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.

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[1] 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}, $i = 1, \ldots, N_1$, $j = 1, \ldots, N_2$, and $t = 1, \ldots, T$. As in economic flows such as trade, capital (FDI), etc., there is some kind of reciprocity, we assume to start with, that $N_1 = N_2 = N$. Implicitly we also assume that the set of individuals in the observation sets i and j are the same, then we relax this assumption later on. The main question is how to formalize the individual and time heterogeneity, in our case the fixed effects. Different forms of heterogeneity yield naturally different models. In theory any fixed effects three-dimensional panel data model can directly be estimated, say for example, by least squares (LS). This involves the explicit incorporation in the model of the fixed effects through dummy variables (see for example formulation (13) later on). The resulting estimator is usually called Least Squares Dummy Variable (LSDV) estimator. However, it is well known that the

[1] 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.[2] 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 theorem, there is always an optimal Within estimator, excatly 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

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

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.

The simplest Within transformation for this model is

$$(y_{ijt} - \bar{y}_{ij} - \bar{y}_t + \bar{y}) \tag{2}$$

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

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

$$(y_{ijt} - \bar{y}_i - \bar{y}_j - \bar{y}_t + 2\bar{y}) \tag{3}$$

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

$$y_{ijt} = \beta' x_{ijt} + \gamma_{ij} + \varepsilon_{ijt} \tag{4}$$

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

$$(y_{ijt} - \bar{y}_{ij}) \quad \text{where} \quad \bar{y}_{ij} = 1/T \sum_t y_{ijt} \tag{5}$$

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 (4) often used in empirical studies is

$$y_{ijt} = \beta' x_{ijt} + \gamma_{ij} + \lambda_t + \varepsilon_{ijt} \tag{6}$$

As model (1) is in fact a special case of this model (6), transformation (2) can be used to clear the fixed effects. While transformation (2) leads to the optimal Within estimator for model (6), 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 (4) and (6) 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

$$y_{ijt} = \beta' x_{ijt} + \alpha_{jt} + \varepsilon_{ijt} \tag{7}$$

The Within transformation which clears the fixed effects is

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

Another variant of this model is

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

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

$$y_{ijt} = \beta' x_{ijt} + \alpha_{it} + \alpha_{jt} + \varepsilon_{ijt} \tag{9}$$

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

$$(y_{ijt} - \bar{y}_{jt} - \bar{y}_{it} + \bar{y}_t) \tag{10}$$

Let us notice here that transformation (10) 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

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

By applying suitable restrictions to model (11) we can obtain the models discussed above. The Within transformation for this model is

$$\begin{aligned}(y_{ijt} - 1/T\sum_t y_{ijt} - 1/N\sum_i y_{ijt} - 1/N\sum_j y_{ijt} + 1/N^2\sum_i\sum_j y_{ijt} \\ + 1/(NT)\sum_i\sum_t y_{ijt} + 1/(NT)\sum_j\sum_t y_{ijt} - 1/(N^2T)\sum_i\sum_j\sum_t y_{ijt})\end{aligned} \tag{12}$$

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 (11) in matrix form is

$$y = X\beta + D_1\gamma + D_2\alpha + D_3\alpha_* + \varepsilon \tag{13}$$

where y, $(N^2\times 1)$ 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,

$$D_1 = I_{N^2}\otimes l_t, \quad D_2 = I_N\otimes l_N\otimes I_T \quad \text{and} \quad D_3 = l_N\otimes I_{NT}$$

l is the vector of ones and I is the identity matrix with the appropriate size in the index. Let $D = (D_1, D_2, D_3)$, $Q_D = D(D'D)^{-1}D'$ and $P_D = I - Q_D$. Using *Davis'* [2002] method it can be shown that $P_D = P_1 - Q_2 - Q_3$ where

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

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

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

The typical element of P_D gives the transformation (12). By appropriate restrictions on the parameters of (13) we get back the previously analysed Within transformations. Now transforming model (13) with transformation (12) leads to

$$\underbrace{P_Dy}_{y_p} = \underbrace{P_DX}_{X_p}\beta + \underbrace{P_DD_1}_{=0}\gamma + \underbrace{P_DD_2}_{=0}\alpha + \underbrace{P_DD_3}_{=0}\alpha_* + \underbrace{P_D\varepsilon}_{\varepsilon_p}$$

and the corresponding (optimal) Within estimator is

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

3. Some Data Problems

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 (3)), (9) and (11). Let us have a closer look at the difficulty. For model (1) and transformation (3), instead of canceled out fixed effects, we end up with the following remaining fixed effects

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

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

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

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

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

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

and for the time effects

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

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

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

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 (2) 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 (9). After the Within transformation (10), instead of canceled out fixed effects we end up with the following remaining fixed effects

$$\begin{aligned}
\alpha_{it}^* &= \alpha_{it} - \frac{1}{N-1}\sum_{i=1;i\neq j}^{N} \alpha_{it} - \frac{1}{N-1}(N-1)\alpha_{it} + \frac{1}{N(N-1)}\sum_{i=1}^{N}(N-1)\alpha_{it} \\
&= -\frac{1}{N(N-1)}\sum_{k=1;k\neq j}^{N} \alpha_{kt} + \frac{1}{N}\alpha_{jt}
\end{aligned}$$

and

$$\begin{aligned}
\gamma_{jt}^* &= \gamma_{jt} - \frac{1}{N-1}(N-1)\gamma_{jt} - \frac{1}{N-1}\sum_{j=1;j\neq i}^{N} \gamma_{jt} + \frac{1}{N(N-1)}\sum_{j=1}^{N}(N-1)\gamma_{jt} \\
&= -\frac{1}{N(N-1)}\sum_{l=1;l\neq i}^{N} \gamma_{lt} + \frac{1}{N}\gamma_{it}
\end{aligned}$$

As long as the α^* and γ^* parameters are not zero, the Within estimators will be biased. Similarly for model (11), the remaining fixed effects are now

$$\begin{aligned}
\gamma_{ij}^* = \gamma_{ij} &- \frac{1}{T}T\cdot\gamma_{ij} - \frac{1}{N-1}\sum_{i=1;i\neq j}^{N}\gamma_{ij} - \frac{1}{N-1}\sum_{j=1;j\neq i}^{N}\gamma_{ij} \\
&+ \frac{1}{N(N-1)}\sum_{i=1}^{N}\sum_{j=1;j\neq i}^{N}\gamma_{ij} + \frac{1}{(N-1)T}\sum_{i=1;i\neq j}^{N}T\gamma_{ij} \\
&+ \frac{1}{(N-1)T}\sum_{j=1;j\neq i}^{N}T\gamma_{ij} - \frac{1}{N(N-1)T}\sum_{i=1}^{N}\sum_{j=1;j\neq i}^{N}T\gamma_{ij} = 0
\end{aligned}$$

but

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

and, finally

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

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.

We can go further along the above lines and see what going to happen if the observation sets i and j are different. If the two set 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.

Like in the case of the usual panel data sets, 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 $t = 1, \dots, T_{ij}$, $\sum_i \sum_j T_{ij} = T$ and T_{ij} is often not equal to $T_{i'j'}$. For models (4), (7), (8) and (9) 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 (11) we are facing trouble.

In the case of model (1) and transformation (2) we get for the fixed effects the following terms (let us remember: this in fact is the optimal transformation for model (6))

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

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

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 (2) is the optimal Within estimator for model (6), this is bad news for the estimation of that model as well. We, unfortunately, get very similar results for transformation (3) too. The good news is, on the other hand, as seen earlier, that for model (1) transformation (10) 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 (11) and transformation (12). The remaining fixed effects are now

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

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

and

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

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 model (11) is quite problematic. The direct estimation of the model by LSDV is not feasible, even for moderate N, as the number of dummies (fixed effects) is becoming very large quite quickly. On the other hand, the estimation of the model with the Within estimator (12) is unbiased only when there are no data problems such as no-self-trade or unbalanced observations. A pragmatic way out of this trap is to follow a two-step procedure. First, transform model (11), (13) (including all dummy variables) with the Within transformation (5) (D_1 will drop out), then estimate this transformed model with OLS (including the transformed D_2 and D_3). Using this procedure we need to use substantially less dummy variables, and it can be shown (after some algebra) that the estimator remains unbiased even in the case of the above data problems.

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.

In order to show the problem, let us start with the simple linear dynamic model with bilateral interaction effects, that is model (4)

$$y_{ijt} = \rho y_{ijt-1} + \gamma_{ij} + \varepsilon_{ijt} \tag{14}$$

With backward substitution we get

$$y_{ijt} = \rho^t y_{ij0} + \frac{1-\rho^t}{1-\rho}\gamma_{ij} + \sum_{k=0}^{t} \rho^k \varepsilon_{ijt-k} \tag{15}$$

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

$$y_{ijt} = \rho y_{ijt-1} + x'_{ijt}\beta + \gamma_{ij} + \varepsilon_{ijt} \tag{16}$$

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 (14) introduced above we get

$$(y_{ijt} - \bar{y}_{trans}) = \rho(y_{ijt-1} - \bar{y}_{trans-1}) + (\varepsilon_{ijt} - \bar{\varepsilon}_{trans}) \tag{17}$$

Using OLS to estimate ρ, we get

$$\widehat{\rho}_t = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N}(y_{ijt-1} - \bar{y}_{trans-1})\cdot(y_{ijt} - \bar{y}_{trans})}{\sum_{i=1}^{N}\sum_{j=1}^{N}(y_{ijt-1} - \bar{y}_{trans-1})^2} \tag{18}$$

So the bias is

$$\begin{aligned}
E\left[\hat{\rho}_t\right] &= E\left[\frac{\sum_{i=1}^{N}\sum_{j=1}^{N}(y_{ijt-1} - \bar{y}_{trans-1})\cdot(\rho(y_{ijt-1} - \bar{y}_{trans-1}) + (\varepsilon_{ijt} - \bar{\varepsilon}_{trans}))}{\sum_{i=1}^{N}\sum_{j=1}^{N}(y_{ijt-1} - \bar{y}_{trans-1})^2}\right] = \\
&= E\left[\frac{\rho\cdot\sum_{i=1}^{N}\sum_{j=1}^{N}(y_{ijt-1} - \bar{y}_{trans-1})^2}{\sum_{i=1}^{N}\sum_{j=1}^{N}(y_{ijt-1} - \bar{y}_{trans-1})^2} + \frac{\sum_{i=1}^{N}\sum_{j=1}^{N}(y_{ijt-1} - \bar{y}_{trans-1})(\varepsilon_{ijt} - \bar{\varepsilon}_{trans})}{\sum_{i=1}^{N}\sum_{j=1}^{N}(y_{ijt-1} - \bar{y}_{trans-1})^2}\right] \\
&= \rho + E\left[\frac{\sum_{i=1}^{N}\sum_{j=1}^{N}(y_{ijt-1} - \bar{y}_{trans-1})(\varepsilon_{ijt} - \bar{\varepsilon}_{trans})}{\sum_{i=1}^{N}\sum_{j=1}^{N}(y_{ijt-1} - \bar{y}_{trans-1})^2}\right] = \rho + \frac{A_t}{B_t}
\end{aligned} \tag{19}$$

Continuing with model (14) and using now the appropriate (5) 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 A_t 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 B_t

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

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

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

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

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

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 (2) 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

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

where

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

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

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 (14).

As seen earlier, the optimal Within transformation for model (2) is in fact (3)

$$(y_{ijt} - \bar{y}_i - \bar{y}_j - \bar{y}_t + 2\bar{y})$$

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

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

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).

Let us now continue with models (7) and (8) which can be considered as the same models from this point of view

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

With the Within transformation we get

$$y_{ijt} - \bar{y}_{jt} = \rho \cdot (y_{ijt-1} - \bar{y}_{jt-1}) + (\alpha_{jt} - \underbrace{\frac{1}{N} \cdot \sum_{i=1}^{N} \alpha_{jt}}_{\frac{1}{N} N \alpha_{jt}}) + (\varepsilon_{ijt} - \bar{\varepsilon}_{jt}),$$

where

$$\bar{y}_{jt} = \frac{1}{N} \cdot \sum_{i=1}^{N} y_{ijt} \quad \bar{y}_{jt-1} = \frac{1}{N} \cdot \sum_{i=1}^{N} y_{ijt-1} \quad \bar{\varepsilon}_{jt} = \frac{1}{N} \cdot \sum_{i=1}^{N} \varepsilon_{ijt}.$$

Following the derivation presented in details in the Appendix the bias for Model (7) is in fact zero, so this Within estimator is unbiased.

Let us carry on with model (9). Using the Within transformation we get

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

The finite sample bias now is (see the Appendix for details), as above, zero, so again, this Within estimator is unbiased.

And finally, let us turn to model (11)

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

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

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

And for the finite sample bias of this model we get

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

where

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

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

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 (4).

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 we had data not only for a given country, but for several, 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 (11), (13) with pair-wise interaction effects:

$$y_{ijst} = x'_{ijst}\beta + \gamma^0_{ijs} + \gamma^1_{ijt} + \gamma^2_{jst} + \gamma^3_{ist} + \varepsilon_{ijst} \tag{20}$$

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

$$\gamma^0_{IS} + \sum_{k=1}^{M} \gamma^M_{i_k,t} \tag{21}$$

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

The Within transformation for model (20) is

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

or in matrix form

$$
\begin{aligned}
P_D = & \left(I_{N_i} - \bar{J}_{N_i}\right) \otimes \left(I_{N_j} - \bar{J}_{N_j}\right) \otimes \left(I_{N_s} - \bar{J}_{N_s}\right) \otimes \left(I_T - \bar{J}_T\right) = \\
& I_{N_i N_j N_s T} - \left(\bar{J}_{N_i} \otimes I_{N_j N_s T}\right) - \left(I_{N_i} \otimes \bar{J}_{N_j} \otimes I_{N_s T}\right) - \left(I_{N_i N_j} \otimes \bar{J}_{N_s} \otimes I_T\right) \\
& - \left(I_{N_i N_j N_s} \otimes \bar{J}_T\right) + \left(\bar{J}_{N_i N_j} \otimes I_{N_s T}\right) + \left(\bar{J}_{N_i} \otimes I_{N_j} \otimes \bar{J}_{N_s} \otimes I_T\right) \\
& + \left(\bar{J}_{N_i} \otimes I_{N_j N_s} \otimes \bar{J}_T\right) + \left(I_{N_i} \otimes \bar{J}_{N_j N_s} \otimes I_T\right) + \left(I_{N_i} \otimes \bar{J}_{N_j} \otimes I_{N_s} \otimes \bar{J}_T\right) \\
& + \left(I_{N_i N_j} \otimes \bar{J}_{N_s T}\right) - \left(\bar{J}_{N_i N_j N_s} \otimes I_T\right) - \left(\bar{J}_{N_i N_j} \otimes I_{N_s} \otimes \bar{J}_T\right) - \left(\bar{J}_{N_i} \otimes I_{N_j} \otimes \bar{J}_{N_s T}\right) \\
& - \left(I_{N_j} \otimes \bar{J}_{N_j N_s T}\right) + \bar{J}_{N_i N_j N_s T}
\end{aligned} \tag{22}
$$

It can be shown easily, that the properties of the Within estimator based on transformation (22) in the case of no-self-trade, unbalanced data and dynamic models are exactly the same as seen earlier for the three dimensional model.

The generalization of this Within estimator for any higher dimensions can be done using the general form (21). 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 $s_1, \ldots s_M$ (in the three-dimensional case $s_1 = i$, $s_2 = j$ and $s_3 = t$). The Within transformation then is

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

where

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

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

6. 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.

Appendix

Finite sample bias derivations for the dynamic model.

Model (1)

In this case the Within transformation (2) leads to

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

Components of the numerator of the bias are

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

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

$$E[y_{ijt-1}\bar{\varepsilon}_t] = 0$$

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

$$E[\bar{y}_{ij-1}\varepsilon_{ijt}] = \frac{\sigma_\varepsilon^2}{T}\frac{1-\rho^{T-t}}{1-\rho}$$

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

$$E[\bar{y}_{ij-1}\bar{\varepsilon}_t] = \frac{\sigma_\varepsilon^2}{N^2T}\frac{1-\rho^{T-t}}{1-\rho}$$

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

$$E[\bar{y}_{t-1}\varepsilon_{ijt}] = 0$$

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

$$E[\bar{y}_{t-1}\bar{\varepsilon}_t] = 0$$

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

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

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

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

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

Considering the signs of the components, we get the following expected value for the numerator

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

Components of the denominator are

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

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

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

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

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

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

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

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

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

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

Thus the expected value of the denominator is

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

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

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

where

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

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

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 (3) leads to the following terms. For the numerator:

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

$$E[y_{ijt-1}\bar{\varepsilon}_i]=E[y_{ijt-1}\bar{\varepsilon}_j]=\frac{\sigma_\varepsilon^2}{NT}\frac{1-\rho^{t-1}}{1-\rho}$$

$$E[y_{ijt-1}\bar{\varepsilon}_t]=0$$

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

$$E[\bar{y}_{i-1}\varepsilon_{ijt}]=E[\bar{y}_{j-1}\varepsilon_{ijt}]=\frac{\sigma_\varepsilon^2}{NT}\cdot\frac{1-\rho^{T-t}}{1-\rho}$$

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

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

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

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

$$E[\bar{y}_{t-1}\varepsilon_{ijt}] = 0$$

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

$$E[\bar{y}_{t-1}\bar{\varepsilon}_t] = 0$$

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

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

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

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

$$E[2\bar{y}_{-1}2\bar{\varepsilon}] = \frac{4\sigma_\varepsilon^2}{N^2T} \cdot \left(\frac{1}{1-\rho} - \frac{1}{T} \cdot \frac{1-\rho^T}{(1-\rho)^2}\right)$$

And for the denominator

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

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

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

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

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

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

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

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

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

$$E[4\bar{y}_{-1}^2] = \frac{4\sigma_\varepsilon^2}{N^2 T(1-\rho)^2} \left(1 - \frac{2\rho(1-\rho^T)}{T(1-\rho^2)} + \frac{2\rho^{T+2} - \rho^{2(T+1)} - \rho^2}{1-\rho^2} \right)$$

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

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

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

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

Models (7) and (8)

Let us continue with models (7) and (8) which can be considered as the same models from this point of view

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

With the Within transformation we get

$$y_{ijt} - \bar{y}_{jt} = \rho \cdot (y_{ijt-1} - \bar{y}_{jt-1}) + (\alpha_{jt} - \underbrace{\frac{1}{N} \cdot \sum_{i=1}^{N} \alpha_{jt}}_{\frac{1}{N} N \alpha_{jt}}) + (\varepsilon_{ijt} - \bar{\varepsilon}_{jt}),$$

where

$$\bar{y}_{jt} = \frac{1}{N} \cdot \sum_{i=1}^{N} y_{ijt} \quad \bar{y}_{jt-1} = \frac{1}{N} \cdot \sum_{i=1}^{N} y_{ijt-1} \quad \bar{\varepsilon}_{jt} = \frac{1}{N} \cdot \sum_{i=1}^{N} \varepsilon_{ijt}.$$

The components of the bias are the following

$$E[y_{ijt-1}\varepsilon_{ijt}] = 0 \quad \text{since they are uncorrelated}$$

$$E[\bar{y}_{jt-1}\varepsilon_{ijt}] = E\left[\left(\frac{1}{N}\cdot\sum_{i=1}^{N}\sum_{k=0}^{t-1}\rho^k\varepsilon_{ijt-k}\right)\varepsilon_{ijt}\right] = 0$$

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

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

The elements in the denominator are

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

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

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

So the bias for Model (7) is nil as the nominator of the bias is zero, and the denominator finite.

Model (9)
Using the Within transformation we get

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

As in the numerator of the bias all elements are zero, while the denominator is finite, this Within estimator is obviously unbiased.

Model (11)
And finally, let us turn to model (11)

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

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

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

The expected value of the components are the following. For the numerator:

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

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

$$E[y_{ijt-1}\bar{\varepsilon}_{it}] = 0$$

$$E[y_{ijt-1}\bar{\varepsilon}_{jt}] = 0$$

$$E[y_{ijt-1}\bar{\varepsilon}_{t}] = 0$$

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

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

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

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

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

$$E[\bar{y}_{ij-1}\bar{\varepsilon}_{jt}] = \frac{\sigma_\varepsilon^2}{NT} \cdot \frac{1-\rho^{T-t}}{1-\rho}$$

$$E[\bar{y}_{ij-1}\bar{\varepsilon}_{it}] = \frac{\sigma_\varepsilon^2}{NT} \cdot \frac{1-\rho^{T-t}}{1-\rho}$$

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

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

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

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

$$E[\bar{y}_{it-1}\varepsilon_{ijt}] = E[\bar{y}_{jt-1}\varepsilon_{ijt}] = 0$$

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

$$E[\bar{y}_{it-1}\bar{\varepsilon}_{it}] = E[\bar{y}_{jt-1}\bar{\varepsilon}_{jt}] = 0$$

$$E[\bar{y}_{it-1}\bar{\varepsilon}_{jt}] = E[\bar{y}_{jt-1}\bar{\varepsilon}_{it}] = 0$$

$$E[\bar{y}_{it-1}\bar{\varepsilon}_t] = E[\bar{y}_{jt-1}\bar{\varepsilon}_t] = 0$$

$$E[\bar{y}_{it-1}\bar{\varepsilon}_i] = E[\bar{y}_{jt-1}\bar{\varepsilon}_j] = \frac{\sigma_\varepsilon^2}{NT} \cdot \frac{1-\rho^{t-1}}{1-\rho}$$

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

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

$$E[\bar{y}_{t-1}\varepsilon_{ijt}] = 0$$

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

$$E[\bar{y}_{t-1}\bar{\varepsilon}_{jt}] = 0$$

$$E[\bar{y}_{t-1}\bar{\varepsilon}_{it}] = 0$$

$$E[\bar{y}_{t-1}\bar{\varepsilon}_t] = 0$$

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

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

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

$$E[\bar{y}_{i-1}\varepsilon_{ijt}] = E[\bar{y}_{j-1}\varepsilon_{ijt}] = \frac{\sigma_\varepsilon^2}{NT} \cdot \frac{1-\rho^{T-t}}{1-\rho}$$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

So the expected value of the numerator, considering the signs of the components is

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

The components of the denominator are

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Thus the expected value of the denominator after taking into account the signs of the components is

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

To sum up the bias we get for this model is

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

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

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, The Econometrics of Panel Data, 3rd edition, Sringer Verlag*, pp. 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, pp. 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.

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.