The Estimation of Multi-dimensional Fixed Effects Panel Data Models

Laszlo Balazsi
Laszlo Matyas
Tom Wansbeek

CESIFO WORKING PAPER NO. 5251
CATEGORY 12: EMPIRICAL AND THEORETICAL METHODS
MARCH 2015

ISSN 2364-1428

CESifo *Working Paper No. 5251*

The Estimation of Multi-dimensional Fixed Effects Panel Data Models

Abstract

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

JEL-Code: C100, C200, C400, F170, F470.

Keywords: panel data, unbalanced panel, dynamic panel data model, multidimensional panel data, fixed effects, trade models, gravity models, FDI.

Laszlo Balazsi
Central European University
Department of Economics
Hungary - Budapest 1051, Nador u. 9
lac.balazsi@gmail.com

*Laszlo Matyas**
Central European University
Department of Economics
Hungary - Budapest 1051, Nador u. 9
matyas@ceu.edu

Tom Wansbeek
Department of Economics, Econometrics
& Finance / University of Groningen
Nettelbosje 2
The Netherlands - 9747 AE Groningen
t.j.wansbeek@rug.nl

*corresponding author

March 04, 2015
Short running title: **Multi-dimensional Fixed Effects Models**.
Support by the Australian Research Council grant DP110103824 is kindly acknowledged. Comments received from the participants of the 18th and 19th Panel Data Conferences, Paris 2012 and London 2013 respectively, are also much appreciated. Last but not least, we would like to thank the editor and two anonymous reviewers for their very helpful comments that substantially improved the quality of the paper.

1. Introduction

Multidimensional panel data sets are becoming more readily available and are used to study a variety of phenomena like: 1) International trade and/or capital flows between countries or regions; 2) The trading volume across several products and stores over time (three panel dimensions) and; 3) The number of passengers between multiple airport hubs for different airlines (four panel dimensions). 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. These models are linear regression models differing in the specification of the fixed effects. As in the case of the familiar two-dimensional (2D) fixed effects panel data models, they can simply be estimated by Ordinary Least Squares (OLS). However, the large number of dummy variables can make this computationally difficult or even impossible which holds a fortiori when the dimensionality of the data is three or more. This problem is usually solved by the within transformation and thus invoking the Frisch-Waugh theorem [see, for example, Gourieroux and Monfort (1995) and Greene (2012)]. This states that in a linear regression model, using matrix notation, with a partitioned regressor set (X, D),

$$y = X\beta + D\pi + \varepsilon \tag{1}$$

the OLS estimator for β can be obtained by regressing $\tilde{y}$ on $\tilde{X}$, with $\tilde{y} = M_D y$, $\tilde{X} = M_D X$ and $M_D = I - D(D^{'}D)^{-}D^{'}$, the matrix that projects into the space orthogonal to D. In our case, D matrix contains the dummy variables corresponding to the fixed effects. When the data set is balanced, computing $\tilde{y}$ and $\tilde{X}$ is a matter of some simple scalar transformations, as is well-known from the ANOVA literature [see, for example, Scheffé (1959)].

This sets the stage for our paper. In Section 2, we line up various fixed effects model specifications proposed in the literature for three-dimensional data. For each of these models, we present specific D and M_D matrices and derive the “tilde” scalar transformations. Often, the data are flow type, where the nature of the observations is such that there are no self-flows. This requires different and more complicated transformations. To get the feeling for what is at stake, in Section 3 we analyze the two-way model, which has not yet been described in the literature, and discuss from there the various three-way cases. Data with no self-flow are a rather well-behaved special kind of unbalanced data as they still allow for fairly simple scalar transformations. The situation is less favorable in the general unbalanced case, as shown in Section 4.

In Sections 2, 3, and 4 only static models were considered. In Section 5, we show how the presence of the lagged dependent variable may render OLS on the transformed data inconsistent, thus generalizing the well-known Nickell (1981) bias. Somewhat surprisingly with three-way data, inconsistency does not occur in all models. For the cases with inconsistency we present the appropriate generalization of the Arellano-Bond estimator. Section 6 concludes.

Throughout the paper, we use the conventional ANOVA notation and indicate the average over an index for a variable by denoting a bar on the variable and a dot on the place of that index. When discussing unbalanced data, a plus sign at the place of an index indicates summation over that index. The matrix M with a subscript denotes projection orthogonal to the space spanned by the subscript.

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}, $i = 1, \ldots, N_i$, $j = 1, \ldots, N_j$, 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 $N_i = N_j = N$. Implicitly, we also assume that the set of individuals in the observation sets i and j are the same, although we relax these assumptions later on. The main question is how to formalize the individual and time heterogeneity — in our case, the fixed effects.

2.1 The Model with Three Effects

The first attempt to properly extend the standard fixed effects panel data model [see, for example, Baltagi (2005) or Balestra and Krishnakumar (2008)] to a multi-dimensional setup was proposed by Matyas (1997). The specification of this model is

$$y_{ijt} = \beta' x_{ijt} + \alpha_i + \gamma_j + \lambda_t + \varepsilon_{ijt} \qquad i = 1, \ldots, N \quad j = 1, \ldots, N, \quad t = 1, \ldots, T, \quad (2)$$

where the α_i, γ_j, and λ_t parameters are the country and time-specific fixed effects, the x_{ijt} variables are the usual covariates, β $(K \times 1)$ is the vector of the structural parameters; and ε_{ijt} are the $i.i.d.(0, \sigma_\varepsilon^2)$ idiosyncratic disturbance terms. We also assume that the covariates and the disturbance terms are uncorrelated. Now, in model (1), y is the vector of the dependent variable of size $(N^2T \times 1)$; X is the matrix of the covariates of size $(N^2T \times K)$;

$$D = (I_N \otimes l_{NT},\ l_N \otimes I_N \otimes l_T,\ l_{N^2} \otimes I_T)$$

is the $(N^2T \times (2N+T))$ dummy matrix with column rank $(2N+T-2)$ corresponding to the fixed effects, with I_N and l_N being the identity matrix and the column vector of ones respectively, with the sizes indicated in the index; and $\pi = (\alpha', \gamma', \lambda')'$ is the $((2N+T) \times 1)$ vector of the fixed effects. Wansbeek (1991) has shown that the column space of D does not change by replacing I_N (or similarly I_T) with any $(G_N, \bar{l}_N)$ orthonormal matrix of order $(N \times (N-1))$, where G_N has to satisfy the following conditions:

$$G_N' l_N = 0, \quad \text{and} \quad G_N' G_N = I_{N-1} \quad \text{with} \quad \bar{l}_N \equiv l_N / \sqrt{N} \, .$$

As matrix D spans the same vector space as the following orthonormal matrix

$$\widetilde{D} \equiv \big((G_N \otimes \bar{l}_N \otimes \bar{l}_T),\ (\bar{l}_N \otimes G_N \otimes \bar{l}_T),\ (\bar{l}_N \otimes \bar{l}_N \otimes G_T),\ (\bar{l}_N \otimes \bar{l}_N \otimes \bar{l}_T) \big) \, ,$$

which has in fact full column rank $(2N+T-2)$, the projection matrix of size $(N^2T \times N^2T)$ to eliminate D is simply

$$\begin{aligned} M_D &\equiv I_{N^2T} - \tilde{D}\tilde{D}' \\ &= I_{N^2T} - (Q_N \otimes \bar{J}_{NT}) - (\bar{J}_N \otimes Q_N \otimes \bar{J}_T) - (\bar{J}_{N^2} \otimes Q_T) - \bar{J}_{N^2T} \\ &= I_{N^2T} - (I_N \otimes \bar{J}_{NT}) - (\bar{J}_N \otimes I_N \otimes \bar{J}_T) - (\bar{J}_{N^2} \otimes I_T) + 2\bar{J}_{N^2T} \, , \end{aligned}$$

with $\bar{J}_N \equiv \bar{l}_N \bar{l}_N' = l_N l_N' / N$ and $Q_N \equiv G_N G_N' = I_N - \bar{J}_N$. This matrix operation defines the scalar transformation

$$\tilde{y}_{ijt} = y_{ijt} - \bar{y}_{i..} - \bar{y}_{.j.} - \bar{y}_{..t} + 2\bar{y}_{...} \tag{3}$$

Note that this optimal within transformation actually numerically gives the same parameter estimates as the direct OLS estimation of model (2). We must emphasize that these Within transformations are usually not unique. For example, a simple transformation that also eliminates the fixed effects from model (2) is

$$\tilde{y}_{ijt} = y_{ijt} - \bar{y}_{ij.} - \bar{y}_{..t} + \bar{y}_{...} \tag{4}$$

This model is suited to deal with purely cross-sectional data as well (that is, when $T = 1$). In this case, there are only the α_i and γ_j fixed effects and the appropriate within transformation is $\tilde{y}_{ijt} = y_{ij} - \bar{y}_{.j} - \bar{y}_{i.} + \bar{y}_{..}$.

2.2 Models with Composite Effects

There is no reason to stop at model (2) after we derive its optimal within transformation, as similar reasoning can be done for models with different fixed effects formulations. We collect them here with their unique D dummy matrices, $\tilde{D}$ orthonormal matrices, and M_D optimal projection matrices.

A model has been proposed by Egger and Pfaffermayr (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. Now

$$\begin{aligned} D =& (I_N \otimes I_N \otimes l_T) \quad \text{of size} \quad (N^2T \times N^2)\,, \\ \tilde{D} =& (I_N \otimes I_N \otimes \bar{l}_T) \quad \text{of size} \quad (N^2T \times N^2)\,, \end{aligned}$$

both with full column ranks N^2. The optimal projection matrix orthogonal to D is

$$M_D = I_{N^2T} - (I_{N^2} \otimes \bar{J}_T),$$

defining the scalar operation

$$\tilde{y}_{ijt} = y_{ijt} - \bar{y}_{ij.} \,. \tag{6}$$

A variant of model (5), proposed by Cheng and Wall (2005), often used in empirical studies is

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

Now

$$\begin{aligned} D &= ((I_N \otimes I_N \otimes l_T),\ (l_N \otimes l_N \otimes I_T)) \quad \text{of size} \quad (N^2T \times (N^2 + T))\,, \\ \tilde{D} &= \big((I_N \otimes I_N \otimes \bar{l}_T),\ (\bar{l}_N \otimes \bar{l}_N \otimes G_T)\big) \quad \text{of size} \quad (N^2T \times (N^2 + T - 1))\,, \end{aligned}$$

each with column ranks $(N^2 + T - 1)$, so

$$M_D = I_{N^2T} - (I_N \otimes I_N \otimes \bar{J}_T) - (\bar{J}_N \otimes \bar{J}_N \otimes I_T) + (\bar{J}_N \otimes \bar{J}_N \otimes \bar{J}_T)\,,$$

defining in fact (4). As model (2) is a special case of model (7), transformation (4) can naturally be used to clear the fixed effects here as well. While transformation (4) leads to the optimal within estimator for model (7), it is clear why it is not optimal for model (2): 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 2D panel data models, where the individuals are now the (ij) pairs.

Baltagi et al. (2003), Baldwin and Taglioni (2006) and Baier and Bergstrand (2007) suggest several other forms of fixed effects. A simpler model is

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

with

$$D = (l_N \otimes I_N \otimes I_T) \quad \text{of size} \quad (N^2T \times NT)\,,$$
$$\tilde{D} = \left(\bar{l}_N \otimes I_N \otimes I_T\right) \quad \text{of size} \quad (N^2T \times NT)\,.$$

both with full column rank (NT). Thus,

$$M_D = I_{N^2T} - (\bar{J}_N \otimes I_N \otimes I_T)$$

defining the simple

$$\tilde{y}_{ijt} = y_{ijt} - \bar{y}_{.jt} \tag{9}$$

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

A variation of this model is

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

where now

$$D = ((I_N \otimes l_N \otimes I_T),\ (l_N \otimes I_N \otimes I_T)) \quad \text{of size} \quad (N^2T \times 2NT)\,,$$
$$\tilde{D} = \left((G_N \otimes \bar{l}_N \otimes I_T),\ (\bar{l}_N \otimes G_N \otimes I_T), (\bar{l}_N \otimes \bar{l}_N \otimes I_T)\right) \quad \text{of size} \quad (N^2T \times (2NT-1))\,,$$

each with column ranks ($2NT-1$), so

$$M_D = I_{N^2T} - (I_N \otimes \bar{J}_N \otimes I_T) - (\bar{J}_N \otimes I_N \otimes I_T) + (\bar{J}_N \otimes \bar{J}_N \otimes I_T)\,.$$

This matrix operation defines the scalar optimal within transformation

$$\tilde{y}_{ijt} = y_{ijt} - \bar{y}_{.jt} - \bar{y}_{i.t} + \bar{y}_{..t}\,. \tag{11}$$

Let us notice here that transformation (11) clears the fixed effects for model (2) as well, but of course the resulting within estimator is not optimal.

The model that encompasses all the above effects is

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

By applying suitable restrictions to model (12), we can obtain all models discussed above. As

$$D = ((I_N \otimes I_N \otimes l_T),\ (I_N \otimes l_N \otimes I_T),\ (l_N \otimes I_N \otimes I_T))$$
$$\tilde{D} = ((G_N \otimes G_N \otimes \bar{l}_T),\ (G_N \otimes \bar{l}_N \otimes G_T),\ (\bar{l}_N \otimes G_N \otimes G_T),\ (G_N \otimes \bar{l}_N \otimes \bar{l}_T),$$
$$(\bar{l}_N \otimes G_N \otimes \bar{l}_T),\ (\bar{l}_N \otimes \bar{l}_N \otimes G_T),\ (\bar{l}_N \otimes \bar{l}_N \otimes \bar{l}_T))$$

of sizes $(N^2T \times (N^2 + 2NT))$ and $(N^2T \times (N^2 + 2NT - 2))$, respectively, and with column ranks $(N^2 + 2N(T-1) - (T-1))$, the projection orthogonal to D is simply $(Q_N \otimes Q_N \otimes Q_T)$, or

$$\begin{aligned} M_D =& I_{N^2T} - (I_{N^2} \otimes \bar{J}_T) - (I_N \otimes \bar{J}_N \otimes I_T) - (\bar{J}_N \otimes I_{NT}) \\ &+ (I_N \otimes \bar{J}_{NT}) + (\bar{J}_N \otimes I_N \otimes \bar{J}_T) + (\bar{J}_{N^2} \otimes I_T) - \bar{J}_{N^2T} \,. \end{aligned}$$

The within transformation for this model, defined by M_D, is as simple as

$$\tilde{y}_{ijt} = y_{ijt} - \bar{y}_{ij.} - \bar{y}_{.jt} - \bar{y}_{i.t} + \bar{y}_{..t} + \bar{y}_{.j.} + \bar{y}_{i..} - \bar{y}_{...} \,. \tag{13}$$

2.3 The Relationship Between the Models

As discussed, six model structures have been outlined in the literature for dummy variables in fixed effects three-way models. Next, we study the relationship between these models. We first see what insight we gain by comparing them to each other. We then show how far these proposed models cover all theoretically possible models. Finally in this section, we demonstrate to what extent a transformation for one particular model clears the dummy variables in each of the other models.

To start with, let us make the structure of the models visible. A clear way to do this is through the projection matrix $\tilde{D}\tilde{D}' = I_{N^2T} - M_D$, which projects the data into the space spanned by the dummy variables D. This matrix can be easily obtained, as $\tilde{D}$ for each model has already been written out explicitly. In elaborating $\tilde{D}\tilde{D}'$, we replace I_N by $Q_N + \bar{J}_N$ and likewise for Q_T. Remember that $\bar{J}_N = \bar{l}_N\bar{l}'_N = l_Nl'_N/N$ and $Q_N = G_NG'_N = I_N - \bar{J}_N$. Results are presented in Table 1. Each column of the table corresponds to one particular model and a + sign indicates which building blocks have to be used to get the appropriate $\tilde{D}\tilde{D}'$.

Table 1: Building blocks in projection matrices
$(I_{N^2T} - M_D) = \tilde{D}\tilde{D}'$

Model	(2)	(5)	(7)	(8)	(10)	(12)
$Q_N \otimes Q_N \otimes Q_T$						
$Q_N \otimes Q_N \otimes \bar{J}_T$		+	+			+
$Q_N \otimes \bar{J}_N \otimes Q_T$					+	+
$\bar{J}_N \otimes Q_N \otimes Q_T$				+	+	+
$Q_N \otimes \bar{J}_N \otimes \bar{J}_T$	+	+	+		+	+
$\bar{J}_N \otimes Q_N \otimes \bar{J}_T$	+	+	+	+	+	+
$\bar{J}_N \otimes \bar{J}_N \otimes Q_T$	+		+	+	+	+
$\bar{J}_N \otimes \bar{J}_N \otimes \bar{J}_T$	+	+	+	+	+	+

We see that the first row of Table 1 is empty. Any model producing a $\tilde{D}\tilde{D}'$ with a non-empty first row would indicate fixed effects with three indices, which evidently does not make sense for three-way data. The fact that the last row of the table does not have empty cells means that all structures have effects that add up to one. The number of models with an empty first row and a full last row is 2^6. As these models are nested into at least one of the six models we cover, we take in fact care of all relevant cases.

Let us now address the question of the extent to which the transformation for one model clears the effects of another one. Model A does so for model B if, in obvious notation, $M_A D_B = 0$. In terms of Table 1, this is the case when the + signs of model A cover those of model B. We see that the transformation of the "all-encompassing" model (12) covers all cases, while the transformations of model (10) covers only models (2) and (8), and so on. Another kind of insight from this exercise is into the possible effect of misspecification error(s). When, for example, the true model is (7) and we use the transformation corresponding to model (10), the effects are not fully cleared, thus leading to a bias in the estimation of β. The argument can go the other way as well. If, for example, transformation (13) is used for model (5), we in effect "over-clear" the fixed effects, thus leading to a loss of efficiency.

2.4 Beyond Three Dimensions

But what if our data is such that variables are observed along four, or even five, dimensions? Take the following example. 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}. If the data at hand is not only for a given country, but for several, with product and firm observations, we would end up with five-dimensional panel data. It is clear that such higher-dimensional setups involve several possible fixed effects specifications (a number that grows radically along with the dimensions), making the full collection of such models non-trivial. We can, however, see how to generalize our results on a four-dimensional benchmark model and this approach can then easily be extended to higher dimensions as well.

Take the four-dimensional extension of the all-encompassing model (12) with pair-wise interaction effects:

$$y_{ijst} = x'_{ijst}\beta + \gamma^0_{ijs} + \gamma^1_{ijt} + \gamma^2_{jst} + \gamma^3_{ist} + \varepsilon_{ijst}\,, \tag{14}$$

with $i = 1 \ldots N_i$, $j = 1 \ldots N_j$, $s = 1 \ldots N_s$, and $t = 1 \ldots T$. Notice, that now

$$D = \big((I_{N_iN_jN_s} \otimes l_T),\ (I_{N_iN_j} \otimes l_{N_s} \otimes I_T),\ (l_{N_i} \otimes I_{N_jN_sT}),\ (I_{N_i} \otimes l_{N_j} \otimes I_{N_sT})\big)$$

is a $(N_iN_jN_sT \times (N_iN_jN_s + N_iN_jT + N_jN_sT + N_iN_sT))$ dummy coefficient matrix with column rank $(N_iN_jN_s + N_iN_jT + N_jN_sT + N_iN_sT - 3)$, and the optimal projection orthogonal to D is the $(N_iN_jN_sT \times N_iN_jN_sT)$ matrix

$$\begin{aligned} M_D = \quad & I_{N_iN_jN_sT} - \left(\bar{J}_{N_i} \otimes I_{N_jN_sT}\right) - \left(I_{N_i} \otimes \bar{J}_{N_j} \otimes I_{N_sT}\right) - \left(I_{N_iN_j} \otimes \bar{J}_{N_s} \otimes I_T\right) \\ & - \left(I_{N_iN_jN_s} \otimes \bar{J}_T\right) + \left(\bar{J}_{N_iN_j} \otimes I_{N_sT}\right) + \left(\bar{J}_{N_i} \otimes I_{N_j} \otimes \bar{J}_{N_s} \otimes I_T\right) \\ & + \left(\bar{J}_{N_i} \otimes I_{N_jN_s} \otimes \bar{J}_T\right) + \left(I_{N_i} \otimes \bar{J}_{N_jN_s} \otimes I_T\right) + \left(I_{N_i} \otimes \bar{J}_{N_j} \otimes I_{N_s} \otimes \bar{J}_T\right) \\ & + \left(I_{N_iN_j} \otimes \bar{J}_{N_sT}\right) - \left(\bar{J}_{N_iN_jN_s} \otimes I_T\right) - \left(\bar{J}_{N_iN_j} \otimes I_{N_s} \otimes \bar{J}_T\right) \\ & - \left(\bar{J}_{N_i} \otimes I_{N_j} \otimes \bar{J}_{N_sT}\right) - \left(I_{N_i} \otimes \bar{J}_{N_jN_sT}\right) + \bar{J}_{N_iN_jN_sT} \, . \end{aligned}$$

It can be seen that M_D in fact defines the optimal scalar within transformation

$$\begin{aligned} \tilde{y}_{ijst} = y_{ijst} - \bar{y}_{.jst} - \bar{y}_{i.st} - \bar{y}_{ij.t} - \bar{y}_{ijs.} + \bar{y}_{..st} + \bar{y}_{.j.t} + \bar{y}_{.js.} \\ + \bar{y}_{i..t} + \bar{y}_{i.s.} + \bar{y}_{ij..} - \bar{y}_{...t} - \bar{y}_{..s.} - \bar{y}_{.j..} - \bar{y}_{i...} + \bar{y}_{....} \end{aligned} \tag{15}$$

needed to eliminate $(\gamma^0_{ijs},\ \gamma^1_{ijt},\ \gamma^2_{jst},\ \gamma^3_{ist})$.

3. No Self-Flow Data

Often the models we study are used to deal with flow types of data like trade and capital movements (FDI) between countries. In such cases i and j index the same entities, $N_i = N_j = N$ and there is, by definition, no self-flow. In terms of the models from Section 2, we have a case of missing data and the transformations that we give can no longer be applied. Fortunately, the pattern of the missing observations is highly structured, allowing for the derivation of optimal transformations that are still quite simple. We start with presenting the derivation of the optimal transformation for the $T = 1$ case in some detail as, to the best of our knowledge, even this has not been studied in the literature. This leads us to the appropriate within transformation and also offers the main tools for deriving the optimal transformation for all models from Section 2 with $T > 1$. The derivations are given in the online supplement of this paper.[2]

3.1 The Cross-sectional Case

In the case when $T = 1$, there is only one relevant model. For $i, j = 1, \dots, N$,

$$y_{ij} = \beta' x_{ij} + \alpha_i + \gamma_j + \varepsilon_{ij}, \tag{16}$$

[2] See: *http://personal.ceu.edu/staff/matyas/BMW-Supplement.pdf.*

or in matrix form,

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

As there are no data with $i = j$, we eliminate these from the model by using the selection matrix L of order $N^2 \times N(N-1)$ to get

$$L'y = L'X\beta + L'D(\alpha', \gamma')' + L'\varepsilon.$$

So the optimal effects-eliminating projection matrix is

$$M_{L'D} = I_{N(N-1)} - L'DW^+D'L,$$

with $W = D'LL'D$ and "+" denoting the Moore-Penrose generalized inverse. We want to have a simple expression for the elements of $M_{L'D}L'y$, indicated by a tilde. When in the data $i = j$ are observed, this expression is

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

Now, the issue gets more complicated. For $i \neq j$, $(e_i \otimes e_j)'D = (e_i \otimes e_j)'LL'D = (e_i', e_j')$, so

$$\begin{aligned} \tilde{y}_{ij} &= (e_i \otimes e_j)'LM_{L'D}L'y \\ &= y_{ij} - (e_i', e_j')W^+D'LL'y, \end{aligned} \qquad (17)$$

with e_i being the ith unit vector of size N. This causes us to further elaborate on W. Since

$$D_\alpha' LL' D_\alpha = D_\gamma' LL' D_\gamma = (N-1)I_N \quad \text{and} \quad D_\alpha' LL' D_\gamma = J_N - I_N$$

and, as before, $\bar{J}_N = J_N/N$ and $Q_N = I_N - \bar{J}_N$, we obtain

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

Since Q_N and $\bar{J}_N$ are idempotent and mutually orthogonal, the Moore-Penrose inverse W^+ of W is

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

with

$$p = \frac{1}{4(N-1)} - \frac{N-1}{N(N-2)} \quad \text{and} \quad q = \frac{1}{4(N-1)} - \frac{1}{N(N-2)} \; .$$

Now, with this updated form of W,

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

Moreover, with Y being the $(N \times N)$ data matrix containing the y_{ij} observations, with zeros filling in the empty diagonal elements,

$$D'LL'y = \begin{pmatrix} Y'l_N \\ Yl_N \end{pmatrix} \; .$$

So, after multiplying $(e_i', e_i')W^{+}$ and $D'LL'y$, and as $y_{++} = l_N'Yl_N = l_N'Y'l_N$, we get

$$\begin{aligned} \tilde{y}_{ij} = y_{ij} &- \frac{N-1}{N(N-2)}\,(y_{i+} + y_{+j}) - \frac{1}{N(N-2)}\,(y_{+i} + y_{j+}) \\ &+ \frac{1}{(N-1)(N-2)} y_{++} \, . \end{aligned} \tag{18}$$

When N grows larger, the effects of the missing diagonal elements becomes smaller, which is reflected in the above expression by the third term at the right-hand side of formula (18) being of lower order than N.

3.2 The Model with Three Effects

Let us turn our attention back to the three-dimensional models. To derive the optimal within transformation for model (2), we start from its matrix form:

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

From this point, with $D = (D_{\alpha_*},\ D_{\gamma_*},\ D_{\lambda})$, the derivations are very similar to those used for pure cross-sections, only appearing slightly more complicated. The optimal within transformation for model (2) in the no self-flow case is

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

3.3 Models with Composite Effects

Now, let us continue with the models with composite effects. In most cases, the optimal within transformation has to be adjusted only moderately, to reflect the missing diagonal elements. For model (5), this reads as

$$\tilde{y}_{ijt} = y_{ijt} - \frac{1}{T} y_{ij+} \, ; \tag{20}$$

for model (7), it is

$$\tilde{y}_{ijt} = y_{ijt} - \frac{1}{T} y_{ij+} - \frac{1}{N(N-1)} y_{++t} + \frac{1}{TN(N-1)} y_{+++} \, ; \tag{21}$$

and for model (8),

$$\tilde{y}_{ijt} = y_{ijt} - \frac{1}{N-1} y_{+jt} \, . \tag{22}$$

Note that all three transformations above are in fact very similar to their complete data counterparts from Section 2. Next, let us move on to model (10). The optimal within transformation, following the method discussed above is obtained as

$$\begin{aligned} \tilde{y}_{ijt} = y_{ijt} &- \frac{N-1}{N(N-2)} \left(y_{i+t} + y_{+jt} \right) - \frac{1}{N(N-2)} \left(y_{+it} + y_{j+t} \right) \\ &+ \frac{1}{(N-1)(N-2)} y_{++t} \, . \end{aligned} \tag{23}$$

For model (12), we follow a somewhat different approach. First, we have to notice that the optimal effects-clearing projection matrix, with $\tilde{L}$ being the selection matrix of order $(N^2T \times N(N-1)T)$, is now

$$\begin{aligned} M_{\tilde{L}'D} &= M_{\tilde{L}'\tilde{D}} \\ &= \tilde{L}' M_D \tilde{L} - \tilde{L}'\tilde{D}\tilde{D}'(H \otimes I_T) V (H \otimes I_T)' \tilde{D}\tilde{D}'\tilde{L} \, , \end{aligned}$$

with

$$H = \sum_i e_i e_i' \otimes e_i \, , \quad \tilde{L}\tilde{L}' = I_{N^2T} - (HH' \otimes I_T) \, , \quad \text{and}$$

$$V = \left(I - (H \otimes I_T)' \tilde{D}\tilde{D}'(H \otimes I_T) \right)^{-} = ((Q_N \cdot Q_N) \otimes Q_T)^{-} \, .$$

Intuitively enough, the first part of the projection corresponds to the transformation used in the case of complete data, while the second term corrects for the missing $i = j$ observations. All elements of $M_{\tilde{L}'\tilde{D}}$ have already been defined and D and $\tilde{D}$ are the model-specific dummy matrices. After some elaboration on the projection, we see that for observation y_{ijt} $(i \neq j)$,

$$(e_i \otimes e_j \otimes e_t)' M_{\tilde{L}'\tilde{D}} y$$

gives

$$
\begin{aligned}
\tilde{y}_{ijt} =& y_{ijt} - \frac{N-3}{N(N-2)}(y_{i+t} + y_{+jt}) + \frac{N-3}{N(N-2)T}(y_{i++} + y_{+j+}) - \frac{1}{T}y_{ij+} \\
&+ \frac{1}{N(N-2)}(y_{+it} + y_{j+t}) - \frac{1}{N(N-2)T}(y_{+i+} + y_{j++}) \\
&+ \frac{N^2-6N+4}{N^2(N-1)(N-2)}(y_{++t} - y_{+++})\,.
\end{aligned} \tag{24}
$$

Note that this method is also flexibly applicable for all fixed effects model formulations as one only has to substitute in the specific D and $\tilde{D}$ dummy matrices corresponding to the given model.

The listed no self-flow transformations can also be generalized to any higher dimensions. In the four-dimensional case we get

$$
\begin{aligned}
\tilde{y}_{ijst} =& y_{ijst} - \frac{1}{N-1}y_{+jst} - \frac{1}{N-1}y_{i+st} - \frac{1}{N_s}y_{ij+t} - \frac{1}{T}y_{ijs+} + \frac{1}{(N-1)^2}y_{++st} \\
&+ \frac{1}{(N-1)N_s}y_{+j+t} + \frac{1}{(N-1)T}y_{+js+} + \frac{1}{(N-1)N_s}y_{i++t} + \frac{1}{(N-1)T}y_{i+s+} \\
&+ \frac{1}{N_sT}y_{ij++} - \frac{1}{(N-1)^2N_s}y_{+++t} - \frac{1}{(N-1)^2T}y_{++s+} - \frac{1}{(N-1)N_sT}y_{+j++} \\
&- \frac{1}{(N-1)N_sT}y_{i+++} + \frac{1}{(N-1)^2N_sT}y_{++++} - \frac{1}{(N-1)N_sT}y_{ji++} \\
&+ \frac{1}{(N-1)T}y_{jis+} + \frac{1}{(N-1)N_s}y_{ji+t} - \frac{1}{N-1}y_{jist}\,.
\end{aligned} \tag{25}
$$

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

Next, we go further along the above lines and see what is going to happen if the observation sets i and j are different. If the two sets are completely disjoint, say for example, if we are modeling export activity between the EU and APEC countries, for all the models considered, the within estimators are unbiased, as the no self-flow problem does not arise. If the two sets are not completely disjoint, say for example in the case of trade between the EU and OECD countries, when the no self-flow problem does arise, we are faced with the same biases that are outlined above. Unfortunately, however, transformations (19), (23) and (24) do not work in this case and there are no obvious transformations that could be worked out for this scenario.

4. Unbalanced Data

As in the case of the usual 2D panel data sets [see Wansbeek and Kapteyn (1989) or Baltagi (2005), for example], just more frequently, one may be faced with a situation in which the data at hand is unbalanced. In our framework of analysis this means that for all the previously studied models, in general $t \in T_{ij}$, for all (ij) pairs, where T_{ij} is a subset of the index set $t \in \{1, \ldots, T\}$, with T being chronologically the last time period in which we have any (i, j) observations. Note that two T_{ij} and $T_{i'j'}$ sets are usually different and also let $R = \sum_{ij} |T_{ij}|$ denote the total number of observations, where $|T_{ij}|$ is the cardinality of the set T_{ij} (the number of observations in the given set).

For models (5) and (8), the unbalanced nature of the data does not cause any problem, the within transformations can be used, and they have exactly the same properties, as in the balanced case. However, for models (2), (7), (10), and (12), we face some problems. As the within transformations fail to fully eliminate the fixed effects for these models (somewhat similarly to the no self-flow case), the resulting within estimators suffer from (potentially severe) biases. However, luckily, the Wansbeek and Kapteyn (1989) approach can be extended to these four cases.

Let us start with model (2). Dummy variable matrix D has to be modified to reflect the unbalanced nature of the data. Let the U_t and V_t $(t = 1 \ldots T)$ be the sequence of $(I_N \otimes l_N)$ and $(l_N \otimes I_N)$ matrices, respectively, in which the following adjustments were made: for each (ij) observation, we leave the row [representing (ij)] in those U_t and V_t matrices untouched, are $t \in T_{ij}$, but delete it from the remaining $T - |T_{ij}|$ matrices. In this way we end up with the following dummy variable setup

$$D_1^a = (U_1', U_2', \ldots, U_T')' \quad \text{of size} \quad (R \times N)\,,$$
$$D_2^a = (V_1', V_2', \ldots, V_T')' \quad \text{of size} \quad (R \times N)\,, \text{ and}$$
$$D_3^a = \text{diag}\{V_1 \cdot l_N, V_2 \cdot l_N \ldots, V_T \cdot l_N\} \quad \text{of size} \quad (R \times T)\,.$$

So the complete dummy variable structure is now $D_a = (D_1^a, D_2^a, D_3^a)$. In this case, let us note here that, just as in Wansbeek and Kapteyn (1989), index t goes "slowly" and ij goes "fast". Now with this modified dummy variable structure, the optimal projection removing the fixed effects can be obtained in three steps:

$$M_{D_a}^{(1)} = I_R - D_1^a (D_1^{a'} D_1^a)^{-1} D_1^{a'}\,,$$
$$M_{D_a}^{(2)} = M_{D_a}^{(1)} - M_{D_a}^{(1)} D_2^a (D_2^{a'} M_{D_a}^{(1)} D_2^a)^{-} D_2^{a'} M_{D_a}^{(1)}\,,$$

and finally

$$M_{D_a} = M^{(3)}_{D_a} = M^{(2)}_{D_a} - M^{(2)}_{D_a} D^a_3 (D^{a\prime}_3 M^{(2)}_{D_a} D^a_3)^{-} D^{a\prime}_3 M^{(2)}_{D_a} \,, \tag{26}$$

where "$^-$" stands for any generalized inverse. It is easy to see that in fact $M_{D_a} D_a = 0$ projects out all three dummy matrices. Note that in the balanced case $(D^{a\prime}_1 D^a_1)^{-1} = I_N/(NT)$, but now

$$(D^{a\prime}_1 D^a_1)^{-1} = \text{diag}\left\{ \frac{1}{\sum_j |T_{1j}|}, \frac{1}{\sum_j |T_{2j}|}, \ldots, \frac{1}{\sum_j |T_{Nj}|} \right\} \quad \text{of size} \quad (N \times N) \,.$$

With this in hand, we only have to calculate two inverses instead of three, $(D^{a\prime}_2 M^{(1)}_{D_a} D^a_2)^-$, and $(D^{a\prime}_3 M^{(2)}_{D_a} D^a_3)^-$, with respective sizes $(N \times N)$ and $(T \times T)$. This is feasible for reasonable sample sizes.

For model (7), the job is essentially the same. Let the W_t $(t = 1 \ldots T)$ be the sequence of $(I_N \otimes I_N)$ matrices, where again for each (ij), we remove the rows corresponding to observation (ij) in those W_t, where $t \notin T_{ij}$. In this way,

$$\begin{aligned} D^b_1 &= (W'_1, W'_2, \ldots, W'_T)' \quad \text{of size} \quad (R \times N^2) \,, \\ D^b_2 &= D^a_3 \quad \text{of size} \quad (R \times T) \,. \end{aligned}$$

The first step in the projection is now

$$M^{(1)}_{D_b} = I_R - D^b_1 (D^{b\prime}_1 D^b_1)^{-1} D^{b\prime}_1 \,,$$

so the optimal projection orthogonal to $D_b = (D^b_1, D^b_2)$ is simply

$$M_{D_b} = M^{(2)}_{D_b} = M^{(1)}_{D_b} - M^{(1)}_{D_b} D^b_2 (D^{b\prime}_2 M^{(1)}_{D_b} D^b_2)^{-} D^{b\prime}_2 M^{(1)}_{D_b} \,. \tag{27}$$

Note that as

$$(D^{b\prime}_1 D^b_1)^{-1} = \text{diag}\left\{ \frac{1}{|T_{11}|}, \frac{1}{|T_{12}|}, \ldots, \frac{1}{|T_{NN}|} \right\} \quad \text{of size} \quad (N^2 \times N^2) \,,$$

we only have to calculate the inverse of a $(T \times T)$ matrix – $D^{b\prime}_2 M^{(1)}_{D_b} D^b_2$ – which is easily doable. Further, as discussed above, given that model (2) is nested in (7), transformation (27) is in fact also valid for model (2).

Let us move on to model (10). Now, after the same adjustments as before,

$$\begin{aligned} D^c_1 &= \text{diag}\{U_1, U_2, \ldots, U_T\} \quad \text{of size} \quad (R \times NT) \quad \text{and} \\ D^c_2 &= \text{diag}\{V_1, V_2, \ldots, V_T\} \quad \text{of size} \quad (R \times NT) \,, \end{aligned}$$

so the stepwise projection, removing $D_c = (D_1^c, D_2^c)$, is

$$M_{D_c}^{(1)} = I_R - D_1^c (D_1^{c'} D_1^c)^{-1} D_1^{c'} ,$$

leading to

$$M_{D_c} = M_{D_c}^{(2)} = M_{D_c}^{(1)} - M_{D_c}^{(1)} D_2^c (D_2^{c'} M_{D_c}^{(1)} D_2^c)^{-} D_2^{c'} M_{D_c}^{(1)} . \tag{28}$$

Note that for M_{D_c}, we have to invert $(NT \times NT)$ matrices, which can be computationally difficult.

The last model to deal with is model (12). Let $D_d = (D_1^d, D_2^d, D_3^d)$, where the adjusted dummy matrices are all defined above:

$$\begin{aligned} D_1^d &= D_1^b \quad \text{of size} \quad (R \times N^2) , \\ D_2^d &= D_1^c \quad \text{of size} \quad (R \times NT) , \\ D_3^d &= D_2^c \quad \text{of size} \quad (R \times NT) . \end{aligned}$$

Defining the partial projector matrices $M_{D_d}^{(1)}$ and $M_{D_d}^{(2)}$ as

$$\begin{aligned} M_{D_d}^{(1)} &= I_R - D_1^d (D_1^{d'} D_1^d)^{-1} D_1^{d'} \text{ and} \\ M_{D_d}^{(2)} &= M_{D_d}^{(1)} - M_{D_d}^{(1)} D_2^{d'} (D_2^{d'} M_{D_d}^{(1)} D_2^d)^{-} D_2^{d'} M_{D_d}^{(1)} , \end{aligned}$$

the appropriate transformation for model (12) is now

$$M_{D_d} = M_{D_d}^{(3)} = M_{D_d}^{(2)} - M_{D_d}^{(2)} D_3^{d'} (D_3^{d'} M_{D_d}^{(2)} D_3^d)^{-} D_3^{d'} M_{D_d}^{(2)} . \tag{29}$$

It can be easily verified that M_{D_d} is idempotent and $M_{D_d} D_d = 0$, so all the fixed effects are indeed eliminated.[3] As model (10) is covered by model (12), projection (29) eliminates the fixed effects from that model as well. Moreover, as suggested above, all three-way fixed effects models are in fact nested into model (12). It is therefore intuitive that transformation (29) clears the fixed effects in all model formulations. Using (29) is not always advantageous, however, as the transformation involves the inversion of potentially large matrices (of order NT). In the case of most models studied, we can find suitable unbalanced transformations at the cost of only inverting $(T \times T)$ matrices; or in some cases, we can even derive scalar transformations. It is good to know, however, that there is a general projection that is universally applicable to all three-way models in the presence of all kinds of data issues.

[3] A STATA program code for transformation (29) with a user-friendly detailed explanation is available at *www.personal.ceu.hu/staff/repec/pdf/stata-program_document-dofile.pdf.* Estimation of model (12) is then easily done for any kind of incompleteness.

It is worth noting that transformations (26), (27), (28), and (29) are all 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 variable matrices.

All transformations detailed above can also be rewritten in a semi-scalar form. Let us show here how this idea works on transformation (29), as all subsequent transformations can be dealt with in the same way. Let

$$\phi = C^{-}\bar{D}'y \quad \text{and} \quad \omega = \tilde{C}^{-}(M_{D_d}^{(2)}D_3^d)'y \quad \xi = C^{-}\bar{D}'D_3^d\omega \,,$$

where

$$C = \left(D_2^d\right)'\bar{D}\,, \quad \bar{D} = \left(I_R - D_1^d(D_1^{d'}D_1^d)^{-1}D_1^{d'}\right)D_2^d\,, \quad \text{and} \quad \tilde{C} = D_3^{d'}M_{D_d}^{(2)}D_3^d\,.$$

Now the scalar representation of transformation (29) is

$$[M_{D_d}y]_{ijt} = y_{ijt} - \frac{1}{|T_{ij}|}\sum_{t\in T_{ij}} y_{ijt} + \frac{1}{|T_{ij}|}a'_{ij}\phi - \phi_{it} - \omega_{jt} + \frac{1}{|T_{ij}|}\tilde{a}'_{ij}\omega + \xi_{it} - \frac{1}{|T_{ij}|}\left(a_{ij}^b\right)'\xi\,,$$

where a_{ij} and $\tilde{a}_{ij}$ are the column vectors corresponding to observations (ij) from matrices $A = D_2^{d'}D_1^d$ and $\tilde{A} = D_3^{d'}D_1^d$, respectively; ϕ_{it} is the (it)-th element of the $(NT \times 1)$ column vector ϕ; ω_{jt} is the (jt)-th element of the $(NT \times 1)$ column vector ω; and finally, ξ_{it} is the element corresponding to the (it)-th observation from the $(NT \times 1)$ column vector, ξ.[4]

Transformation (29) can also be generalized into a four-dimensional setup. Let the dummy variables matrices for the four fixed effects in (14) be denoted by $D_e = (D_1^e, D_2^e, D_3^e, D_4^e)$ and let $M_{D_e}^{(k)}$ be the transformation that clears out the first k fixed effects; namely, $M_{D_e}^{(k)} \cdot (D_1^e, \ldots, D_k^e) = (0, \ldots, 0)$ for $k = 1 \ldots 4$. The appropriate within transformation to clear out the first k fixed effects is then

$$M_{D_e}^{(k)} = M_{D_e}^{(k-1)} - \left(M_{D_e}^{(k-1)}D_k^e\right)\left[\left(M_{D_e}^{(k-1)}D_k^e\right)'\left(M_{D_e}^{(k-1)}D_k^e\right)\right]^{-}\left(M_{D_e}^{(k-1)}D_k^e\right)'\,, \quad (30)$$

where the first step in the iteration is

$$M_{D_e}^{(1)} = I - D_1^e\left((D_1^e)'D_1^e\right)^{-1}(D_1^e)'\,,$$

[4] From a computational point of view, the calculation of matrix M_{D_d} is by far the most resource requiring as we have to invert $(NT \times NT)$ size matrices. Simplifications related to this can dramatically reduce CPU and storage requirements. This topic, however, is well beyond the scope of this paper.

and the iteration should be processed until $k = 4$. Note that none of this hinges on the model specification and can be done to any other multi-dimensional fixed effects model.

5. 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 first derive a general semi-asymptotic bias formulae, then we make it specific for each of the models introduced in Section 2; and lastly, we propose consistent estimators for the problematic models.

5.1 Nickell Biases

The models of Section 2 can all be written in the general dynamic form

$$y = \rho y_{-1} + D\pi + \varepsilon \,, \tag{31}$$

where D and π correspond to any of the specific D and π discussed in Section 2. With M_D, the projection matrix orthogonal to D,

$$\hat{\rho} = \frac{y_{-1}' M_D y}{y_{-1}' M_D y_{-1}} = \rho + \frac{\mathrm{tr}(M_D \varepsilon y_{-1}')}{\mathrm{tr}(M_D y_{-1} y_{-1}')} \,, \tag{32}$$

where y and y_{-1} are the column vectors of dependent and lagged dependent variables, respectively, of size $N^2 T$. Let

$$B_0 = \begin{pmatrix} 0 & 0 \\ I_{T-1} & 0 \end{pmatrix} \quad \text{of size} \quad (T \times T) \,,$$

$$\Gamma_0 \equiv (I_T - \rho B_0)^{-1} = \begin{pmatrix} 1 & \dots & \dots & 0 \\ \rho & \ddots & & \vdots \\ & \ddots & \ddots & \vdots \\ \rho^{T-1} & & \rho & 1 \end{pmatrix} \quad \text{of size} \quad (T \times T) \,,$$

$$\Psi_0 = \begin{pmatrix} 1 & \rho & & \rho^{T-1} \\ \rho & \ddots & \ddots & \\ & \ddots & \ddots & \rho \\ \rho^{T-1} & & \rho & 1 \end{pmatrix} = I_T + \rho(\Gamma_0 B_0 + (\Gamma_0 B_0)') \quad \text{of size} \quad (T \times T) \,,$$

and let $B = I_{N^2} \otimes B_0, \Gamma = I_{N^2} \otimes \Gamma_0, \Psi = I_{N^2} \otimes \Psi_0$ define matrices necessary for the general bias formulae. With e_1, the first unit vector of size $(T \times 1)$, and y_0 having N^2 elements [the initial values of the y_{ijt} for all (ij) pair],

$$By = y_{-1} - y_0 \otimes e_1 \,.$$

Therefore, model (31) can be rewritten as

$$y = \rho B y + \rho y_0 \otimes e_1 + D\pi + \varepsilon \,, \quad \text{or} \quad (I_{N^2T} - \rho B)y = \rho y_0 \otimes e_1 + D\pi + \varepsilon \,,$$

which ultimately leads to

$$y = \rho\Gamma(y_0 \otimes e_1) + \Gamma D\pi + \Gamma\epsilon \,.$$

Let ε_+ be ε advanced by one time period. Then, under the stationarity of ε_{ijt},

$$\mathrm{E}(y_{-1}\varepsilon') = \mathrm{E}(y\varepsilon'_+) = \Gamma\mathrm{E}(\varepsilon\varepsilon'_+) = \sigma^2_\varepsilon \Gamma B \,.$$

So, for the expectation of the numerator in (32), we obtain

$$\mathrm{E}(\mathrm{tr}(M_D \varepsilon y'_{-1})) = \sigma^2_\varepsilon \mathrm{tr}(M_D \Gamma B) = \frac{\sigma^2_\varepsilon}{2\rho}(\mathrm{tr}(M_D\Psi) - \mathrm{tr}(M_D)) \,,$$

with $\Psi = (I_{N^2T} + \rho(\Gamma B + (\Gamma B)'))$. For the denominator in (32),

$$\begin{aligned} \mathrm{E}(\mathrm{tr}(M_D y_{-1} y'_{-1})) &= \mathrm{E}(\mathrm{tr}(M_D y y')) \\ &= \rho^2 \mathrm{E}(\mathrm{tr}(M_D y_{-1} y'_{-1})) + \sigma^2_\varepsilon tr(M_D) + 2\mathrm{E}(\mathrm{tr}(M_D \varepsilon y'_{-1})) \,, \end{aligned}$$

so, as $\mathrm{E}(\mathrm{tr}(M_D \varepsilon y'_{-1})) = \sigma^2_\varepsilon tr(M_D \Gamma B)$,

$$\mathrm{E}(\mathrm{tr}(M_D y_{-1} y'_{-1})) = \frac{\sigma^2_\varepsilon}{1-\rho^2}(\mathrm{tr}(M_D) + 2\mathrm{tr}(M_D \Gamma B)) = \frac{\sigma^2_\varepsilon}{1-\rho^2}\mathrm{tr}(M_D \Psi) \,.$$

Combining the expressions for the numerator and denominator, we get

$$\operatorname*{plim}_{N\to\infty} \hat{\rho} = \rho + \frac{1-\rho^2}{2\rho}\left(1 - \operatorname*{plim}_{N\to\infty} \frac{\mathrm{tr}(M_D)}{\mathrm{tr}(M_D\Psi)}\right) \,. \tag{33}$$

As for the specific models, given that $\mathrm{tr}(I_T\Psi_0) = \mathrm{tr}(I_T)$, the traces are summarized in Table 2.

Table 2: Traces for the models considered

Model	$\mathrm{tr}(M_D)$	$\mathrm{tr}(M_D\Psi)$
(2)	$(N-1)(NT+T-2)$	$(N-1)(NT+T-2\theta)$
(5)	$N^2(T-1)$	$N^2(T-\theta)$
(7)	$(N^2-1)(T-1)$	$(N^2-1)(T-\theta)$
(8)	$NT(N-1)$	$NT(N-1)$
(10)	$(N-1)^2T$	$(N-1)^2T$
(12)	$(N-1)^2(T-1)$	$(N-1)^2(T-\theta)$

with

$$\theta = \mathrm{tr}(\bar{J}_T\Psi_0) = 1 + 2\frac{\rho}{1-\rho}\left(1 - \frac{1}{T}\frac{1-\rho^T}{1-\rho}\right).$$

Therefore, following (33), the individual asymptotic biases are given in Table 3.

Table 3: Asymptotic biases for the models considered

Model	$\underset{N\to\infty}{\mathrm{plim}}\ \hat{\rho} - \rho$
(2)	$\frac{1-\rho^2}{2\rho}\left(1 - \underset{N\to\infty}{\mathrm{plim}}\frac{NT+T-2}{NT+T-2\theta}\right) = 0$
(5), (7), (12)	$\frac{1-\rho^2}{2\rho}\left(1 - \frac{T-1}{T-\theta}\right)$
(8), (10)	0

5.2 Arellano–Bond Estimation

As seen above, we have problems with the N inconsistency of models (5), (7) and (12) in the dynamic case. 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 estimators: the Arellano and Bond IV estimator [see Arellano and Bond (1991) and Harris et al. (2008), p. 260] for the estimation of model (5).

The model is written up in first differences, such as

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

or

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

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

$$z_{ij} = \begin{pmatrix} y_{ij1} & 0 & \cdots & \cdots & 0 & \cdots & 0 \\ 0 & y_{ij1} & y_{ij2} & 0 & 0 & \cdots & 0 \\ \vdots & \cdots & & \vdots & \vdots & \cdots & \vdots \\ 0 & \cdots & 0 & 0 & y_{ij1} & \cdots & y_{ijT-2} \end{pmatrix}$$

of size $((T-2) \times (T-1)(T-2)/2)$. 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

$$\Sigma = \begin{pmatrix} 2 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & -1 & 2 & -1 \\ 0 & \cdots & 0 & -1 & 2 \end{pmatrix} \quad \text{of size } ((T-2) \times (T-2)) \,.$$

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

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

$$\Delta y_{ijt} = \rho \Delta y_{ijt-1} + \Delta \lambda_t + \Delta \varepsilon_{ijt} \,.$$

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

$$\Delta \tilde{y}_{ijt} = \left(\Delta y_{ijt} - \frac{1}{N} \sum_{i=1}^{N} \Delta y_{ijt} \right) \,,$$

or introduce the notation $\Delta\bar{y}_{.jt} = \frac{1}{N}\sum_{i=1}^{N}\Delta y_{ijt}$. We also notice that the λ-s had been eliminated from the model:

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

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

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

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

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

$$\left(\Delta y_{ijt} - \frac{1}{N}\sum_{i=1}^{N}\Delta y_{ijt} - \frac{1}{N}\sum_{j=1}^{N}\Delta y_{ijt} + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\Delta y_{ijt}\right)\,.$$

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

$$\begin{aligned}(\Delta y_{ijt} - \Delta\bar{y}_{.jt} - \Delta\bar{y}_{i.t} + \Delta\bar{y}_{..t}) &= \\ &= \rho(\Delta y_{ijt-1} - \Delta\bar{y}_{.jt-1} - \Delta\bar{y}_{i.t-1} + \Delta\bar{y}_{..t-1}) + \left(\Delta\varepsilon_{ijt} - \Delta\bar{\varepsilon}_{.jt} - \Delta\bar{\varepsilon}_{i.t} + \Delta\bar{\varepsilon}_{..t}\right).\end{aligned}$$

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

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 models increase dramatically. As a consequence, even when the number of individuals and time periods is not too large, the OLS 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 seen in the usual two dimensional panel data models, they are quite useful in these higher dimensional setups. However, unlike in two dimensions, they are biased in the case of some very relevant data problems, such as the lack of self-flows or unbalanced observations. These properties must be taken into account by all researchers relying on these methods. Also, in dynamic models, for some, but not all fixed effects formulations, the Within estimators are biased and inconsistent. Therefore, appropriate estimation methods need to be derived to deal with these cases.

References

Arellano, M., Bond, S. (1991). Some Tests of Specification for Panel Data: Monte-Carlo Evidence and an Application to Employment Equations. *Review of Economic Studies*, 58: 277-297.

Baier, S. L., Bergstrand, J. H. (2007). Do Free Trade Agreements Actually Increase Members International Trade? *Journal of International Economics*, 71: 72-95.

Balestra, P., Krishnakumar, J. (2008). Fixed Effects and Fixed Coefficients Models. In: Matyas and Sevestre (eds.), *The Econometrics of Panel Data.* 3rd edition, Springer Verlag, 23-48.

Baltagi, B. H., Egger, P., Pfaffermayr, M. (2003). A generalized Design for Bilateral Trade Flow Models. *Economics Letters*, 80: 391-397.

Baltagi, B. H. (2005). *Econometric Analysis of Panel Data.* 3rd edition, John Wiley & Sons.

Baldwin, R., Taglioni, D. (2006). Gravity for Dummies and Dummies for the Gravity Equations, NBER Working Paper 12516.

Cheng, I., Wall, H.J. (2005). Controlling for Heterogeneity in Gravity Models of Trade and Integration. *Federal Reserve Bank of St. Louis Review*, 87(1): 49-63.

Egger, P., Pfaffermayr, M. (2003). The Proper Econometric Specification of the Gravity Equation: 3-Way Model with Bilateral Interaction Effects. *Empirical Economics*, 28: 571-580.

Gourieroux, C., Monfort, A. (1995). *Statistics and Econometric Models.* Cambridge University Press.

Greene, W. H. (2012). *Econometric Analysis.* 7th edition, Pearson Education Ltd, 73-74.

Harris, M. N., Matyas, L., Sevestre, P. (2008). Dynamic Models for Short Panels. In: Matyas and Sevestre (eds.) *The Econometrics of Panel Data.* 3rd edition, Springer Verlag.

Matyas, L. (1997). Proper Econometric Specification of the Gravity Model. *The World Economy*, 20: 363-369.

Nickell, S. (1981). Biases in Dynamic Models with Fixed Effects. *Econometrica*, 49: 1417-1426.

Scheffé, H. (1959). *The Analysis of Variance.* John Wiley & Sons.

Wansbeek, T., Kapteyn, A. (1989). Estimation of Error-Components Model With Incomplete Panels. *Journal of Econometrics*, 41: 341-361.

Wansbeek, T. (1991). Singular value decomposition of design matrices in balanced ANOVA models. *Statistics & Probability Letters*, 11: 33-36.