes l_T' - (l_N' e_i)^2 \sum_i e_i \otimes l_T' \\
&= (N-1) l_N l_T' \\
D'_{\gamma_*}\tilde{L}\tilde{L}'D_{\lambda} &= (N-1) l_N l_T' \\
D'_{\lambda}\tilde{L}\tilde{L}'D_{\alpha_*} &= (N-1) l_T l_N' \text{ and} \\
D'_{\lambda}\tilde{L}\tilde{L}'D_{\gamma_*} &= (N-1) l_T l_N',
\end{aligned}
$$

so

$$
W_{(2)} = \left(\begin{array}{cc|c} (N-1)TI_N & T(J_N - I_N) & (N-1) l_N l_T' \\ T(J_N - I_N) & (N-1)TI_N & (N-1) l_N l_T' \\ \hline (N-1) l_T l_N' & (N-1) l_T l_N' & N(N-1) I_T \end{array} \right) = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}.
$$

It is slightly more challenging to derive $W_{(2)}^-$. One has to use partial inverse calculations with the four boxes defined by the dashed lines. It turns out however, that the only "direct inverse" we have to calculate, A_{11}^{-1}, is actually $1/T$ times the $W_{(1)}^-$ from the two dimensional model (15). After lengthy derivations,

$$
W_{(2)}^- = \begin{pmatrix} \frac{1}{TN(N-2)} \begin{pmatrix} N-1 & 1 \\ 1 & N-1 \end{pmatrix} \otimes I_N + \begin{pmatrix} r & s \\ s & r \end{pmatrix} \otimes J_N & -\frac{1}{TN(N-1)} \iota_{2N} l_T' \\ -\frac{1}{TN(N-1)} l_T l_{2N}' & \frac{1}{N(N-1)} (I_T + 1/T J_T) \end{pmatrix}
$$

with

$$
\begin{aligned}
r &\equiv -\frac{-N^2 + 2N - 4}{4N^2(N-1)(N-2)T} \\
s &\equiv \frac{3N^2 - 10N + 4}{4N^2(N-1)(N-2)T}.
\end{aligned}
$$

Appendix 3

Matrix form of transformation (21)

$$
\begin{aligned}
&\big(I_{N(N-1)T} - \bar{J}_{N-1} \otimes I_{NT} - I_N \otimes \bar{J}_{N-1} \otimes I_T - I_{N(N-1)} \otimes \bar{J}_T \\
&\quad + \bar{J}_{N(N-1)} \otimes I_T + \bar{J}_{N-1} \otimes I_N \otimes \bar{J}_T + I_N \otimes \bar{J}_{(N-1)T} - \bar{J}_{N(N-1)T} \\
&\quad - \frac{1}{N-1}\bar{J}_{N(N-1)T} + \frac{1}{N-1}\bar{J}_{N(N-1)} \otimes I_T + \frac{1}{N-1}K_{N(N-1)} \otimes \bar{J}_T \\
&\quad - \frac{1}{N-1}K_{N(N-1)} \otimes I_T\big) = \\
&\big(I_{N(N-1)T} - \bar{J}_{N-1} \otimes I_{NT} - I_N \otimes \bar{J}_{N-1} \otimes I_T - I_{N(N-1)} \otimes \bar{J}_T \\
&\quad + \frac{N}{N-1}\bar{J}_{N(N-1)} \otimes I_T + \bar{J}_{N-1} \otimes I_N \otimes \bar{J}_T + I_N \otimes \bar{J}_{(N-1)T} - \frac{N}{N-1}\bar{J}_{N(N-1)T} \\
&\quad + \frac{1}{N-1}K_{N(N-1)} \otimes \bar{J}_T - \frac{1}{N-1}K_{N(N-1)} \otimes I_T\big)
\end{aligned}
$$

where $K_{N(N-1)}$ is the matrix with the following rows: the row corresponding to observation (ij) is a row of 0-s with 1 in the (ji)th place, that is the (ij)th row is in fact

$$\left[0, 0, \ldots 0, \underbrace{1}_{(ji)-\text{th element}}, 0, \ldots, 0\right]$$

Appendix 4

Derivation of the bias formulae in case of incomplete data

The Within estimator is

$$\hat{\beta} = \frac{\sum_{ijt} \tilde{x}_{ijt}\tilde{y}_{ijt}}{\sum_{ijt} \tilde{x}_{ijt}\tilde{x}'_{ijt}},$$

where $\tilde{x}$ and $\tilde{y}$ denotes the transformed dependent and explanatory variables respectively. Now

$$E\left[\hat{\beta} - \beta\right] = \frac{\sum_{ijt} \tilde{x}_{ijt} \cdot FE_{ijt}}{\sum_{ijt} \tilde{x}_{ijt}\tilde{x}'_{ijt}},$$

where FE_{ijt} stands for the non-eliminated fixed effects parameters for observation (ijt) after the appropriate transformation. The formula above serves as some general scheme as the Within transformation and remained fixed effects parameters uniquely determine the exact value of the bias. For example, in the case of model (10), with transformation (11) the no self-flow bias takes the form

$$E\left[\hat{\beta}-\beta\right] = \frac{\sum_{ijt}\left(x_{ijt}-\bar{x}_{.jt}-\bar{x}_{i.t}+\bar{x}_{..t}\right)'\cdot A}{\sum_{ijt}\left(x_{ijt}-\bar{x}_{.jt}-\bar{x}_{i.t}+\bar{x}_{..t}\right)'\cdot\left(x_{ijt}-\bar{x}_{.jt}-\bar{x}_{i.t}+\bar{x}_{..t}\right)}$$

$$\text{with} \quad A = (-\frac{1}{N-1}\sum_{i=1;\, i\neq j}^{N}\alpha_{it} + \frac{1}{N}\sum_{i=1}^{N}\alpha_{it} - \frac{1}{N-1}\sum_{j=1;\, j\neq i}^{N}\alpha^*_{jt} + \frac{1}{N}\sum_{j=1}^{N}\alpha^*_{jt})$$

Appendix 5

Derivation of the semi-asymptotic bias for the dynamic model

All possible plim terms resulting from the bias derivation can be assigned to one of the groups listed below. The $\operatorname*{plim}_{N\to\infty}\frac{1}{N^2}\sum_{ij}$ of all the components in a group is the same.

$$\begin{aligned}
&y_{ijt-1}\bar{\varepsilon}_{i.t}; \quad y_{ijt-1}\bar{\varepsilon}_{.jt}; \quad y_{ijt-1}\bar{\varepsilon}_{..t}; \quad \bar{y}_{i.t-1}\varepsilon_{ijt}; \quad \bar{y}_{.jt-1}\varepsilon_{ijt}; \quad \bar{y}_{i.t-1}\bar{\varepsilon}_{i.t};\\
&\bar{y}_{.jt-1}\bar{\varepsilon}_{.jt}; \quad \bar{y}_{i.t-1}\bar{\varepsilon}_{.jt}; \quad \bar{y}_{.jt-1}\bar{\varepsilon}_{i.t}; \quad \bar{y}_{i.t-1}\bar{\varepsilon}_{..t}; \quad \bar{y}_{.jt-1}\bar{\varepsilon}_{..t};\\
&\bar{y}_{..t-1}\varepsilon_{ijt}; \quad \bar{y}_{..t-1}\bar{\varepsilon}_{.jt}; \quad \bar{y}_{..t-1}\bar{\varepsilon}_{i.t}; \quad \bar{y}_{..t-1}\bar{\varepsilon}_{..t}\\
&\Rightarrow \quad 0
\end{aligned}$$

$$\begin{aligned}
&y_{ijt-1}\bar{\varepsilon}_{i..}; \quad y_{ijt-1}\bar{\varepsilon}_{.j.}; \quad \bar{y}_{i.t-1}\bar{\varepsilon}_{ij.}; \quad \bar{y}_{.jt-1}\bar{\varepsilon}_{ij.}; \quad \bar{y}_{i.t-1}\bar{\varepsilon}_{i..}; \quad \bar{y}_{.jt-1}\bar{\varepsilon}_{.j.}\\
&\Rightarrow \quad \operatorname*{plim}_{N\to\infty}\frac{\sigma^2_\varepsilon}{NT}\cdot\frac{1-\rho^{t-1}}{1-\rho} = 0
\end{aligned}$$

$$\begin{aligned}
&y_{ijt-1}\bar{\varepsilon}_{...}; \quad \bar{y}_{i.t-1}\bar{\varepsilon}_{.j.}; \quad \bar{y}_{.jt-1}\bar{\varepsilon}_{i..}; \quad \bar{y}_{i.t-1}\bar{\varepsilon}_{.j.}; \quad \bar{y}_{.jt-1}\bar{\varepsilon}_{i..}; \quad \bar{y}_{i.t-1}\bar{\varepsilon}_{...};\\
&\bar{y}_{.jt-1}\bar{\varepsilon}_{...}; \quad \bar{y}_{..t-1}\bar{\varepsilon}_{ij.}; \quad \bar{y}_{..t-1}\bar{\varepsilon}_{i..}; \quad \bar{y}_{..t-1}\bar{\varepsilon}_{.j.}; \quad \bar{y}_{..t-1}\bar{\varepsilon}_{...}\\
&\Rightarrow \quad \operatorname*{plim}_{N\to\infty}\frac{\sigma^2_\varepsilon}{N^2T}\cdot\frac{1-\rho^{t-1}}{1-\rho} = 0
\end{aligned}$$

$$\begin{aligned}
&\bar{y}_{ij.-1}\bar{\varepsilon}_{.jt}; \quad \bar{y}_{ij.-1}\bar{\varepsilon}_{i.t}; \quad \bar{y}_{i..-1}\varepsilon_{ijt}; \quad \bar{y}_{.j.-1}\varepsilon_{ijt}; \quad \bar{y}_{i..-1}\bar{\varepsilon}_{i.t}; \quad \bar{y}_{.j.-1}\bar{\varepsilon}_{.jt}\\
&\Rightarrow \quad \operatorname*{plim}_{N\to\infty}\frac{\sigma^2_\varepsilon}{NT}\cdot\frac{1-\rho^{T-t}}{1-\rho} = 0
\end{aligned}$$

$$\bar{y}_{ij.-1}\bar{\varepsilon}_{..t};\quad \bar{y}_{i..-1}\bar{\varepsilon}_{.jt};\quad \bar{y}_{.j.-1}\bar{\varepsilon}_{i.t};\quad \bar{y}_{i..-1}\bar{\varepsilon}_{..t};\quad \bar{y}_{.j.-1}\bar{\varepsilon}_{..t};\quad \bar{y}_{...-1}\varepsilon_{ijt};$$

$$\bar{y}_{...-1}\bar{\varepsilon}_{.jt};\quad \bar{y}_{...-1}\bar{\varepsilon}_{i.t};\quad \bar{y}_{...-1}\bar{\varepsilon}_{..t}$$

$$\Rightarrow \quad \operatorname*{plim}_{N\to\infty} \frac{\sigma_\varepsilon^2}{N^2T}\cdot\frac{1-\rho^{T-t}}{1-\rho} = 0$$

$$\bar{y}_{ij.-1}\bar{\varepsilon}_{.j.};\quad \bar{y}_{ij.-1}\bar{\varepsilon}_{i..};\quad \bar{y}_{i..-1}\bar{\varepsilon}_{ij.};\quad \bar{y}_{.j.-1}\bar{\varepsilon}_{ij.};\quad \bar{y}_{i..-1}\bar{\varepsilon}_{i..};\quad \bar{y}_{.j.-1}\bar{\varepsilon}_{.j.}$$

$$\Rightarrow \quad \operatorname*{plim}_{N\to\infty} \frac{\sigma_\varepsilon^2}{NT}\cdot\left(\frac{1}{1-\rho}-\frac{1}{T}\cdot\frac{1-\rho^T}{(1-\rho)^2}\right) = 0$$

$$\bar{y}_{ij.-1}\bar{\varepsilon}_{...};\quad \bar{y}_{i..-1}\bar{\varepsilon}_{.j.};\quad \bar{y}_{.j.-1}\bar{\varepsilon}_{i..};\quad \bar{y}_{i..-1}\bar{\varepsilon}_{...};\quad \bar{y}_{.j.-1}\bar{\varepsilon}_{...};\quad \bar{y}_{...-1}\bar{\varepsilon}_{ij.};$$

$$\bar{y}_{...-1}\bar{\varepsilon}_{i..};\quad \bar{y}_{...-1}\bar{\varepsilon}_{.j.};\quad \bar{y}_{...-1}\bar{\varepsilon}_{...}$$

$$\Rightarrow \quad \operatorname*{plim}_{N\to\infty} \frac{\sigma_\varepsilon^2}{N^2T}\cdot\left(\frac{1}{1-\rho}-\frac{1}{T}\cdot\frac{1-\rho^T}{(1-\rho)^2}\right) = 0$$

$$y_{ijt-1}\bar{y}_{i.t-1};\quad y_{ijt-1}\bar{y}_{.jt-1};\quad \bar{y}^2_{i.t-1} = \bar{y}^2_{.jt-1}$$

$$\Rightarrow \quad \operatorname*{plim}_{N\to\infty} \frac{\sigma_\varepsilon^2}{N}\cdot\frac{1-\rho^{2t}}{1-\rho^2} = 0$$

$$y_{ijt-1}\bar{y}_{..t-1};\quad \bar{y}_{i.t-1}\bar{y}_{.jt-1};\quad \bar{y}_{i.t-1}\bar{y}_{..t-1};\quad \bar{y}_{.jt-1}\bar{y}_{..t-1};\quad \bar{y}^2_{..t-1}$$

$$\Rightarrow \quad \operatorname*{plim}_{N\to\infty} \frac{\sigma_\varepsilon^2}{N^2}\cdot\frac{1-\rho^{2t}}{1-\rho^2} = 0$$

$$y_{ijt-1}\bar{y}_{i..-1};\quad y_{ijt-1}\bar{y}_{.j.-1};\quad \bar{y}_{ij.-1}\bar{y}_{i.t-1};\quad \bar{y}_{ij.-1}\bar{y}_{.jt-1};\quad \bar{y}_{i.t-1}\bar{y}_{i..-1};\quad \bar{y}_{.jt-1}\bar{y}_{.j.-1}$$

$$\Rightarrow \quad \operatorname*{plim}_{N\to\infty} \frac{\sigma_\varepsilon^2}{NT(1-\rho^2)}\left((1+\rho)\cdot\frac{1-\rho^t}{1-\rho}+\rho\cdot\frac{\rho^{T+t}-\rho^{T-t}}{1-\rho}\right) = 0$$

$$y_{ijt-1}\bar{y}_{...-1};\quad \bar{y}_{ij.-1}\bar{y}_{..t-1};\quad \bar{y}_{i.t-1}\bar{y}_{.j.-1};\quad \bar{y}_{.jt-1}\bar{y}_{i..-1};\quad \bar{y}_{i.t-1}\bar{y}_{...-1};\quad \bar{y}_{.jt-1}\bar{y}_{...-1};$$

$$\bar{y}_{..t-1}\bar{y}_{i..-1};\quad \bar{y}_{..t-1}\bar{y}_{.j.-1};\quad \bar{y}_{..t-1}\bar{y}_{...-1}$$

$$\Rightarrow \quad \operatorname*{plim}_{N\to\infty} \frac{\sigma_\varepsilon^2}{N^2T(1-\rho^2)}\left((1+\rho)\cdot\frac{1-\rho^t}{1-\rho}+\rho\cdot\frac{\rho^{T+t}-\rho^{T-t}}{1-\rho}\right) = 0$$

$\bar{y}_{ij.-1}\bar{y}_{i..-1}; \quad \bar{y}_{ij.-1}\bar{y}_{.j.-1}; \quad \bar{y}^2_{i..-1}; \quad \bar{y}^2_{.j.-1}$

$$\Rightarrow \quad \operatorname*{plim}_{N\to\infty} \frac{\sigma_\varepsilon^2}{NT(1-\rho)^2}\left(1 - \frac{\rho\cdot(1-\rho^T)\cdot(2+\rho-\rho^{T+1})}{T(1-\rho^2)}\right) = 0$$

$\bar{y}_{ij.-1}\bar{y}_{...-1}; \quad \bar{y}_{i..-1}\bar{y}_{.j.-1}; \quad \bar{y}_{i..-1}\bar{y}_{...-1}; \quad \bar{y}_{.j.-1}\bar{y}_{...-1} = \bar{y}^2_{...-1}$

$$\Rightarrow \quad \operatorname*{plim}_{N\to\infty} \frac{\sigma_\varepsilon^2}{N^2T(1-\rho)^2}\left(1 - \frac{\rho\cdot(1-\rho^T)\cdot(2+\rho-\rho^{T+1})}{T(1-\rho^2)}\right) = 0$$

In order to compute the bias for each model, we only have to collect the terms present in the corresponding transformation and add them up (if necessary with the appropriate multiplicity).

Appendix 6

Matrix form of transformation (34)

$$\begin{aligned}
&\big(I_{N(N-1)N_sT} - \bar{J}_{N-1}\otimes I_{NN_sT} - I_N\otimes\bar{J}_{N-1}\otimes I_{N_sT} - I_{N(N-1)}\otimes\bar{J}_{N_s}\otimes I_T - I_{N(N-1)N_s}\otimes\bar{J}_T \\
&\quad + \frac{N}{N-1}\bar{J}_{N(N-1)}\otimes I_{N_sT} + \bar{J}_{N-1}\otimes I_N\otimes\bar{J}_{N_s}\otimes I_T + \bar{J}_{N-1}\otimes I_{NN_s}\otimes\bar{J}_T \\
&\quad + I_N\otimes\bar{J}_{(N-1)N_s}\otimes I_T + I_N\otimes\bar{J}_{N-1}\otimes I_{N_s}\otimes\bar{J}_T + I_{N(N-1)}\otimes\bar{J}_{N_sT} \\
&\quad - \frac{N}{N-1}\bar{J}_{N(N-1)N_s}\otimes I_T - \frac{N}{N-1}\bar{J}_{N(N-1)}\otimes I_{N_s}\otimes\bar{J}_T \\
&\quad - \bar{J}_{N-1}\otimes I_N\otimes\bar{J}_{N_sT} - I_N\otimes\bar{J}_{(N-1)N_sT} + \frac{N}{N-1}\bar{J}_{N(N-1)N_sT} \\
&\quad - K_{N(N-1)}\otimes\bar{J}_{N_sT} + \frac{1}{N-1}K_{N(N-1)}\otimes\bar{J}_{N_s}\otimes I_T + \frac{1}{N-1}K_{N(N-1)}\otimes I_{N_s}\otimes\bar{J}_T \\
&\quad - \frac{1}{N-1}K_{N(N-1)}\otimes I_{N_sT}\big)
\end{aligned}$$

References

Manuel Arellano and Steve Bond [1991]: Some Tests of Specification for Panel Data: Monte-Carlo Evidence and an Application to Employment Equations, *Review of Economic Studies*, 58, 127-134

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

Pietro Balestra and Jayalakshmi Krishnakumar [2008]: Fixed Effects and Fixed Coefficients Models, in Matyas and Sevestre (eds.), *The Econometrics of Panel Data*, 3rd edition, Sringer Verlag, 23-48.

Badi Baltagi [1995]: *Econometric Analysis of Panel Data*, John Wiley & Sons.

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

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

Peter Davis [2002]: Estimating multi-way error components models with unbalanced data structures, *Journal of Econometrics*, 106, 67-95.

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

Christian Gourieroux and Alain Monfort [1989]: *Statistique et Modeles Econometriques*, Economica, 184-185.

Mark N. Harris, Laszlo Matyas and Patrick Sevestre [2005]: *Dynamic Models for Short Panels*; in Matyas and Sevestre (eds.) The Econometrics of Panel Data, 3rd edition, Springer Verlag, 2005.

Cecilia Hornok [2011]: Gravity or Dummies? The Limits of Identification in Gravity Estimations, *CeFig Working Papers*, 15, September 2011.

Laszlo Matyas [1997]: Proper Econometric Specification of the Gravity Model, *The World Economy*, 20, 363-369.

Laszlo Matyas, Mark N. Harris and Laszlo Konya [2011]: Within Transformations for Three-Way Fixed Effects Models of Trade, unpublished manuscript, 2011/05.

Stephen Nickell [1981]: Biases in Dynamic Models with Fixed Effects, *Econometrica*, 49, 1417-1426.

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

Table 1: The Behaviour of the Proposed Within Transformations in Case of Some Data Problems

Models		**(1)**			**(5)**	**(7)**		**(8)**	**(10)**		**(12)**			**(31)**		
Transformation		Opt (4)	(19)	(11)	(6)	Opt (3)	(22)	(9)	Opt (11)	(20)	Opt (13)	(21)	(23)	Opt (33)	(34)	(35)
CD	**Finite** N, T	+	-	+	+	+	+	+	+	-	+	-	+	+	-	+
	$N \to \infty$	+	-	+	+	+	+	+	+	-	+	-	+	+	-	+
NSF	**Finite** N, T	-	+	-	+	+	+	+	-	+	-	+	+	-	+	+
	$N \to \infty$	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
UBD	**Finite** N, T	-	-	+	+	-	+	+	+	-	-	-	+	-	-	+
	$N \to \infty$	-	-	+	+	-	+	+	+	-	-	-	+	-	-	+

Where: + stands for no bias, *CD*, *NSF* and *UBD* stand for *Complete Data*, *No-Self-Flow Data* and *Unbalanced Data* respectively.

Table 2: The Behaviour of the Proposed Within Transformations in Case of Dynamic Models

Models	(1)		(5)	(7)	(8)	(10)	(12)
Transformation	(3)	(4)	(6)	(3)	(9)	(11)	(13)
Finite T, $N \rightarrow \infty$	-	+	-	-	+	+	-

Where: + stands for no bias.