DEPARTMENT OF ECONOMICS
WORKING PAPERS

economics.ceu.hu

The Estimation of Multi-dimensional Fixed Effects Panel Data Models

by

Laszlo Matyas[1] and Laszlo Balazsi[2]

2012/2

[1] Department of Economics, Central European University, matyas@ceu.hu
[2] Department of Economics, Central European University

Abstract

The paper introduces for the most frequently used three-dimensional fixed effects panel data models the appropriate within estimators. It analyzes the behaviour of these estimators in the case of no-self-flow data, unbalanced data and dynamic autoregressive models. Then the main results are generalised 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.

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

<u>Acknowledgment</u>

Support by Australian Research Council grant DP110103824 is kindly acknowledged, as well as the hospitality of Monash University for Laszlo Matyas. Comments received from the participants of the 18th Panel Data Conference, Paris 2012, were 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 and finally, we draw some conclusions in Section 6.

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}, $i = 1, \dots, N_1$, $j = 1, \dots, N_2$, and $t = 1, \dots, T$. 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 (13) later on). The resulting estimator is usually called Least Squares Dummy Variable (LSDV) estimator. However, it is well known that the

[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.

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 the 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)

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

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

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

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

$$\bar{y}_i = 1/(NT) \sum_j \sum_t y_{ijt}$$
$$\bar{y}_j = 1/(NT) \sum_i \sum_t y_{ijt}$$

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 (11) 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

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

Another variant of this model is

$$y_{ijt} = \beta' x_{ijt} + \alpha_{it} + \varepsilon_{ijt} \tag{10}$$

Here the Within transformation which clears the fixed effects is

$$(y_{ijt} - \bar{y}_{it}) \quad \text{where} \quad \bar{y}_{it} = 1/N \sum_j y_{ijt}$$

The most frequently used variation of this model is

$$y_{ijt} = \beta' x_{ijt} + \alpha_{it} + \alpha_{jt} + \varepsilon_{ijt} \tag{11}$$

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

$$(y_{ijt} - \bar{y}_{jt} - \bar{y}_{it} + \bar{y}_t) \tag{12}$$

Let us notice here that transformation (12) 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

$$y_{ijt} = \beta' x_{ijt} + \gamma_{ij} + \alpha_{it} + \alpha_{jt} + \varepsilon_{ijt} \tag{13}$$

By applying suitable restrictions to model (13) we can obtain the models discussed above. The Within transformation for this model is

$$\begin{aligned}(y_{ijt} - 1/T\sum_t y_{ijt} - 1/N\sum_i y_{ijt} - 1/N\sum_j y_{ijt} + 1/N^2\sum_i\sum_j y_{ijt}\\ + 1/(NT)\sum_i\sum_t y_{ijt} + 1/(NT)\sum_j\sum_t y_{ijt} - 1/(N^2T)\sum_i\sum_j\sum_t y_{ijt})\end{aligned} \tag{14}$$

or in a shorter form

$$(y_{ijt} - \bar{y}_{ij} - \bar{y}_{jt} - \bar{y}_{it} + \bar{y}_t + \bar{y}_j + \bar{y}_i - \bar{y})$$

We can write up these Within transformations in a more compact matrix form using *Davis'* [2002] and *Hornok's* [2011] approach. Model (13) in matrix form is

$$y = X\beta + \tilde{D}_1\gamma + \tilde{D}_2\alpha + \tilde{D}_3\alpha_* + \varepsilon \tag{15}$$

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 $D = (\tilde{D}_1, \tilde{D}_2, \tilde{D}_3)$, $Q_D = D(D'D)^{-1}D'$ and $P_D = I - Q_D$. 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 (14). By appropriate restrictions on the parameters of (15) we get back the previously analysed Within transformations. Now transforming model (15) with transformation (14) 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 and eliminate the fixed effects. However, this is not the case unfortunately for models (1) (transformation (4)), (11) and (13). Let us have a closer look at the difficulty. For model (1) and transformation (4), instead of canceled out fixed effects, we end up with the following remaining fixed effects

$$\begin{aligned}
\alpha_i^* = \alpha_i &- \frac{1}{(N-1)T} \cdot (N-1)T \cdot \alpha_i - \frac{1}{(N-1)T} \sum_{i=1;\, i\neq j}^{N} T \cdot \alpha_i \\
&- \frac{1}{N(N-1)} \sum_{i=1}^{N} (N-1) \cdot \alpha_i + \frac{2}{N(N-1)T} \sum_{i=1}^{N} (N-1)T \cdot \alpha_i \\
= \alpha_i - \alpha_i &- \frac{1}{N-1} \sum_{i=1;\, i\neq j}^{N} \alpha_i + \frac{1}{N} \sum_{i=1}^{N} \alpha_i = \frac{1}{N}\alpha_j - \frac{1}{N(N-1)} \sum_{i=1;\, i\neq j}^{N} \alpha_i
\end{aligned}$$

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

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

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

and for the time effects

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

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

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

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 $N \to \infty$. On the other hand, luckily, transformation (3) as seen earlier, although not optimal, leads to an unbiased Within estimator for model (1) and remains so even in the lack of self flow data.

Now let us continue with model (11). After the Within transformation (12), instead of canceled out fixed effects we end up with the following remaining fixed effects

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

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

and

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

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

As long as the α^* and γ^* parameters are not zero, the Within estimators will be biased. Similarly for model (13), the remaining fixed effects are now

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

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

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

but

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

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

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

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

and, finally

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

$$+ \frac{1}{N(N-1)}\sum_{j=1}^{N}(N-1)\tilde{\alpha}_{jt} + \frac{1}{(N-1)T}\sum_{t=1}^{T}(N-1)\tilde{\alpha}_{jt}$$

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

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

where in order to avoid confusion with the two similar α fixed effects α_{jt} is now denoted by $\tilde{\alpha}_{jt}$. 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 go to zero, which implies that the bias of the Within estimators go to zero as well.

Fortunately, however, there is good news as well. For both models (11) and (13) there is a transformations which wipes out the fixed effects, and so remains unbiased even in this case. For model (11) this can be written up as

$$\begin{aligned} &y_{ijt} - \bar{y}_{it} - \bar{y}_{jt} + \bar{y}_t + \frac{1}{N-1}\bar{y}_t - \frac{1}{N-1}y_{jit} = \\ &y_{ijt} - \bar{y}_{it} - \bar{y}_{jt} + \frac{N}{N-1}\bar{y}_t - \frac{1}{N-1}y_{jit} \end{aligned} \tag{16}$$

or in matrix form

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

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 jith place, that is the ijth row is in fact

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

For model (13) the appropriate transformation is

$$\begin{aligned} &y_{ijt} - \bar{y}_{jt} - \bar{y}_{it} - \bar{y}_{ij} + \bar{y}_t + \bar{y}_j + \bar{y}_i - \bar{y} - \frac{1}{N-1}\bar{y} + \frac{1}{N-1}\bar{y}_t + \frac{1}{N-1}\bar{y}_{ji} - \frac{1}{N-1}y_{jit} = \\ &y_{ijt} - \bar{y}_{jt} - \bar{y}_{it} - \bar{y}_{ij} + \frac{N}{N-1}\bar{y}_t + \bar{y}_j + \bar{y}_i - \frac{N}{N-1}\bar{y} + \frac{1}{N-1}\bar{y}_{ji} - \frac{1}{N-1}y_{jit} \end{aligned} \tag{17}$$

or, again, in matrix form

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

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 (16) and (17) 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), (10) and (11) 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 (13) 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))

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

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

and

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

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 α_i^*, γ_j^* and λ_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 (12) 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 (13) and transformation (14). The remaining fixed effects are now

$$\gamma_{ij}^* = \gamma_{ij} - \frac{1}{T_{ij}}\sum_{t=1}^{T_{ij}}\gamma_{ij} - \frac{1}{N}\sum_{i=1}^{N}\gamma_{ij} - \frac{1}{N}\sum_{j=1}^{N}\gamma_{ij} + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\gamma_{ij} +$$
$$+ \frac{1}{\sum_{i=1}^{N}T_{ij}}\sum_{i=1}^{N}\sum_{t=1}^{T_{ij}}\gamma_{ij} + \frac{1}{\sum_{j=1}^{N}T_{ij}}\sum_{j=1}^{N}\sum_{t=1}^{T_{ij}}\gamma_{ij} - \frac{1}{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T_{ij}}\gamma_{ij}$$
$$= \gamma_{ij} - \gamma_{ij} - \frac{1}{N}\sum_{i=1}^{N}\gamma_{ij} - \frac{1}{N}\sum_{j=1}^{N}\gamma_{ij} + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\gamma_{ij} + \frac{1}{\sum_{i=1}^{N}T_{ij}}\sum_{i=1}^{N}\gamma_{ij}T_{ij} +$$
$$+ \frac{1}{\sum_{j=1}^{N}T_{ij}}\sum_{j=1}^{N}\gamma_{ij}T_{ij} - \frac{1}{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\gamma_{ij}T_{ij}$$
$$= -\frac{1}{N}\sum_{i=1}^{N}\gamma_{ij} - \frac{1}{N}\sum_{j=1}^{N}\gamma_{ij} + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\gamma_{ij} + \frac{1}{\sum_{i=1}^{N}T_{ij}}\sum_{i=1}^{N}\gamma_{ij}T_{ij} +$$
$$+ \frac{1}{\sum_{j=1}^{N}T_{ij}}\sum_{j=1}^{N}\gamma_{ij}T_{ij} - \frac{1}{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\gamma_{ij}T_{ij}$$

$$\alpha_{it}^* = \alpha_{it} - \frac{1}{T_{ij}}\sum_{t=1}^{T_{ij}}\alpha_{it} - \frac{1}{N}\sum_{i=1}^{N}\alpha_{it} - \frac{1}{N}\sum_{j=1}^{N}\alpha_{it} + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_{it} +$$
$$+ \frac{1}{\sum_{i=1}^{N}T_{ij}}\sum_{i=1}^{N}\sum_{t=1}^{T_{ij}}\alpha_{it} + \frac{1}{\sum_{j=1}^{N}T_{ij}}\sum_{j=1}^{N}\sum_{t=1}^{T_{ij}}\alpha_{it} - \frac{1}{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T_{ij}}\alpha_{it}$$
$$= \alpha_{it} - \frac{1}{T_{ij}}\sum_{t=1}^{T_{ij}}\alpha_{it} - \frac{1}{N}\sum_{i=1}^{N}\alpha_{it} - \alpha_{it} + \frac{1}{N}\sum_{i=1}^{N}\alpha_{it} +$$
$$+ \frac{1}{\sum_{i=1}^{N}T_{ij}}\sum_{i=1}^{N}\sum_{t=1}^{T_{ij}}\alpha_{it} + \frac{1}{\sum_{j=1}^{N}T_{ij}}\sum_{j=1}^{N}\sum_{t=1}^{T_{ij}}\alpha_{it} - \frac{1}{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T_{ij}}\alpha_{it}$$
$$= -\frac{1}{T_{ij}}\sum_{t=1}^{T_{ij}}\alpha_{it} + \frac{1}{\sum_{i=1}^{N}T_{ij}}\sum_{i=1}^{N}\sum_{t=1}^{T_{ij}}\alpha_{it} + \frac{1}{\sum_{j=1}^{N}T_{ij}}\sum_{j=1}^{N}\sum_{t=1}^{T_{ij}}\alpha_{it} - \frac{1}{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T_{ij}}\alpha_{it}$$

and

$$\alpha^*_{jt} = \alpha_{jt} - \frac{1}{T_{ij}}\sum_{t=1}^{T_{ij}}\alpha_{jt} - \frac{1}{N}\sum_{i=1}^{N}\alpha_{jt} - \frac{1}{N}\sum_{j=1}^{N}\alpha_{jt} + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_{jt} +$$

$$+ \frac{1}{\sum_{i=1}^{N} T_{ij}}\sum_{i=1}^{N}\sum_{t=1}^{T_{ij}}\alpha_{jt} + \frac{1}{\sum_{j=1}^{N} T_{ij}}\sum_{j=1}^{N}\sum_{t=1}^{T_{ij}}\alpha_{jt} - \frac{1}{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T_{ij}}\alpha_{jt}$$

$$= \alpha_{jt} - \frac{1}{T_{ij}}\sum_{t=1}^{T_{ij}}\alpha_{jt} - \alpha_{jt} - \frac{1}{N}\sum_{i=1}^{N}\alpha_{jt} + \frac{1}{N}\sum_{i=1}^{N}\alpha_{jt} + \frac{1}{\sum_{i=1}^{N} T_{ij}}\sum_{i=1}^{N}\sum_{t=1}^{T_{ij}}\alpha_{jt} +$$

$$+ \frac{1}{\sum_{j=1}^{N} T_{ij}}\sum_{j=1}^{N}\sum_{t=1}^{T_{ij}}\alpha_{jt} - \frac{1}{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T_{ij}}\alpha_{jt}$$

$$= -\frac{1}{T_{ij}}\sum_{t=1}^{T_{ij}}\alpha_{jt} + \frac{1}{\sum_{i=1}^{N} T_{ij}}\sum_{i=1}^{N}\sum_{t=1}^{T_{ij}}\alpha_{jt} + \frac{1}{\sum_{j=1}^{N} T_{ij}}\sum_{j=1}^{N}\sum_{t=1}^{T_{ij}}\alpha_{jt} - \frac{1}{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T_{ij}}\alpha_{jt}$$

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 $N \to \infty$. 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 (13) 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 (15), $\tilde{D}_1$ 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, $(t = 1 \dots T^*)$ 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

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

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

$$P^a = \left(I_T - D_1 \Delta_{N^2}^{-1} D_1' \right) - \bar{D}^a Q^{a-} \bar{D}^{a'} \tag{18}$$

where Q^{a-} denotes the 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 (18) 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 (13) 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,

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

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

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

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

$$B \equiv I_T - D_1 (D_1' D_1)^{-1} D_1' \text{ and}$$

$$C \equiv B - (B D_2^b)[(B D_2^b)'(B D_2^b)]^{-}(B D_2^b)'$$

the appropriate transformation for model (13) now is

$$P^b \equiv C - (C D_3)[(C D_3)'(C D_3)]^{-}(C D_3)' \tag{19}$$

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

It is worth noting that both transformations (18) and (19) 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 (19) 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 (19) is

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

where a_{ij}^b and $\tilde{a}_{ij}$ are the column vectors corresponding to observations ij from matrixes $A^b \equiv \left(D_2^b\right)' D_1$ and $\tilde{A} \equiv D_3' D_1$ respectively. $\bar{\phi}_{it}^b$ 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, $\bar{\xi}$.[4]

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 finite sample bias for each of the models introduced in Section 2.

[4] 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.

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

$$y_{ijt} = \rho y_{ijt-1} + \gamma_{ij} + \varepsilon_{ijt} \tag{20}$$

With backward substitution we get

$$y_{ijt} = \rho^t y_{ij0} + \frac{1-\rho^t}{1-\rho}\gamma_{ij} + \sum_{k=0}^{t} \rho^k \varepsilon_{ijt-k} \tag{21}$$

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 (20) after applying the appropriate Within transformation, that is the correlation between $(y_{ijt-1} - \bar{y}_{ij-1})$ where $\bar{y}_{ijt-1} = 1/T \sum_t y_{ijt-1}$ and $(\varepsilon_{ijt} - \bar{\varepsilon}_{ij})$ where $\bar{\varepsilon}_{ij} = 1/T \sum_t \varepsilon_{ij}$. 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, not even 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.

Model (20) can of course be expanded to have exogenous explanatory variables as well

$$y_{ijt} = \rho y_{ijt-1} + x'_{ijt}\beta + \gamma_{ij} + \varepsilon_{ijt} \tag{22}$$

Let us turn now to the derivation of the finite sample bias and denote in general any of the above Within transformations by $\bar{y}_{trans}$. Using this notation we can derive the general form of the bias using *Nickell–type* calculations. Starting from the simple first order autoregressive model (20) introduced above we get

$$(y_{ijt} - \bar{y}_{trans}) = \rho(y_{ijt-1} - \bar{y}_{trans-1}) + (\varepsilon_{ijt} - \bar{\varepsilon}_{trans}) \tag{23}$$

Using OLS to estimate ρ, we get

$$\widehat{\rho}_t = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} (y_{ijt-1} - \bar{y}_{trans-1}) \cdot (y_{ijt} - \bar{y}_{trans})}{\sum_{i=1}^{N}\sum_{j=1}^{N} (y_{ijt-1} - \bar{y}_{trans-1})^2} \tag{24}$$

So in the expectations we have

$$E(\widehat{\rho} - \rho) = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} E\,(y_{ijt-1} - \bar{y}_{trans-1})\,(\varepsilon_{ijt} - \bar{\varepsilon}_{trans})}{\sum_{i=1}^{N}\sum_{j=1}^{N} E\,(y_{ijt-1} - \bar{y}_{trans-1})^2} \tag{25}$$

Continuing with model (20) 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 (25) from above we get

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

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

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

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

And for the denominator of the bias in (25)

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

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

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

So the finite sample bias for this model is

$$E\left[\hat{\rho} - \rho\right] = \frac{-\frac{\sigma_\varepsilon^2}{T}\cdot\left(\frac{1-\rho^{t-1}}{1-\rho}\right) - \frac{\sigma_\varepsilon^2}{T}\cdot\left(\frac{1-\rho^{T-t}}{1-\rho}\right) + \frac{\sigma_\varepsilon^2}{T}\cdot\left(\frac{1}{1-\rho} - \frac{1}{T}\cdot\frac{1-\rho^T}{(1-\rho)^2}\right)}{\sigma_\varepsilon^2\cdot\left(\frac{1-\rho^{2t}}{1-\rho^2}\right) - A^* + B^*}$$

where

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

and

$$B^* = \frac{\sigma_\varepsilon^2}{T(1-\rho)^2}\left(1 - \frac{2\rho(1-\rho^T)}{T(1-\rho^2)} + \frac{2\rho^{T+2} - \rho^{2(T+1)} - \rho^2}{1-\rho^2}\right)$$

It can be seen that these results are very similar to the original Nickell results, and the bias is persistent even in the semi-asymptotic case when $N \to \infty$.

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 the Appendix) we get for the finite sample bias

$$E\left[\hat{\rho} - \rho\right] = \frac{\left(\frac{1-N^2}{N^2}\right)\frac{\sigma_\varepsilon^2}{T}\frac{1-\rho^{t-1}}{1-\rho} + \left(\frac{1-N^2}{N^2}\right)\frac{\sigma_\varepsilon^2}{T}\frac{1-\rho^{T-t}}{1-\rho} + A^*}{\left(\frac{N^2-1}{N^2}\right)\cdot\sigma_\varepsilon^2\frac{1-\rho^{2t}}{1-\rho^2} - B^* + C^*}$$

where

$$A^* = \left(\frac{N^2-1}{N^2}\right)\frac{\sigma_\varepsilon^2}{T}\left(\frac{1}{1-\rho} - \frac{1}{T}\frac{1-\rho^T}{(1-\rho)^2}\right)$$

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

and

$$C^* = \left(\frac{N^2-1}{N^2}\right)\frac{\sigma_\varepsilon^2}{T(1-\rho)^2}\left(1 - \frac{2\rho(1-\rho^T)}{T(1-\rho^2)} + \frac{2\rho^{T+2} - \rho^{2(T+1)} - \rho^2}{1-\rho^2}\right)$$

It is worth noticing that in the semi-asymptotic case as $N \to \infty$ we get back the bias derived above for model (20).

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 the Appendix)

$$E\left[\hat{\rho} - \rho\right] = \frac{\left(\frac{2-2N}{N^2}\right)\cdot\frac{\sigma_\epsilon^2}{T}\cdot\frac{1-\rho^{t-1}}{1-\rho} + \left(\frac{2-2N}{N^2}\right)\cdot\frac{\sigma_\epsilon^2}{T}\cdot\frac{1-\rho^{T-t}}{1-\rho} + A^{**}}{\left(\frac{N^2-1}{N^2}\right)\cdot\sigma_\varepsilon^2\frac{1-\rho^{2t}}{1-\rho^2} + B^{**} + C^{**}}$$

where

$$A^{**} = \left(\frac{2N-2}{N^2}\right) \cdot \frac{\sigma_\epsilon^2}{T} \cdot \left(\frac{1}{1-\rho} - \frac{1}{T} \cdot \frac{1-\rho^T}{(1-\rho)^2}\right)$$

$$B^{**} = \left(\frac{4-4N}{N^2}\right) \cdot \frac{\sigma_\epsilon^2}{T(1-\rho^2)} \left(\frac{1-\rho^t}{1-\rho} + \rho\frac{1-\rho^{T-t}}{1-\rho} - \rho^{t+1} \cdot \frac{1+\rho^T}{1-\rho}\right)$$

and

$$C^{**} = \left(\frac{2N-4}{N^2}\right) \frac{\sigma_\epsilon^2}{T(1-\rho)^2} \left(1 - \frac{2\rho(1-\rho^T)}{T(1-\rho^2)} + \frac{2\rho^{T+2} - \rho^{2(T+1)} - \rho^2}{1-\rho^2}\right)$$

It can be seen as $N \to \infty$ 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), (10) and (11) 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

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

The bias then can be expressed as

$$E\left[\bar{\rho} - \rho\right] = \frac{E\left[y_{ijt-1}\varepsilon_{ijt}\right] - E\left[y_{ijt-1}\bar{\varepsilon}_{jt}\right] - E\left[\bar{y}_{jt-1}\varepsilon_{ijt}\right] + E\left[\bar{y}_{jt-1}\bar{\varepsilon}_{jt}\right]}{E\left[y_{ijt-1}^2\right] - 2 \cdot E\left[y_{ijt-1}\bar{y}_{jt-1}\right] + E\left[\bar{y}_{jt-1}^2\right]}$$

It can easily be seen that the expected value of the numerator is 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 bias is in fact nil. The same arguments are valid for models (10) and (11) as well.

And finally, let us turn to model (13)

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

The Within transformation gives

$$\left(y_{ijt} - \bar{y}_{ij} - \bar{y}_{jt} - \bar{y}_{it} + \bar{y}_t + \bar{y}_j + \bar{y}_i - \bar{y}\right)$$

so we get

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

And for the finite sample bias of this model we get

$$
E[\hat{\rho} - \rho] = \frac{\left(\frac{-(N-1)^2}{N^2}\right) \cdot \frac{\sigma_\varepsilon^2}{T} \cdot \frac{1-\rho^{t-1}}{1-\rho} + \left(\frac{-(N-1)^2}{N^2}\right) \cdot \frac{\sigma_\varepsilon^2}{T} \cdot \frac{1-\rho^{T-t}}{1-\rho} + A^*}{\left(\frac{(N-1)^2}{N^2}\right) \sigma_\varepsilon^2 \frac{1-\rho^{2t}}{1-\rho^2} + B^* + C^*}
$$

where

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

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

and

$$
C^* = \left(\frac{(N-1)^2}{N^2}\right) \frac{\sigma_\varepsilon^2}{T(1-\rho)^2} \left(1 - \frac{2\rho(1-\rho^T)}{T(1-\rho^2)} + \frac{2\rho^{T+2} - \rho^{2(T+1)} - \rho^2}{1-\rho^2}\right)
$$

It is clear that if N goes to infinity and T is finite, then we get back the bias of model (5).

As seen above, we have problems with the estimation and N inconsistency of models (5), (7) and (13) 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

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

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

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

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

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

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

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

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

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

Continuing now with model (13), 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

$$(\Delta y_{ijt} - \Delta\tilde{y}_{jt} - \Delta\tilde{y}_{it} + \Delta\tilde{y}_t) = \rho(\Delta y_{ijt-1} - \Delta\tilde{y}_{jt-1} - \Delta\tilde{y}_{it-1} + \Delta\tilde{y}_{t-1}) + (\Delta\varepsilon_{ijt} - \Delta\tilde{\varepsilon}_{jt} - \Delta\tilde{\varepsilon}_{it} + \Delta\tilde{\varepsilon}_t)$$

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 (13), (15) with pair-wise interaction effects:

$$y_{ijst} = x'_{ijst}\beta + \gamma^0_{ijs} + \gamma^1_{ijt} + \gamma^2_{jst} + \gamma^3_{ist} + \varepsilon_{ijst} \tag{26}$$

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

$$\gamma^0_{IS} + \sum_{k=1}^{M} \gamma^M_{i_k,t} \tag{27}$$

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 (26) $M = 3$). In the case of unbalanced panel data $t = 1, \ldots, T_{IS}$.

The Within transformation for model (26) is

$$(y_{ijst} - \bar{y}_{jst} - \bar{y}_{ist} - \bar{y}_{ijt} - \bar{y}_{ijs} + \bar{y}_{st} + \bar{y}_{jt} + \bar{y}_{js} + \bar{y}_{it} + \bar{y}_{is} + \bar{y}_{ij} - \bar{y}_t - \bar{y}_s - \bar{y}_j - \bar{y}_i + \bar{y})$$

where

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

$$\bar{y}_i = \frac{1}{N_j N_s T} \sum_{j=1}^{N} \sum_{s=1}^{N_s} \sum_{t=1}^{T} y_{ijst}$$

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

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

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

$$
\begin{aligned}
P_D = &\left(I_{N_i} - \bar{J}_{N_i}\right) \otimes \left(I_{N_j} - \bar{J}_{N_j}\right) \otimes \left(I_{N_s} - \bar{J}_{N_s}\right) \otimes \left(I_T - \bar{J}_T\right) = \\
&\Big(I_{N_iN_jN_sT} - \left(\bar{J}_{N_i} \otimes I_{N_jN_sT}\right) - \left(I_{N_i} \otimes \bar{J}_{N_j} \otimes I_{N_sT}\right) - \left(I_{N_iN_j} \otimes \bar{J}_{N_s} \otimes I_T\right) \\
&- \left(I_{N_iN_jN_s} \otimes \bar{J}_T\right) + \left(\bar{J}_{N_iN_j} \otimes I_{N_sT}\right) + \left(\bar{J}_{N_i} \otimes I_{N_j} \otimes \bar{J}_{N_s} \otimes I_T\right) \\
&+ \left(\bar{J}_{N_i} \otimes I_{N_jN_s} \otimes \bar{J}_T\right) + \left(I_{N_i} \otimes \bar{J}_{N_jN_s} \otimes I_T\right) + \left(I_{N_i} \otimes \bar{J}_{N_j} \otimes I_{N_s} \otimes \bar{J}_T\right) \\
&+ \left(I_{N_iN_j} \otimes \bar{J}_{N_sT}\right) - \left(\bar{J}_{N_iN_jN_s} \otimes I_T\right) - \left(\bar{J}_{N_iN_j} \otimes I_{N_s} \otimes \bar{J}_T\right) - \left(\bar{J}_{N_i} \otimes I_{N_j} \otimes \bar{J}_{N_sT}\right) \\
&- \left(I_{N_j} \otimes \bar{J}_{N_jN_sT}\right) + \bar{J}_{N_iN_jN_sT}\Big)
\end{aligned}
\tag{28}
$$

The generalization of the Within transformation for model (26) for any higher dimensions can be done using the general form (27). There are basically two types of fixed effect, γ^0_{IS}, 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^0_{IS}, 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 $s_1, \dots s_M$ (in the three-dimensional case $s_1 = i$, $s_2 = j$ and $s_3 = t$). The Within transformation then is

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

where

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

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 (28) 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 both transformations (16) and (17) can also be generalized to any higher dimension. Let us see, for example transformation (17) in the four dimensional case (with $N_i = N_j = N$), which now is

$$(y_{ijst} - \bar{y}_{jst} - \bar{y}_{ist} - \bar{y}_{ijt} - \bar{y}_{ijs} + \frac{N}{N-1}\bar{y}_{st} + \bar{y}_{jt} + \bar{y}_{js} + \bar{y}_{it} + \bar{y}_{is} + \bar{y}_{ij} - \frac{N}{N-1}\bar{y}_t$$
$$- \frac{N}{N-1}\bar{y}_s - \bar{y}_j - \bar{y}_i + \frac{N}{N-1}\bar{y} - \frac{1}{N-1}\bar{y}_{ji} + \frac{1}{N-1}\bar{y}_{jis} + \frac{1}{N-1}\bar{y}_{jit} - \frac{1}{N-1}y_{jist})$$

or in matrix form

$$\begin{aligned}
\big(& I_{N(N-1)N_sT} - \bar{J}_{N-1} \otimes I_{NN_sT} - I_N \otimes \bar{J}_{N-1} \otimes I_{N_sT} - I_{N(N-1)} \otimes \bar{J}_{N_s} \otimes I_T - I_{N(N-1)N_s} \otimes \bar{J}_T \\
& + \frac{N}{N-1}\bar{J}_{N(N-1)} \otimes I_{N_sT} + \bar{J}_{N-1} \otimes I_N \otimes \bar{J}_{N_s} \otimes I_T + \bar{J}_{N-1} \otimes I_{NN_s} \otimes \bar{J}_T \\
& + I_N \otimes \bar{J}_{(N-1)N_s} \otimes I_T + I_N \otimes \bar{J}_{N-1} \otimes I_{N_s} \otimes \bar{J}_T + I_{N(N-1)} \otimes \bar{J}_{N_sT} \\
& - \frac{N}{N-1}\bar{J}_{N(N-1)N_s} \otimes I_T - \frac{N}{N-1}\bar{J}_{N(N-1)} \otimes I_{N_s} \otimes \bar{J}_T \\
& - \bar{J}_{N-1} \otimes I_N \otimes \bar{J}_{N_sT} - I_N \otimes \bar{J}_{(N-1)N_sT} + \frac{N}{N-1}\bar{J}_{N(N-1)N_sT} \\
& - K_{N(N-1)} \otimes \bar{J}_{N_sT} + \frac{1}{N-1}K_{N(N-1)} \otimes \bar{J}_{N_s} \otimes I_T + \frac{1}{N-1}K_{N(N-1)} \otimes I_{N_s} \otimes \bar{J}_T \\
& - \frac{1}{N-1}K_{N(N-1)} \otimes I_{N_sT}\big)
\end{aligned} \tag{29}$$

We can go further along the above lines, and generalize transformation (19) into any higher dimensional setup. Using the dummy variable structure (27), let the dummy variables matrixes for the $M+1$ fixed effects be denoted by $D^c_1, D^c_2, \dots D^c_{M+1}$ respectively, and let $P^{(k)}$ be the transformation which clears out the first k fixed effects, namely $P^{(k)} \cdot (D^c_1, D^c_2, \dots, D^c_k) = (0, 0, \dots, 0)$. The appropriate within transformation to clear out the first $k+1$ fixed effects then is

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

where

$$P^{(1)} = I_T - D^c_1 \left((D^c_1)' D^c_1\right)^{-1} (D^c_1)'$$

Now, it should be clear from the induction, that the appropriate within transformation for model (26)–(27) is

$$P^c = P^{(M+1)} = P^{(M)} - \left(P^{(M)} D^c_{M+1}\right) \left[\left(P^{(M)} D^c_{M+1}\right)' \left(P^{(M)} D^c_{M+1}\right)\right]^{-} \left(P^{(M)} D^c_{M+1}\right)' \tag{30}$$

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. So far the approach has been to transform the models in such a way that the fixed effects drop out, and then estimate the transformed models with OLS. Now, however, after the appropriate transformation the model has to be estimated by Feasible GLS (FGLS) instead of OLS, as we have to take into account its covariance structure.

First, we 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

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

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

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

For transformation (6) we have

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

and

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

For transformation (9) we get

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

and

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

For transformation (12) 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

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

And finally, for transformation (14) 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, so difficulties lay ahead. Let us start with the simples 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 $\rho_{(1)}$

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

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

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

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

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

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

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

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 (12):

$$E\left[u_{ijt}^2\right] = E\left[(\varepsilon_{ijt} - \bar{\varepsilon}_{ij} - \bar{\varepsilon}_t + \bar{\varepsilon})^2\right] =$$

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

$$E\left[u_{ijt}^{*2}\right] = E\left[(\varepsilon_{ijt} - \bar{\varepsilon}_{it} - \bar{\varepsilon}_{jt} + \bar{\varepsilon}_t)^2\right] =$$

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

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

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

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

$$E\left[\bar{u}_{jt}^2\right] = E\left[(\bar{\varepsilon}_{jt} - \bar{\varepsilon}_j - \bar{\varepsilon}_t + \bar{\varepsilon})^2\right] =$$

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

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

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

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

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

$$E\left[\bar{u}_{it}^2\right] = E\left[\left(\bar{\varepsilon}_{it} - \bar{\varepsilon}_i - \bar{\varepsilon}_t + \bar{\varepsilon}\right)^2\right] = \frac{(N-2)(T-1)}{N^2T}\sigma_{\varepsilon i}^2 + \frac{T-1}{N^3T}\sum_{i=1}^{N}\sigma_{\varepsilon i}^2 -$$

$$- \frac{(N-1)(T-1)}{N^2T}\rho_{(2)} + \frac{(N-1)^2(T-1)}{N^2T}\rho_{(1)}$$

And so

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

which leads to

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

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

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)

$$E\left[(u_{ijt}^a)^2\right] = E\left[\left(\varepsilon_{ijt} - \bar{\varepsilon}_{ij} - \bar{\varepsilon}_{it} + \bar{\varepsilon}_i\right)^2\right] = \frac{(N-1)(T-1)}{NT}\sigma_{\varepsilon i}^2 - \frac{(N-1)(T-1)}{NT}\rho_{(1)}$$

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

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

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

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

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

It can be noticed that

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

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

The other components can easily be derived

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

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

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

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

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

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

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

In this way we can estimate ρ

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

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

$$\begin{aligned}
\hat{\sigma}_{\varepsilon i}^2 &= \frac{N^2}{N-2}\left\{E\left[\bar{\hat{u}}_{it}^2\right] - \frac{1}{N^2(N-1)}\sum_{i=1}^{N}E\left[\hat{u}_{ijt}^2\right]\right\} - \frac{N(N-1)(N-2)+1}{N^3}\hat{\rho} = \\
&= \frac{N^2}{(N-2)T}\sum_{t=1}^{T}\left(\frac{1}{N}\sum_{j=1}^{N}\hat{u}_{ijt}\right)^2 - \frac{1}{N(N-1)(N-2)T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\hat{u}_{ijt}^2 - \\
&\quad - \frac{N(N-1)(N-2)+1}{N^3}\hat{\rho} = \\
&= \frac{1}{(N-2)T}\sum_{t=1}^{T}\left(\sum_{j=1}^{N}\hat{u}_{ijt}\right)^2 - \frac{1}{N(N-1)(N-2)T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\hat{u}_{ijt}^2 - \\
&\quad - \frac{N(N-1)(N-2)+1}{N^3}\hat{\rho}
\end{aligned}$$

Let us continue next with model (10). Now we have

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

We can transform u_{ijt} further by taking the averages with respect to i and then compute the respective variances

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

So we can estimate ρ as

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

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

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

and $\sigma_{\varepsilon i}^2$ as

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

We still have two models, namely (11) and (13), 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 $\sigma_{\varepsilon i}^2$ variances.

For model (11) we have

$$E\left[u_{ijt}^2\right] = E\left[\left(\varepsilon_{ijt} - \bar{\varepsilon}_{it} - \bar{\varepsilon}_{jt} + \bar{\varepsilon}_t\right)^2\right] = \frac{(N-1)(N-2)}{N^2}\sigma_{\varepsilon i}^2 + \frac{N-1}{N^3}\sum_{i=1}^{N}\sigma_{\varepsilon i}^2$$

Taking the averages with respect to i

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

As a result,

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

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

We can proceed likewise for model (13)

$$E\left[u_{ijt}^2\right] = E\left[\left(\varepsilon_{ijt} - \bar{\varepsilon}_{it} - \bar{\varepsilon}_{jt} - \bar{\varepsilon}_{ij} + \bar{\varepsilon}_i + \bar{\varepsilon}_j + \bar{\varepsilon}_t - \bar{\varepsilon}\right)^2\right] =$$
$$= \frac{(N-1)(N-2)(T-1)}{N^2T}\sigma_{\varepsilon i}^2 + \frac{(N-1)(T-1)}{N^3T}\sum_{i=1}^{N}\sigma_{\varepsilon i}^2$$

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

Finite sample bias derivations for the dynamic model.

Model (1)

Now for model (1) transformation (3) leads to

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

Deriving the expected values, in the numerator we find

$$\begin{aligned}
E[y_{ijt-1}\varepsilon_{ijt}] &= E[y_{ijt-1}\bar{\varepsilon}_t] = E[\bar{y}_{t-1}\varepsilon_{ijt}] = E[\bar{y}_{t-1}\bar{\varepsilon}_t] = 0 \\
E[y_{ijt-1}\bar{\varepsilon}_{ij}] &= \frac{\sigma_\varepsilon^2}{T}\frac{1-\rho^{t-1}}{1-\rho} \\
E[y_{ijt-1}\bar{\varepsilon}] &= E[\bar{y}_{t-1}\bar{\varepsilon}_{ij}] = E[\bar{y}_{t-1}\bar{\varepsilon}] = \frac{\sigma_\varepsilon^2}{N^2T}\frac{1-\rho^{t-1}}{1-\rho} \\
E[\bar{y}_{ij-1}\varepsilon_{ijt}] &= \frac{\sigma_\varepsilon^2}{T}\frac{1-\rho^{T-t}}{1-\rho} \\
E[\bar{y}_{ij-1}\bar{\varepsilon}_t] &= E[\bar{y}_{-1}\varepsilon_{ijt}] = E[\bar{y}_{-1}\bar{\varepsilon}_t] = \frac{\sigma_\varepsilon^2}{N^2T}\frac{1-\rho^{T-t}}{1-\rho} \\
E[\bar{y}_{ij-1}\bar{\varepsilon}_{ij}] &= \frac{\sigma_\varepsilon^2}{T} \cdot \left(\frac{1}{1-\rho} - \frac{1}{T} \cdot \frac{1-\rho^T}{(1-\rho)^2}\right) \\
E[\bar{y}_{ij-1}\bar{\varepsilon}] &= E[\bar{y}_{-1}\bar{\varepsilon}_{ij}] = E[\bar{y}_{-1}\bar{\varepsilon}] = \frac{\sigma_\varepsilon^2}{N^2T} \cdot \left(\frac{1}{1-\rho} - \frac{1}{T} \cdot \frac{1-\rho^T}{(1-\rho)^2}\right)
\end{aligned}$$

and in the denominator

$$\begin{aligned}
E[y_{ijt-1}^2] &= \sigma_\varepsilon^2 \cdot \frac{1-\rho^{2t}}{1-\rho^2} \\
E[y_{ijt-1}\bar{y}_{t-1}] &= E[\bar{y}_{t-1}^2] = \frac{\sigma_\varepsilon^2}{N^2} \cdot \frac{1-\rho^{2t}}{1-\rho^2} \\
E[y_{ijt-1}\bar{y}_{ij-1}] &= \frac{\sigma_\varepsilon^2}{T(1-\rho^2)}\left(\frac{1-\rho^t}{1-\rho} + \rho\frac{1-\rho^{T-t}}{1-\rho} - \rho^{t+1} \cdot \frac{1+\rho^T}{1-\rho}\right) \\
E[y_{ijt-1}\bar{y}_{-1}] &= E[\bar{y}_{ij-1}\bar{y}_{t-1}] = \\
&\qquad E[\bar{y}_{t-1}\bar{y}_{-1}] = \frac{\sigma_\varepsilon^2}{N^2T(1-\rho^2)}\left(\frac{1-\rho^t}{1-\rho} + \rho\frac{1-\rho^{T-t}}{1-\rho} - \rho^{t+1} \cdot \frac{1+\rho^T}{1-\rho}\right) \\
E[\bar{y}_{ij-1}^2] &= \frac{\sigma_\varepsilon^2}{T(1-\rho)^2}\left(1 - \frac{2\rho(1-\rho^T)}{T(1-\rho^2)} + \frac{2\rho^{T+2} - \rho^{2(T+1)} - \rho^2}{1-\rho^2}\right) \\
E[\bar{y}_{ij-1}\bar{y}_{-1}] &= E[\bar{y}_{-1}^2] = \frac{\sigma_\varepsilon^2}{N^2T(1-\rho)^2}\left(1 - \frac{2\rho(1-\rho^T)}{T(1-\rho^2)} + \frac{2\rho^{T+2} - \rho^{2(T+1)} - \rho^2}{1-\rho^2}\right)
\end{aligned}$$

The bias of this Within estimator for (1) is therefore the following

$$E\left[\hat{\rho}-\rho\right]=\frac{\left(\frac{1-N^2}{N^2}\right)\frac{\sigma_\varepsilon^2}{T}\frac{1-\rho^{t-1}}{1-\rho}+\left(\frac{1-N^2}{N^2}\right)\frac{\sigma_\varepsilon^2}{T}\frac{1-\rho^{T-t}}{1-\rho}+\left(\frac{N^2-1}{N^2}\right)\frac{\sigma_\varepsilon^2}{T^2}\cdot A^*}{\left(\frac{N^2-1}{N^2}\right)\cdot\sigma_\varepsilon^2\frac{1-\rho^{2t}}{1-\rho^2}-B^*+C^*}$$

where

$$A^*=\left(\frac{N^2-1}{N^2}\right)\frac{\sigma_\varepsilon^2}{T}\left(\frac{1}{1-\rho}-\frac{1}{T}\frac{1-\rho^T}{(1-\rho)^2}\right)$$

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

and

$$C^*=\left(\frac{N^2-1}{N^2}\right)\frac{\sigma_\varepsilon^2}{T(1-\rho)^2}\left(1-\frac{2\rho(1-\rho^T)}{T(1-\rho^2)}+\frac{2\rho^{T+2}-\rho^{2(T+1)}-\rho^2}{1-\rho^2}\right)$$

Now for the same model (1) transformation (4) leads to the following terms. For the numerator

$$y_{ijt}-\bar{y}_i-\bar{y}_j-\bar{y}_t+2\bar{y}=\rho\left(y_{ijt-1}-\bar{y}_{i-1}-\bar{y}_{j-1}-\bar{y}_{t-1}+2\bar{y}_{-1}\right)+\left(\varepsilon_{ijt}-\bar{\varepsilon}_i-\bar{\varepsilon}_j-\bar{\varepsilon}_t+2\bar{\varepsilon}\right)$$

which yields the following terms. For the numerator

$$\begin{aligned}
E[y_{ijt-1}\varepsilon_{ijt}]=E[y_{ijt-1}\bar{\varepsilon}_t]=E[\bar{y}_{t-1}\varepsilon_{ijt}]=E[\bar{y}_{t-1}\bar{\varepsilon}_t]&=0\\
E[y_{ijt-1}\bar{\varepsilon}_i]=E[y_{ijt-1}\bar{\varepsilon}_j]&=\frac{\sigma_\varepsilon^2}{NT}\frac{1-\rho^{t-1}}{1-\rho}\\
E[y_{ijt-1}\bar{\varepsilon}]=E[\bar{y}_{t-1}\bar{\varepsilon}_i]=E[\bar{y}_{t-1}\bar{\varepsilon}_j]=E[\bar{y}_{t-1}\bar{\varepsilon}]&=\frac{\sigma_\varepsilon^2}{N^2T}\frac{1-\rho^{t-1}}{1-\rho}\\
E[\bar{y}_{i-1}\varepsilon_{ijt}]=E[\bar{y}_{j-1}\varepsilon_{ijt}]&=\frac{\sigma_\varepsilon^2}{NT}\cdot\frac{1-\rho^{T-t}}{1-\rho}\\
E[\bar{y}_{i-1}\bar{\varepsilon}_t]=E[\bar{y}_{j-1}\bar{\varepsilon}_t]=E[\bar{y}_{-1}\varepsilon_{ijt}]=E[\bar{y}_{-1}\bar{\varepsilon}_t]&=\frac{\sigma_\varepsilon^2}{N^2T}\cdot\frac{1-\rho^{T-t}}{1-\rho}\\
E[\bar{y}_{i-1}\bar{\varepsilon}_i]=E[\bar{y}_{j-1}\bar{\varepsilon}_j]&=\frac{\sigma_\varepsilon^2}{NT}\cdot\left(\frac{1}{1-\rho}-\frac{1}{T}\cdot\frac{1-\rho^T}{(1-\rho)^2}\right)\\
E[\bar{y}_{i-1}\bar{\varepsilon}_j]=E[\bar{y}_{j-1}\bar{\varepsilon}_i]=E[\bar{y}_{i-1}\bar{\varepsilon}]=E[\bar{y}_{j-1}\bar{\varepsilon}]=E[\bar{y}_{-1}\bar{\varepsilon}_i]&=E[\bar{y}_{-1}\bar{\varepsilon}_j]=\\
E[\bar{y}_{-1}\bar{\varepsilon}]&=\frac{\sigma_\varepsilon^2}{N^2T}\cdot\left(\frac{1}{1-\rho}-\frac{1}{T}\cdot\frac{1-\rho^T}{(1-\rho)^2}\right)
\end{aligned}$$

and for the denominator

$$E[y_{ijt-1}^2] = \sigma_\varepsilon^2 \cdot \frac{1-\rho^{2t}}{1-\rho^2}$$

$$E[y_{ijt-1}\bar{y}_{t-1}] = E[\bar{y}_{t-1}^2] = \frac{\sigma_\varepsilon^2}{N^2} \cdot \frac{1-\rho^{2t}}{1-\rho^2}$$

$$E[y_{ijt-1}\bar{y}_{i-1}] = E[y_{ijt-1}\bar{y}_{j-1}] = \frac{\sigma_\varepsilon^2}{NT(1-\rho^2)}\left(\frac{1-\rho^t}{1-\rho} + \rho\frac{1-\rho^{T-t}}{1-\rho} - \rho^{t+1}\cdot\frac{1+\rho^T}{1-\rho}\right)$$

$$E[y_{ijt-1}\bar{y}_{-1}] = E[\bar{y}_{i-1}\bar{y}_{t-1}] = E[\bar{y}_{j-1}\bar{y}_{t-1}] =$$

$$E[\bar{y}_{t-1}\bar{y}_{-1}] = \frac{\sigma_\varepsilon^2}{N^2T(1-\rho^2)}\left(\frac{1-\rho^t}{1-\rho} + \rho\frac{1-\rho^{T-t}}{1-\rho} - \rho^{t+1}\cdot\frac{1+\rho^T}{1-\rho}\right)$$

$$E[\bar{y}_{i-1}^2] = E[\bar{y}_{j-1}^2] = \frac{\sigma_\varepsilon^2}{NT(1-\rho)^2}\left(1 - \frac{2\rho(1-\rho^T)}{T(1-\rho^2)} + \frac{2\rho^{T+2}-\rho^{2(T+1)}-\rho^2}{1-\rho^2}\right)$$

$$E[\bar{y}_{i-1}\bar{y}_{-1}] = E[\bar{y}_{j-1}\bar{y}_{-1}] = E[\bar{y}_{-1}^2] =$$

$$\frac{\sigma_\varepsilon^2}{N^2T(1-\rho)^2}\left(1 - \frac{2\rho(1-\rho^T)}{T(1-\rho^2)} + \frac{2\rho^{T+2}-\rho^{2(T+1)}-\rho^2}{1-\rho^2}\right)$$

Taking into account the sign and the frequency of the above elements the bias of this Within estimator is

$$E\left[\hat{\rho}-\rho\right] = \frac{\left(\frac{2-2N}{N^2}\right)\cdot\frac{\sigma_\varepsilon^2}{T}\cdot\frac{1-\rho^{t-1}}{1-\rho} + \left(\frac{2-2N}{N^2}\right)\cdot\frac{\sigma_\varepsilon^2}{T}\cdot\frac{1-\rho^{T-t}}{1-\rho} + A^{**}}{\left(\frac{N^2-1}{N^2}\right)\cdot\sigma_\varepsilon^2\frac{1-\rho^{2t}}{1-\rho^2} + B^{**} + C^{**}}$$

where

$$A^{**} = \left(\frac{2N-2}{N^2}\right)\cdot\frac{\sigma_\varepsilon^2}{T}\cdot\left(\frac{1}{1-\rho} - \frac{1}{T}\cdot\frac{1-\rho^T}{(1-\rho)^2}\right)$$

$$B^{**} = \left(\frac{4-4N}{N^2}\right)\cdot\frac{\sigma_\varepsilon^2}{T(1-\rho^2)}\left(\frac{1-\rho^t}{1-\rho} + \rho\frac{1-\rho^{T-t}}{1-\rho} - \rho^{t+1}\cdot\frac{1+\rho^T}{1-\rho}\right)$$

and

$$\left(\frac{2N-4}{N^2}\right)\frac{\sigma_\varepsilon^2}{T(1-\rho)^2}\left(1 - \frac{2\rho(1-\rho^T)}{T(1-\rho^2)} + \frac{2\rho^{T+2}-\rho^{2(T+1)}-\rho^2}{1-\rho^2}\right)$$

Model (13)

Finally, let us turn to model (13)

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

The Within transformation gives

$$(y_{ijt} - \bar{y}_{ij} - \bar{y}_{jt} - \bar{y}_{it} + \bar{y}_t + \bar{y}_j + \bar{y}_i - \bar{y}),$$

so we get

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

The expected value of the components in the numerator are the following

$$
\begin{aligned}
&E[y_{ijt-1}\varepsilon_{ijt}] = E[y_{ijt-1}\bar{\varepsilon}_{it}] = E[y_{ijt-1}\bar{\varepsilon}_{jt}] = E[y_{ijt-1}\bar{\varepsilon}_t] = E[\bar{y}_{it-1}\varepsilon_{ijt}] = E[\bar{y}_{jt-1}\varepsilon_{ijt}] = \\
&\quad E[\bar{y}_{it-1}\bar{\varepsilon}_{it}] = E[\bar{y}_{jt-1}\bar{\varepsilon}_{jt}] = E[\bar{y}_{it-1}\bar{\varepsilon}_{jt}] = E[\bar{y}_{jt-1}\bar{\varepsilon}_{it}] = E[\bar{y}_{it-1}\bar{\varepsilon}_t] = E[\bar{y}_{jt-1}\bar{\varepsilon}_t] = \\
&\quad E[\bar{y}_{t-1}\varepsilon_{ijt}] = E[\bar{y}_{t-1}\bar{\varepsilon}_{jt}] = E[\bar{y}_{t-1}\bar{\varepsilon}_{it}] = E[\bar{y}_{t-1}\bar{\varepsilon}_t] = 0 \\
&E[y_{ijt-1}\bar{\varepsilon}_{ij}] = \frac{\sigma_\varepsilon^2}{T} \cdot \frac{1-\rho^{t-1}}{1-\rho} \\
&\quad E[y_{ijt-1}\bar{\varepsilon}_i] = E[y_{ijt-1}\bar{\varepsilon}_j] = E[\bar{y}_{it-1}\bar{\varepsilon}_{ij}] = E[\bar{y}_{jt-1}\bar{\varepsilon}_{ij}] = E[\bar{y}_{it-1}\bar{\varepsilon}_i] = E[\bar{y}_{jt-1}\bar{\varepsilon}_j] = \\
&\qquad \frac{\sigma_\varepsilon^2}{NT} \cdot \frac{1-\rho^{t-1}}{1-\rho} \\
&\quad E[y_{ijt-1}\bar{\varepsilon}] = E[\bar{y}_{it-1}\bar{\varepsilon}_j] = E[\bar{y}_{jt-1}\bar{\varepsilon}_i] = E[\bar{y}_{it-1}\bar{\varepsilon}_j] = E[\bar{y}_{jt-1}\bar{\varepsilon}_i] = E[\bar{y}_{it-1}\bar{\varepsilon}] = \\
&\quad E[\bar{y}_{jt-1}\bar{\varepsilon}] = E[\bar{y}_{t-1}\bar{\varepsilon}_{ij}] = E[\bar{y}_{t-1}\bar{\varepsilon}_i] = E[\bar{y}_{t-1}\bar{\varepsilon}_j] = E[\bar{y}_{t-1}\bar{\varepsilon}] = \frac{\sigma_\varepsilon^2}{N^2T} \cdot \frac{1-\rho^{t-1}}{1-\rho} \\
&E[\bar{y}_{ij-1}\varepsilon_{ijt}] = \frac{\sigma_\varepsilon^2}{T} \cdot \frac{1-\rho^{T-t}}{1-\rho}
\end{aligned}
$$

$$
\begin{aligned}
&E[\bar{y}_{ij-1}\bar{\varepsilon}_{jt}] = E[\bar{y}_{ij-1}\bar{\varepsilon}_{it}] = E[\bar{y}_{i-1}\varepsilon_{ijt}] = E[\bar{y}_{j-1}\varepsilon_{ijt}] = E[\bar{y}_{i-1}\bar{\varepsilon}_{it}] = \\
&\quad E[\bar{y}_{j-1}\bar{\varepsilon}_{jt}] = \frac{\sigma_\varepsilon^2}{NT} \cdot \frac{1-\rho^{T-t}}{1-\rho} \\
&E[\bar{y}_{ij-1}\bar{\varepsilon}_t] = E[\bar{y}_{i-1}\bar{\varepsilon}_{jt}] = E[\bar{y}_{j-1}\bar{\varepsilon}_{it}] = E[\bar{y}_{i-1}\bar{\varepsilon}_t] = E[\bar{y}_{j-1}\bar{\varepsilon}_t] = \\
&E[\bar{y}_{-1}\varepsilon_{ijt}] = E[\bar{y}_{-1}\bar{\varepsilon}_{jt}] = E[\bar{y}_{-1}\bar{\varepsilon}_{it}] = E[\bar{y}_{-1}\bar{\varepsilon}_t] = \frac{\sigma_\varepsilon^2}{N^2T} \cdot \frac{1-\rho^{T-t}}{1-\rho} \\
&E[\bar{y}_{ij-1}\bar{\varepsilon}_{ij}] = \frac{\sigma_\varepsilon^2}{T} \cdot \left(\frac{1}{1-\rho} - \frac{1}{T} \cdot \frac{1-\rho^T}{(1-\rho)^2}\right) \\
&E[\bar{y}_{ij-1}\bar{\varepsilon}_j] = E[\bar{y}_{ij-1}\bar{\varepsilon}_i] = E[\bar{y}_{i-1}\bar{\varepsilon}_{ij}] = E[\bar{y}_{j-1}\bar{\varepsilon}_{ij}] = \\
&\quad E[\bar{y}_{i-1}\bar{\varepsilon}_i] = E[\bar{y}_{j-1}\bar{\varepsilon}_j] = \frac{\sigma_\varepsilon^2}{NT} \cdot \left(\frac{1}{1-\rho} - \frac{1}{T} \cdot \frac{1-\rho^T}{(1-\rho)^2}\right) \\
&E[\bar{y}_{ij-1}\bar{\varepsilon}] = E[\bar{y}_{i-1}\bar{\varepsilon}_j] = E[\bar{y}_{j-1}\bar{\varepsilon}_i] = E[\bar{y}_{i-1}\bar{\varepsilon}] = E[\bar{y}_{j-1}\bar{\varepsilon}] = E[\bar{y}_{-1}\bar{\varepsilon}_{ij}] = \\
&\quad E[\bar{y}_{-1}\bar{\varepsilon}_i] = E[\bar{y}_{-1}\bar{\varepsilon}_j] = E[\bar{y}_{-1}\bar{\varepsilon}] = \frac{\sigma_\varepsilon^2}{N^2T} \cdot \left(\frac{1}{1-\rho} - \frac{1}{T} \cdot \frac{1-\rho^T}{(1-\rho)^2}\right)
\end{aligned}
$$

And in the denominator

$$E[y_{ijt-1}^2] = \sigma_\varepsilon^2 \cdot \frac{1-\rho^{2t}}{1-\rho^2}$$

$$E[y_{ijt-1}\bar{y}_{it-1}] = E[y_{ijt-1}\bar{y}_{jt-1}] = E[\bar{y}_{it-1}^2] = E[\bar{y}_{jt-1}^2] = \frac{\sigma_\varepsilon^2}{N} \cdot \frac{1-\rho^{2t}}{1-\rho^2}$$

$$E[y_{ijt-1}\bar{y}_{t-1}] = E[\bar{y}_{it-1}\bar{y}_{jt-1}] = E[\bar{y}_{it-1}\bar{y}_{t-1}] = E[\bar{y}_{jt-1}\bar{y}_{t-1}] = E[\bar{y}_{t-1}^2] = \frac{\sigma_\varepsilon^2}{N^2} \cdot \frac{1-\rho^{2t}}{1-\rho^2}$$

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

$$E[y_{ijt-1}\bar{y}_{i-1}] = E[y_{ijt-1}\bar{y}_{j-1}] = E[\bar{y}_{ij-1}\bar{y}_{it-1}] = E[\bar{y}_{ij-1}\bar{y}_{jt-1}] =$$

$$E[\bar{y}_{it-1}\bar{y}_{i-1}] = E[\bar{y}_{jt-1}\bar{y}_{j-1}] = \frac{\sigma_\varepsilon^2}{NT(1-\rho^2)} \left(\frac{1-\rho^t}{1-\rho} + \rho \frac{1-\rho^{T-t}}{1-\rho} - \rho^{t+1} \cdot \frac{1+\rho^T}{1-\rho} \right)$$

$$E[y_{ijt-1}\bar{y}_{-1}] = E[\bar{y}_{ij-1}\bar{y}_{t-1}] = E[\bar{y}_{it-1}\bar{y}_{j-1}] = E[\bar{y}_{jt-1}\bar{y}_{i-1}] = E[\bar{y}_{it-1}\bar{y}_{-1}] = E[\bar{y}_{jt-1}\bar{y}_{-1}] =$$

$$E[\bar{y}_{t-1}\bar{y}_{i-1}] = E[\bar{y}_{t-1}\bar{y}_{j-1}] = E[\bar{y}_{t-1}\bar{y}_{-1}] = \frac{\sigma_\varepsilon^2}{N^2T(1-\rho^2)} \left(\frac{1-\rho^t}{1-\rho} + \rho \frac{1-\rho^{T-t}}{1-\rho} - \rho^{t+1} \cdot \frac{1+\rho^T}{1-\rho} \right)$$

$$E[\bar{y}_{ij-1}^2] = \frac{\sigma_\varepsilon^2}{T(1-\rho)^2} \left(1 - \frac{2\rho(1-\rho^T)}{T(1-\rho^2)} + \frac{2\rho^{T+2} - \rho^{2(T+1)} - \rho^2}{1-\rho^2} \right)$$

$$E[\bar{y}_{ij-1}\bar{y}_{i-1}] = E[\bar{y}_{ij-1}\bar{y}_{j-1}] = E[\bar{y}_{i-1}^2] = E[\bar{y}_{j-1}^2] = \frac{\sigma_\varepsilon^2}{NT(1-\rho)^2} \left(1 - \frac{2\rho(1-\rho^T)}{T(1-\rho^2)} + \frac{2\rho^{T+2} - \rho^{2(T+1)} - \rho^2}{1-\rho^2} \right)$$

$$E[\bar{y}_{ij-1}\bar{y}_{-1}] = E[\bar{y}_{i-1}\bar{y}_{j-1}] = E[\bar{y}_{i-1}\bar{y}_{-1}] = E[\bar{y}_{j-1}\bar{y}_{-1}] = E[\bar{y}_{-1}^2] = \frac{\sigma_\varepsilon^2}{N^2T(1-\rho^2)} \left(\frac{1-\rho^t}{1-\rho} + \rho \frac{1-\rho^{T-t}}{1-\rho} - \rho^{t+1} \cdot \frac{1+\rho^T}{1-\rho} \right)$$

To sum up the bias we get for this model is

$$E[\widehat{\rho} - \rho] = \frac{\left(\frac{-(N-1)^2}{N^2}\right) \cdot \frac{\sigma_\varepsilon^2}{T} \cdot \frac{1-\rho^{t-1}}{1-\rho} + \left(\frac{-(N-1)^2}{N^2}\right) \cdot \frac{\sigma_\varepsilon^2}{T} \cdot \frac{1-\rho^{T-t}}{1-\rho} + A^*}{\left(\frac{(N-1)^2}{N^2}\right) \sigma_\varepsilon^2 \frac{1-\rho^{2t}}{1-\rho^2} + B^* + C^*}$$

where

$$A^* = \frac{(N-1)^2}{N^2} \cdot \frac{\sigma_\varepsilon^2}{T} \left(\frac{1}{1-\rho} - \frac{1}{T} \cdot \frac{1-\rho^T}{(1-\rho)^2} \right)$$

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

and

$$C^* = \left(\frac{(N-1)^2}{N^2} \right) \frac{\sigma_\varepsilon^2}{T(1-\rho)^2} \left(1 - \frac{2\rho(1-\rho^T)}{T(1-\rho^2)} + \frac{2\rho^{T+2} - \rho^{2(T+1)} - \rho^2}{1-\rho^2} \right)$$

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)**			**(4)**	**(6)**		**(7)**	**(8)**	**(9)**		**(11)**			**4-D (25)**		
Transformation		Opt (3)	(2)	(10)	Opt	Opt (2)	(16)	Opt	Opt	Opt (10)	(14)	Opt (12)	(15)	(17)	Opt (26)	(27)	(28)
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)		(4)	(6)	(7)	(8)	(9)	(11)
Transformation	(2)	(3)	(5)	(2)				(12)
Finite N, T	-	-	-	-	+	+	+	-
Finite T, $N \to \infty$	-	+	-	-	+	+	-	-

Where: + stands for no bias.