GMM Estimation of Short Dynamic Panel Data Models With Error Cross Section Dependence

Vasilis Sarafidis*

This version: March 2008

Abstract

This paper considers the issue of GMM estimation of a short dynamic panel data model when the errors are correlated across individuals. We focus particularly on the conditions required in the cross-sectional dimension of the error process for the dynamic panel GMM estimator to remain consistent. To this end, we demonstrate that cross section independence (or uncorrelatedness) is not necessary − rather, it suffices that, if there is such correlation in the errors, this is weak. We define a stochastic scalar sequence to be weakly correlated at any given point in time if random variables sufficiently far apart in the sequence exhibit very little correlation. Spatial dependence satisfies this condition but factor structure dependence does not. Consequently, the dynamic panel GMM estimator is consistent only in the first case. Under weakly correlated errors, an additional set of moment conditions becomes relevant for each i − specifically, instruments with respect to the individual which unit i is correlated with, denoted by j. We demonstrate that these extra moment conditions can be particularly useful when the errors are subject to both weak and strong correlations, a situation that is likely to arise in practice. Simulated experiments show that the resulting method of moments estimators largely outperform the conventional ones in terms of both bias and RMSE.

Key Words: dynamic panel data, spatial dependence, factor structure dependence, Generalised Method of Moments.
JEL Classification: C13; C31; C33.

1 Introduction

In developing the theory of GMM estimation of short dynamic panel data models, it is commonly assumed that the regression errors are independently distributed across individuals (see e.g. Anderson and Hsiao, 1981, pg. 598, Arellano and Bond, 1991, pg. 278, Arellano, 1993, pg. 88, Ahn and Schmidt, 1995, pg. 7, Blundell and Bond, 1998, page 118, and others). This assumption is usually made for identification purposes

*Discipline of Econometrics and Business Statistics, University of Sydney, NSW 2006, Australia. Tel: +61-2-9036 9120; e-mail: v.sarafidis@econ.usyd.edu.au.

rather than descriptive accuracy with the hope, presumably, that by conditioning on a sufficient number of explanatory variables, what is left over can be treated as a purely idiosyncratic disturbance that is uncorrelated across individuals. On the other hand, in empirical applications of GMM estimation this rather strong assumption is somewhat relaxed by allowing for common variations in the dependent variable at any given point in time using two-way error components disturbances (e.g. Arellano and Bond, 1991, pg. 288, Blundell and Bond, 1998, pg. 137, Bover and Watson, 2005, pg. 1975). In practice, however, a $\alpha_i + f_t + \varepsilon_{it}$ formulation is unlikely to be adequate to remove all correlated behaviour in the errors and this may result in misleading inferences and even inconsistent GMM estimators (Sarafidis and Robertson, 2007)[1].

Error cross section dependence may arise for various reasons in practice; for example, it may be due to the presence of spatial correlations specified on the basis of economic and social distance (Conley, 1999) or relative location (Anselin, 1988), as well as due to the presence of unobserved components that give rise to a common factor specification in the disturbances with a fixed number of factors (e.g. Goldberger, 1972, and Jöreskog and Goldberger, 1975). Methods that account for a multi-factor error structure have been proposed by Robertson and Symons (2000), Coakley, Fuertes and Smith (2002), Phillips and Sul (2003), Moon and Perron (2004), Bai (2005), Pesaran (2006) and others. However, these methods are theoretically justified in panels where the number of time series observations (T) is large. To the best of our knowledge, no study exists that accounts for spatial correlations in a short dynamic panel data model.

The present paper deals specifically with the issue of GMM estimation of a short dynamic panel data model when the errors are not independent across individuals. A major focus lies on the conditions required in the cross-sectional dimension of the error process for the dynamic panel GMM estimator to remain consistent. To this end, we demonstrate that independence, or uncorrelatedness, is not necessary for GMM consistency or asymptotic efficiency − rather, it is sufficient that, if there is such correlation in the errors, this is weak. We define a stochastic scalar sequence to be weakly correlated at any given point in time if random variables sufficiently far apart in the sequence exhibit very little correlation. Therefore, a weakly correlated sequence is asymptotically uncorrelated. Conversely, a sequence is strongly correlated if random variables remain correlated no matter how far apart the lie in the sequence. We show that the spatial approach to modelling error cross section dependence, which typically assumes uniform boundedness of the row and column sums of the weighting matrix, satisfies asymptotic uncorrelatedness, although it is more restrictive in the sense that the latter does not require uniform boundedness. On the other hand, under factor structure dependence the errors are strongly correlated and therefore the GMM estimator is not consistent. The two-way error components model violates asymptotic uncorrelatedness too[2], albeit the problem can be dealt in this case via time-demeaning of the observations. However, careful analysis needs to be made in this case because the aforementioned transformation

[1] In an influential paper, Phillips and Sul (2007) analyse the impact of error cross section dependence on the dynamic Fixed Effects (FE) estimator.

[2] Assuming that the unobserved time-specific individual-invariant effect is treated as stochastic.

induces some dependency among the N individual equations and therefore the moment conditions are not valid anymore for finite N, a result that is usually ignored in the literature.

In addition, this paper shows that when the errors are weakly correlated in the way defined above, then for each individual i there is an additional set of moment conditions that becomes relevant − in particular, instruments with respect to the individual which unit i is correlated with, denoted by j. We demonstrate that these extra moment conditions can be particularly useful when the errors are subject to both weak and strong correlations, a situation that is likely to arise in practice. Pesaran and Tosetti (2007) consider this situation as well, for a model with no lags of the dependent variable on the right-hand side and T sufficiently large.

The structure of the paper is as follows. The following section specifies the panel regression model in a way that encompasses common factors and spatial dependence. Section 3 reviews the standard moment conditions used in GMM estimation under two-way error components disturbances. Section 4 addresses the issue of consistency for the dynamic panel GMM estimator when the independence assumption across individuals is relaxed. Section 5 shows that under weakly correlated errors, additional moment conditions become relevant for each individual i, which arise from the individual(s) which unit i is correlated with. Section 6 demonstrates the validity of these extra moment conditions under both weakly and strongly correlated errors and the following section analyses the properties of the resulting GMM estimators, including cases where the problem of weak instruments applies. The performance of these estimators is investigated in Section 8 using simulated data. A final section concludes.

2 Model Specification

We focus on dynamic panel data models of the following first-order autoregressive form:

$$\begin{aligned} y_{i,t} &= \lambda y_{i,t-1} + \upsilon_{i,t}, \quad i = 1, ..., N \text{ and } t = 2, ..., T \\ \upsilon_{i,t} &= \alpha_i + u_{i,t} \\ u_{i,t} &= \sum_{m=1}^{M} \theta_i^m \sum_{j=1}^{N} w_{i,j}^m \xi_{j,t}^m + \varepsilon_{i,t} = \boldsymbol{\theta}_i' \left[diag\left(\mathbf{W}_i \times \boldsymbol{\xi}_t'\right) \times \mathbf{i}_M \right] + \varepsilon_{i,t} \end{aligned} \tag{1}$$

where $y_{i,t}$ is the dependent variable of individual i at time t, λ is a fixed parameter to be estimated with $|\lambda| < 1$, and υ_{it} is a composite error term that consists of an individual-specific time-invariant unobserved effect and a weighted sum of purely idiosyncratic components, where $\boldsymbol{\theta}_i = \left(\theta_i^1, ..., \theta_i^M\right)'$ is an $M \times 1$ vector,

$$\mathbf{W}_i = \begin{bmatrix} w_{i,1}^1 & w_{i,2}^1 & \cdots & w_{i,N}^1 \\ w_{i,1}^2 & w_{i,2}^2 & \cdots & w_{i,N}^2 \\ . & . & . & . \\ w_{i,1}^M & w_{i,2}^M & \cdots & w_{i,N}^M \end{bmatrix}; M \times N, \quad \boldsymbol{\xi}_t = \begin{bmatrix} \xi_{1,t}^1 & \xi_{2,t}^1 & \cdots & \xi_{N,t}^1 \\ \xi_{1,t}^2 & \xi_{2,t}^2 & \cdots & \xi_{N,t}^2 \\ . & . & . & . \\ \xi_{1,t}^M & \xi_{2,t}^M & \cdots & \xi_{N,t}^M \end{bmatrix}; M \times N, \tag{2}$$

and $\mathbf{i}_M$ is a $M \times 1$ column vector of ones.

We make the following assumptions:

Assumption 1: $\alpha_i \sim iid\left(0, \sigma_\alpha^2\right)$.

Assumption 2: $\xi_{it}^m \sim iid\left(0, \sigma_{\xi_m}^2\right)$ and $\varepsilon_{it} \sim iid\left(0, \sigma_\varepsilon^2\right)$.

Assumption 3: $E\left(y_{i,1}\varepsilon_{i,t}\right) = 0$, for $i = 1, ..., N$ and $t = 2, 3, ..., T$.

Assumption 4: θ_i^m is non-stochastic and bounded with $\lim_{N\to\infty}\left[\frac{1}{N}\sum_{i=1}^N \boldsymbol{\theta}_i\right] = \boldsymbol{\mu}_\theta \neq 0$, $\lim_{N\to\infty}\left[\frac{1}{N}\sum_{i=1}^N \underline{\boldsymbol{\theta}}_i\underline{\boldsymbol{\theta}}_j'\right] = \boldsymbol{\Sigma}_\theta$ for $i = j$ and $\mathbf{0}$ otherwise, where $\underline{\boldsymbol{\theta}}_i = \boldsymbol{\theta}_i - \overline{\boldsymbol{\theta}}$, $\boldsymbol{\Sigma}_\theta$ is a diagonal positive semi-definite matrix and $\boldsymbol{\mu}_\theta$ is a $M \times 1$ vector such that $||\boldsymbol{\mu}_\theta|| < B_\theta < \infty$.

Assumptions 1-3 are standard in the GMM literature. Assumption 2 can be easily relaxed by allowing $\varepsilon_{it} \sim MA(k)$, where k is a small positive integer. Assumption 3 ensures that sufficiently lagged values of y_{it} will be uncorrelated with the first-difference of ε_{it} and thus they will be available as instruments. Assumption 4 is equivalent to requiring that $\boldsymbol{\theta}_i$ has finite mean and variance and it is uncorrelated across i for all m if $\boldsymbol{\theta}_i$ were stochastic, which would be satisfied under $-$ say $-$ $\boldsymbol{\theta}_i \sim iid\left(\boldsymbol{\mu}_\theta, \boldsymbol{\Sigma}_\theta\right)$.

Note that all the results discussed below extend in an obvious fashion to higher order autoregressive processes as well as to panel autoregressive distributed lag models. Model (1) can be written in a more compact form as follows:

$$\mathbf{y} = \lambda\mathbf{y}_{-1} + \boldsymbol{\alpha} + \mathbf{u}, \quad \mathbf{u} = \sum_{m=1}^M \boldsymbol{\theta}^m \left(\mathbf{W}_N^m \otimes I_{T-1}\right)\boldsymbol{\xi}^m + \boldsymbol{\varepsilon} \tag{3}$$

where $\mathbf{y} = (\mathbf{y}_1, ..., \mathbf{y}_N)'$ is a $N \times (T-1)$ matrix with $\mathbf{y}_i = (y_{i,2}, ..., y_{i,T})'$, $\mathbf{y}_{-1} = (\mathbf{y}_{1,-1}, ..., \mathbf{y}_{N,-1})'$ is a $N \times (T-1)$ matrix with $\mathbf{y}_{i,-1} = (y_{i,1}, ..., y_{i,T-1})'$, $\boldsymbol{\alpha} = \left[(\alpha_1, ..., \alpha_N)' \otimes \mathbf{i}\right]$ with $\mathbf{i}_{T-1}$ being a $(T-1) \times 1$ column vector of ones, $\boldsymbol{\theta}^m = diag\left[(\theta_1^m, ..., \theta_N^m)' \otimes \mathbf{i}_{T-1}\right]$ is a $N(T-1) \times 1$ vector, $\mathbf{W}_N^m$ is a $N \times N$ weighting matrix, $\boldsymbol{\xi}^m = (\boldsymbol{\xi}_1^m, ..., \boldsymbol{\xi}_N^m)'$ with $\boldsymbol{\xi}_i^m = \left(\xi_{i,2}^m, ..., \xi_{i,T}^m\right)'$ and $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1, ..., \boldsymbol{\varepsilon}_N)'$ with $\boldsymbol{\varepsilon}_i = (\varepsilon_{i2}, ..., \varepsilon_{i,T})'$.

The composite error term, $\mathbf{u}$, has a flexible structure in that it can characterise various forms of cross section dependence, which include dependence that is due to the presence of unobserved common factors as well as spatial correlations in the error term, depending on the structure of $\mathbf{W}_N^m$. Specifically, the multi-factor structure arises from (3) by setting all elements of $\mathbf{W}_N^m$, denoted by $w_{i,j}^m$, equal to

$$w_{i,j}^m = N^{-1/2} \text{ for } i = 1, ..., N, \ m = 1, ..., M \tag{4}$$

such that

$$\mathbf{u} = \sum_{m=1}^M \boldsymbol{\theta}^m \mathbf{f}^m + \boldsymbol{\varepsilon}, \text{ where } \mathbf{f}^m = \left[\mathbf{i} \otimes (f_2^m, ..., f_T^m)'\right] \text{ and } f_t^m = N^{-1/2}\sum_{i=1}^N \xi_{i,t}^m. \tag{5}$$

In this case we have $E\left(f_t^m\right)=0$, $var\left(f_t^m\right)=\sigma_{\xi_m}^2$ and $cov\left(f_t^m, f_{t-\tau}^m\right)=0$ for $\tau>0$ and all m. Therefore, $E\left(u_{i,t}\right)=0$, $var\left(u_{i,t}\right)=\sum_{m=1}^M\left(\theta_i^m\right)^2\sigma_{\xi_m}^2+\sigma_{\varepsilon}^2$ and

$$cov\left(u_{i,t}, u_{i+k,s}\right)=\begin{cases}\sum_{m=1}^M\theta_i^m\theta_{i+k}^m\sigma_{\xi_m}^2 & \text{for } t=s\\ 0 & \text{otherwise}\end{cases} \tag{6}$$

The Spatial Moving Average (SMA) process arises from (3) by setting $M=1$, $\theta_i^1=\theta$ with $|\theta|<1$, $\boldsymbol{\xi}^1=\boldsymbol{\varepsilon}$ and $\mathbf{W}_N^1=\mathbf{W}_N$ equal to a sparse matrix populated primarily with zeros. For instance in a circular[3] SMA(1) process, $\mathbf{u}$ equals

$$\mathbf{u}=\theta\left(\mathbf{W}_N\otimes I_{T-1}\right)\boldsymbol{\varepsilon}+\boldsymbol{\varepsilon} \tag{7}$$

with $\mathbf{W}_N$ given by

$$\mathbf{W}_N=\begin{bmatrix}0&1&0&.&.&.&.&.&0\\0&0&1&0&.&.&.&.&0\\0&0&0&1&0&.&.&.&0\\.&.&.&.&.&.&.&.&.\\0&.&.&.&.&0&0&0&1\\1&.&.&.&.&.&0&0&0\end{bmatrix} \tag{8}$$

In this case we have $E\left(u_{i,t}\right)=0$, $var\left(u_{i,t}\right)=\sigma_{\varepsilon}^2\left(1+\theta^2\right)$ and

$$cov\left(u_{i,t}, u_{i+k,s}\right)=\begin{cases}\theta\sigma_{\varepsilon}^2 & \text{for } \kappa=i\,(\operatorname{mod}N)+1 \text{ and } t=s\\ 0 & \text{otherwise}\end{cases} \tag{9}$$

where $i\,(\operatorname{mod}N)$ is the modulo operator, defined as the remainder after numerical division of i by N to obtain integer values. Thus, for $i=1,...,N-1$, $i\,(\operatorname{mod}N)+1=i+1$ and for $i=N$, $N\,(\operatorname{mod}N)+1=1$. SMA processes of higher order can be accomodated straightforwardly. Assuming invertibility, the Spatial Autoregressive (SAR) form can be obtained using an infinite SMA representation[4].

The Spatial Error Components (SEC) form arises in a way similar to a SMA form with the only difference being that $\boldsymbol{\xi}^1\neq\boldsymbol{\varepsilon}$ in (7) while (8) includes non-zero values on the main diagonal. In this case we have $E\left(u_{i,t}\right)=0$, $var\left(u_{i,t}\right)=\sigma_{\varepsilon}^2+\theta^2\sigma_{\xi_1}^2$ and

$$cov\left(u_{i,t}, u_{i+k,s}\right)=\begin{cases}\theta\sigma_{\xi_1}^2 & \text{for } \kappa=i\,(\operatorname{mod}N)+1 \text{ and } t=s\\ 0 & \text{otherwise}\end{cases} \tag{10}$$

Finally, it follows that by imposing appropriate restrictions on θ_i^m, $w_{i,j}^m$ and $\xi_{i,t}^m$, (3) can easily accomodate mixture cases too, where both spatial correlations and common unobserved factors are present in the errors. We consider estimation of this type of models in Section 6.

[3] See e.g. Baltagi, Bresson and Pirotte (2007).

[4] In this case, $\mathbf{W}_N$ is not sparse, however its elements will decline with a distance measure that increases sufficiently rapidly as the sample increases. For instance, Stetzer (1982) models the distance decay by a negative exponential function, $w_{ij}=\exp\left(-\theta d_{ij}\right)$, $0<\delta<\infty$, with d_{ij} denoting the distance between individuals i and j.

3 Moment Conditions in Standard GMM Estimation

Typical GMM estimation of linear dynamic panel data models of the form given in (1) imposes $\theta_i^m = 0$ for all i and m, such that any form of dependence in the error process across individuals, whether this is spatial or subject to a factor structure, is ruled out[5]. Consequently, applying first-differences in (1) yields

$$\Delta y_{i,t} = \lambda \Delta y_{i,t-1} + \Delta \upsilon_{i,t}, \quad i = 1, ..., N \text{ and } t = 3, ..., T \tag{11}$$

Under Assumptions 1-3 the following $\zeta = (T-1)(T-2)/2$ moment conditions become available

$$E\left(y_{i,t-s} \Delta \upsilon_{i,t}\right) = E\left(y_{i,t-s} \Delta \varepsilon_{i,t}\right) = 0; \text{ for } t = 3, ..., T \text{ and } s = 2, ..., t-1. \tag{12}$$

On the other hand, in empirical applications it is common practice to generalise the error structure by allowing for common variations in the dependent variable using a two-way error components formulation[6]:

$$\upsilon_{i,t} = \alpha_i + f_t + \varepsilon_{i,t}. \tag{13}$$

If f_t is treated as non-stochastic, the moment conditions given in (12) are not valid anymore because the expectation of f_t is equal to f_t itself. Hence

$$\begin{aligned} E\left(y_{i,t-s} \Delta \upsilon_{i,t}\right) &= E\left[\left(\frac{\alpha_i}{1-\lambda} + \sum_{\tau=0}^{\infty} \lambda^\tau \varepsilon_{i,t-s-\tau} + \sum_{\tau=0}^{\infty} \lambda^\tau f_{t-s-\tau}\right)\left(\Delta f_t + \Delta \varepsilon_{i,t}\right)\right] \\ &= \sum_{\tau=0}^{\infty} \lambda^\tau f_{t-s-\tau} \Delta f_t \neq 0. \end{aligned} \tag{14}$$

If, instead, f_t is treated as stochastic and serially uncorrelated, the moment conditions given above remain valid because $E\left[\sum_{\tau=0}^{\infty} \lambda^\tau f_{t-s-\tau} \Delta f_t\right] = 0$. However, the sample counterpart of (14) does not converge to its expectation for finite T for reasons that will become clear in the next section. Instead,

$$\frac{1}{N} \sum_{i=1}^{N} \left[y_{i,t-s} \Delta \upsilon_{i,t}\right] - \sum_{\tau=0}^{\infty} \lambda^\tau f_{t-s-\tau} \Delta f_t \xrightarrow{p} 0 \tag{15}$$

Transforming the observations in terms of deviations from time-specific averages eliminates both of these problems by removing the common time effect from the regression error:

$$\underline{\upsilon}_{it} = \upsilon_{it} - \overline{\upsilon}_t = (\alpha_i - \overline{\alpha}) + (f_t - f_t) + (\varepsilon_{it} - \overline{\varepsilon}_t) = \underline{\alpha}_i + \underline{\varepsilon}_{it}. \tag{16}$$

However, the cross-sectionally demeaned transformation induces some dependency among the N individual equations and therefore the moment conditions on the transformed observations are not valid for finite N, i.e. $E\left(\underline{y}_{i,t-s} \Delta \underline{\upsilon}_{i,t}\right) = E\left(\underline{y}_{i,t-s} \Delta \underline{\varepsilon}_{i,t}\right) \neq 0$,

[5] See Section 1 for related references.

[6] Viz. footnote 5.

where $\underline{y}_{i,t-s} = y_{i,t-s} - \overline{y}_{t-s}$, and similarly for the remaining variables. As N grows large, this dependency disappears; in particular, defining $\underline{y}^o_{i,t-s} = y_{i,t-s} - \widetilde{\mu}_{y_{t-s}}$ and $\Delta\underline{\upsilon}^o_{it} = \Delta\left[\upsilon_{i,t} - \widetilde{\mu}_{\upsilon_t}\right]$, where $\widetilde{\mu}_{y_{t-s}} = \sum_{\tau=0}^{\infty}\lambda^\tau f_{t-s-\tau}$ and $\widetilde{\mu}_{\upsilon_t} = f_t$, we have

$$\begin{aligned}
&\sqrt{N}\frac{1}{N}\sum_{i=1}^{N}\left[\underline{y}_{i,t-s}\Delta\underline{\upsilon}_{i,t}\right] \\
= \;& \sqrt{N}\frac{1}{N}\sum_{i=1}^{N}\left[\left(y_{i,t-s} - \widetilde{\mu}_{y_{t-s}}\right) - \left(\overline{y}_{t-s} - \widetilde{\mu}_{y_{t-s}}\right)\right]\Delta\left[\left(\upsilon_{i,t} - \widetilde{\mu}_{\upsilon_t}\right) - \left(\overline{\upsilon}_t - \widetilde{\mu}_{\upsilon_t}\right)\right] \\
= \;& \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[\underline{y}^o_{i,t-s}\Delta\underline{\upsilon}^o_{i,t}\right] + o_p(1) \qquad (17)
\end{aligned}$$

since $\Delta\left(\overline{\upsilon}_t - \widetilde{\mu}_{\upsilon_t}\right) = O_p\left(N^{-1/2}\right)$, $\overline{y}_{t-s} - \widetilde{\mu}_{y_{t-s}} = O_p\left(N^{-1/2}\right)$, $N^{-1/2}\sum_{i=1}^{N}\underline{y}^o_{i,t-s} = O_p(1)$ and $N^{-1/2}\sum_{i=1}^{N}\Delta\underline{\upsilon}^o_{it} = O_p(1)$. Therefore, given that $\underline{y}^o_{i,t-s}\Delta\underline{\upsilon}^o_{i,t}$ are independent across i, a suitable CLT (Central Limit Theorem) ensures that

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[\underline{y}^o_{i,t-s}\Delta\underline{\upsilon}^o_{i,t}\right] \xrightarrow{d} N\left(0, Var\left(\frac{1}{\sqrt{N}}\underline{y}^o_{i,t-s}\Delta\underline{\upsilon}^o_{i,t}\right)\right) \qquad (18)$$

Hence, defining

$$\underline{Z}_i = \begin{bmatrix} \underline{y}_{i,1} & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \underline{y}_{i,1} & \underline{y}_{i,2} & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & \underline{y}_{i,1} & \underline{y}_{i,2} & \cdots & \underline{y}_{i,T-2} \end{bmatrix};\ \Delta\underline{\upsilon}_i = \begin{bmatrix} \Delta\underline{\upsilon}_{i,3} \\ \Delta\underline{\upsilon}_{i,4} \\ \vdots \\ \Delta\underline{\upsilon}_{i,T} \end{bmatrix}, \qquad (19)$$

and $\underline{\mathbf{Z}} = \left(\underline{\mathbf{Z}}_1, \underline{\mathbf{Z}}_2, ..., \underline{\mathbf{Z}}_N\right)'$, the first-differenced GMM estimator equals

$$\widehat{\lambda}_{DIF\ GMM} = \left[\frac{1}{N}\Delta\underline{\mathbf{y}}'_{-1}\underline{\mathbf{Z}}\widehat{\mathbf{A}}_N\underline{\mathbf{Z}}'\Delta\underline{\mathbf{y}}_{-1}\right]^{-1}\left[\frac{1}{N}\Delta\underline{\mathbf{y}}'_{-1}\underline{\mathbf{Z}}\widehat{\mathbf{A}}_N\underline{\mathbf{Z}}'\Delta\underline{\mathbf{y}}\right] \qquad (20)$$

with $\Delta\underline{\mathbf{y}} = \left(\Delta\mathbf{y}_1, ..., \Delta\mathbf{y}_N\right)'$, $\Delta\mathbf{y}_i = \left(\Delta y_{i,3}, ..., \Delta y_{i,T}\right)'$, $\Delta\underline{\mathbf{y}}_{,-1} = \left(\Delta\mathbf{y}_{1,-1}, ..., \Delta\mathbf{y}_{N,-1}\right)'$ and $\Delta\mathbf{y}_{i,-1} = \left(\Delta y_{i,2}, ..., \Delta y_{i,T-1}\right)'$. $\widehat{\mathbf{A}}_N$ is some weighting matrix that satisfies

$$\widehat{\mathbf{A}}_N - \mathbf{A}_N \xrightarrow{p} 0 \qquad (21)$$

where $\mathbf{A}_N$ is a non-stochastic sequence of positive definite matrices. Alternative choices of $\widehat{\mathbf{A}}_N$ lead to different GMM estimators, which are all consistent but they differ in terms of efficiency. The asymptotically efficient DIF GMM estimator sets $\widehat{\mathbf{A}}_N$ equal to the inverse of the covariance matrix of the moment conditions[7] – that is, $\widehat{\mathbf{A}}_N^{-1} =$

[7] See Hansen (1982).

$Est.Asy.Var\left(\sqrt{N}\frac{1}{N}\underline{\mathbf{Z}}'\boldsymbol{\Delta}\underline{\boldsymbol{\upsilon}}\right)$, assuming that this matrix exists and is finite positive definite. When ε_{it} is homoscedastic $\widehat{\mathbf{A}}_N^{-1}$ can be approximated by $\frac{1}{N}\underline{\mathbf{Z}}'\mathbf{H}\underline{\mathbf{Z}}$, where

$$\mathbf{H} = I_N \otimes H_i \tag{22}$$

and

$$H_i = \begin{bmatrix} 2 & -1 & \cdots & & 0 \\ -1 & 2 & \cdots & & 0 \\ \vdots & \vdots & & & \vdots \\ . & . & & & -1 \\ 0 & 0 & \cdots & -1 & 2 \end{bmatrix} \tag{23}$$

Hence it is clear that the weighting matrix is a function of the Kronecker product between two distinct matrices, the former of which reflects cross section dependence in the error structure (which, for large N, is zero in the present case and hence the use of the identity matrix) while the latter reflects time series dependence in the error structure, and in particular first-order serial correlation, which is induced by first-differencing the observations. Note that since the individual equations are independent across i for large N, $\widehat{\mathbf{A}}_N^{-1}$ can also be written as $\widehat{\mathbf{A}}_N^{-1} = \frac{1}{N}\sum_i \underline{Z}_i' H_i \underline{Z}_i$ and therefore an equivalent expression for (20) is given by

$$\begin{aligned} &\widehat{\lambda}_{DIF\ GMM} = \\ = \; &\left[\frac{1}{N}\left(\sum_{i=1}^{N}\Delta\underline{\mathbf{y}}_{i,-1}'\underline{\mathbf{Z}}_i\right)\widehat{\mathbf{A}}_N\left(\sum_{i=1}^{N}\underline{\mathbf{Z}}_i'\Delta\underline{\mathbf{y}}_{i,-1}\right)\right]^{-1}\left[\frac{1}{N}\left(\sum_{i=1}^{N}\Delta\underline{\mathbf{y}}_{i,-1}'\underline{\mathbf{Z}}_i\right)\widehat{\mathbf{A}}_N\left(\sum_{i=1}^{N}\underline{\mathbf{Z}}_i'\Delta\underline{\mathbf{y}}_i\right)\right] \end{aligned} \tag{24}$$

When the individual observations are not independent across i, (24) is not equivalent to (20).

The standard first-differenced GMM (DIF GMM) estimator may have poor finite sample properties in terms of bias and precision when $\lambda \to 1$ or $\sigma_\alpha^2/\sigma_\varepsilon^2 \to \infty$. As a result, Blundell and Bond (1998) developed an approach outlined in Arellano and Bover (1995), which combines the equations in first-differences with the equations in levels, using $\Delta y_{i,t-1}$ as an instrument for the lagged dependent variable, $y_{i,t-1}$:

$$E\left(\Delta y_{i,t-1}\upsilon_{i,t}\right) = 0; \text{ for } t = 3, 4, ..., T \tag{25}$$

This approach gives rise to a system GMM (SYS GMM) estimator, which is valid provided that the deviations of the initial observations from the long-run convergent values are uncorrelated with the individual-specific, time-invariant effects − that is,

$$E\left[\alpha_i\left(y_{i,1} - \frac{\alpha_i}{1-\lambda}\right)\right] = 0. \tag{26}$$

If common time effects are included in the error process, what is required is that

$$\text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\left[\underline{\alpha}_i\left(\underline{y}_{i,1} - \frac{\underline{\alpha}_i}{1-\lambda}\right)\right] = 0. \tag{27}$$

Thus, defining

$$\underline{Z}_i^{sys} = \begin{bmatrix} \underline{Z}_i & 0 & \cdots & 0 \\ 0 & \Delta \underline{y}_{i,2} & & 0 \\ . & . & \ddots & . \\ 0 & 0 & \cdots & \Delta \underline{y}_{i,T} \end{bmatrix}; \ \underline{\upsilon}_i^{sys} = \begin{bmatrix} \Delta \underline{\upsilon}_i \\ \underline{\upsilon}_i \end{bmatrix},$$

and $\underline{\mathbf{Z}}^{sys} = \left(\underline{Z}_1^{sys}, \underline{Z}_2^{sys}, ..., \underline{Z}_N^{sys}\right)'$, the SYS GMM estimator is given by

$$\widehat{\lambda}_{SYS\ GMM} = \left[\frac{1}{N}\underline{\mathbf{Y}}'_{-1}\underline{\mathbf{Z}}^{sys}\widehat{\mathbf{A}}_{\mathbf{N},\mathbf{sys}}\underline{\mathbf{Z}}^{sys\prime}\underline{\mathbf{Y}}_{-1}\right]^{-1}\left[\frac{1}{N}\underline{\mathbf{Y}}'_{-1}\underline{\mathbf{Z}}^{sys}\widehat{\mathbf{A}}_{\mathbf{N},\mathbf{sys}}\underline{\mathbf{Z}}^{sys\prime}\underline{\mathbf{Y}}\right] \tag{28}$$

where $\underline{\mathbf{Y}} = (\underline{\mathbf{Y}}_1, \underline{\mathbf{Y}}_2, ..., \underline{\mathbf{Y}}_N)'$, $\underline{\mathbf{Y}}_i = \left(\Delta \underline{y}_{i,3}, ..., \Delta \underline{y}_{i,T}, \underline{y}_{i,3}, ..., \underline{y}_{i,T}\right)$, $\underline{\mathbf{Y}}_{-1} = \left(\underline{\mathbf{Y}}_{1,-1}, \underline{\mathbf{Y}}_{2,-1}, ..., \underline{\mathbf{Y}}_{N,-1}\right)'$, and $\underline{\mathbf{Y}}_{i,-1} = \left(\Delta \underline{y}_{i,2}, ..., \Delta \underline{y}_{i,T-1}, \underline{y}_{i,2}, ..., \underline{y}_{i,T-1}\right)$. With homoscedastic errors the optimal choice of $\widehat{\mathbf{A}}_{\mathbf{N},\mathbf{sys}}$ is given by[8]

$$\widehat{\mathbf{A}}_{\mathbf{N},\mathbf{sys}} = \left[\frac{1}{N}\underline{\mathbf{Z}}^{sys\prime}\mathbf{H}^{sys}\underline{\mathbf{Z}}^{\mathbf{sys}}\right]^{-1} \tag{29}$$

where $\mathbf{H}^{sys}$ equals

$$\mathbf{H}^{sys} = \begin{bmatrix} \mathbf{H} & \mathbf{C} \\ \mathbf{C}' & \mathbf{I}_{N(T-2)} \end{bmatrix} \tag{30}$$

with $\mathbf{I}_{N(T-2)} = I_N \otimes I_{T-2}$ and $\mathbf{C} = I_N \otimes C_1$, where C_1 takes the value of 1 on the main diagonal, -1 on the first lower off-diagonal and zero otherwise.

The next section addresses the issue of consistency for the dynamic panel GMM estimator when the restriction $\theta_i^m = 0\ \forall\ i$, m is relaxed.

4 The Consistency of the Dynamic Panel GMM Estimator under Error Cross Section Dependence

When $\boldsymbol{\theta}^m$ is bounded and different from zero in (1), the structure of $\mathbf{W}_N^m$ will be critical upon the asymptotic properties of the GMM estimator. Without loss of generality, we will impose $M = 1$ for the remaining of this section, such that the error process becomes equal to $\upsilon_{i,t} = \alpha_i + \theta_i \sum_{j=1}^{N} w_{i,j}\xi_{j,t} + \varepsilon_{i,t}$ with $|\theta_i| < B_\theta < \infty$, $E(\upsilon_{i,t}) = 0$ and $E\left(\upsilon_{i,t}^2\right) < B_\upsilon < \infty$. We firstly define the concept of a weakly correlated process. Let $\left\{\upsilon_i^t,\ i \geq 1\right\}$ be the scalar sequence $(\upsilon_{1,t}, \upsilon_{2,t}, \upsilon_{3,t}, ...)$. There are $(T-1)$ such scalar sequences, for $t = 2, ..., T$.

[8] For $\sigma_\alpha^2 = 0$; see Windmeijer (2000) and Kiviet (2007).

Definition 1 *The scalar sequence* $\left\{\upsilon_i^t,\, i \geq 1\right\}$ *is said to be weakly correlated if there exist non-negative constants* $\left\{\rho_\kappa^{t,s}, \kappa \geq 0\right\}$, *where* $0 \leq \rho_\kappa^{t,s} \leq 1$ *and*

$$\rho_\kappa^{t,s} \geq E\left(\upsilon_{i,t}, \upsilon_{i+\kappa,s}\right) / \left[E\left(\upsilon_{i,t}^2\right) E\left(\upsilon_{i+\kappa,s}^2\right)\right]^{1/2} \text{ for all } \kappa \geq 0, \tag{31}$$

such that

$$\sum_{\kappa=0}^{\infty} \rho_\kappa^{t,s} < \infty \tag{32}$$

for all t *and* s.

Notice that $\rho_\kappa^{t,s}$ is merely an upper bound for the correlation between $\upsilon_{i,t}$ and $\upsilon_{i+\kappa,s}$, assuming that $E\left(\upsilon_{i,t}\right) = 0$ for $i \geq 1$ and all t, s. Since it is only positive correlation that matters, if $\upsilon_{i,t}$ and $\upsilon_{i+\kappa,s}$ are negatively correlated, we can set $\rho_\kappa^{t,s} = 0$. Thus, Definition 1 implies that random variables sufficiently far apart in the sequence exhibit very little correlation.

Remark 2 *Observe that for* $\sum_{\kappa=0}^{\infty} \rho_\kappa^{t,s} < \infty$, *it is necessary that* $\rho_\kappa^{t,s} = o(1)$ *and it is sufficient that* $\rho_\kappa^{t,s} = o\left(\kappa^{-1}\right)$. *Therefore, a weakly correlated process is asymptotically uncorrelated. Conversely, a sequence that is not asymptotically uncorrelated* $-$ *that is, where random variables remain correlated no matter how far apart they lie in the sequence, is said to be strongly correlated.*

Theorem 3 *Let* $\left\{\upsilon_i^t,\, i \geq 1\right\}$ *be the scalar sequence* $(\upsilon_{1,t}, \upsilon_{2,t}, \upsilon_{3,t}, ...)$, *where* $\upsilon_{i,t} = \alpha_i + \theta_i \sum_{j=1}^{N} w_{i,j} \xi_{j,t} + \varepsilon_{i,t}$, *with* $|\theta_i| < B_\theta < \infty$, $E\left(\upsilon_{i,t}^2\right) < B_\upsilon < \infty$ *and*

$$||\mathbf{W}||_\infty = \max_i \sum_{j=1}^{N} |w_{i,j}| = o\left(N^{1/2}\right) \tag{33}$$

Then $\left\{\upsilon_i^t\right\}$ *is weakly correlated, or asymptotically uncorrelated.*

Proof. See Appendix A. ■

Note that condition (33) in Theorem (3) is more general than a uniform boundedness condition for the row and column sums of $\mathbf{W}_N$ (typically employed in spatial models), which is stated as follows[9]:

$$\sum_{i=1}^{N} |w_{i,j}| \leq B_w < \infty \;\forall\, j \text{ and } \sum_{j=1}^{N} |w_{i,j}| \leq B_w < \infty \;\forall\, i \tag{34}$$

This is because uniform boundedness is subject to (33) but not vice versa. For instance, we may have $|w_{i,j}| = N^{-2/3}$ $\forall\, i,\, j$, in which case the row and column sums of $\mathbf{W}$ are not bounded because $\sum_{i=1}^{N} |w_{i,j}| = N^{1/3}$ and thefore it is growing with N.

[9] See e.g. Kapoor, Kelejian and Prucha (2007, pg. 106) and Lee (2007, pg. 491).

However, condition (33) is still satisfied. As a result, any spatially correlated process that satisfies (34) is weakly correlated, or asymptotically uncorrelated. On the other hand, the factor structure sets $|w_{i,j}| = N^{-1/2}$ $\forall$ i, j and so it violates (33). Hence it provides an example of a process that is not weakly correlated. As a matter of fact, when $|w_{i,j}| = N^{-1/2}$ there are M unobserved variables, $f_t^m = N^{-1/2} \left(\sum \xi_{1t}^m + ... + \xi_{Nt}^m\right)$, which are common for all i and therefore their effect does not diminish no matter how far in the sequence two random variables, $\upsilon_{i,t}$ and $\upsilon_{i+\kappa,t}$, are. As a result, the factor structure dependence is an example of a strongly correlated process. The two-way error components model with stochastic f_t is a restricted case of the single-factor structure because it sets $\theta_i = 1$ for all i although it retains the same form for $w_{i,j}$. Therefore, it provides another example of a strongly correlated process, albeit the correlation can be removed in this case for large N by transforming the data in terms of deviations from time-specific averages.

Notice that condition (33) does not imply that the $\left\{\upsilon_i^t, \, i \geq 1\right\}$ sequence is spatially ergodic because the row sums of $\mathbf{W}$ need not necessarily be the same, in which case the elements of the sequence are not identically distributed. Furthermore, condition (33) does not require that the sequence is a mixing process either in the sense that the elements of the sequence can be asymptotically uncorrelated but not asymptotically independent[10].

Remark 4 *Pesaran and Tosetti (2007) define the scalar sequence* $\left\{z_i^t, \, i \geq 1\right\}$ *to be weakly dependent at any given* t *if its (weighted) average converges to its expectation in quadratic mean. Specifically, let* w_i^{t-1} *denote a weight that satisfies* $\sum_i \left(w_i^{t-1}\right)^2 = O\left(N^{-1/2}\right)$ *and* $w_i^{t-1}\left[\sum_i \left(w_i^{t-1}\right)^2\right]^{-1} = O\left(N^{-1/2}\right)$ *for any* $i \leq N$*, and let* I_{t-1} *be the information set at time* $t-1$ *containing at least* $\mathbf{z}^{t-1}$, $\mathbf{z}^{t-2}, ...$ *and* $\mathbf{w}^{t-1}$, $\mathbf{w}^{t-2}, ...$, *where* $\mathbf{z}^{t-1} = \left(z_1^{t-1}, ..., z_N^{t-1}\right)'$ *and* $\mathbf{w}^{t-1} = \left(w_1^{t-1}, ..., w_N^{t-1}\right)'$. *Then the sequence* $\left\{z_i^t, \, i \geq 1\right\}$ *is weakly dependent if*

$$\lim_{N \to \infty} var\left(\sum_i w_i^{t-1} z_i^t \middle| I_{t-1}\right) = 0.$$

Under this definition, the following factor structure process

$$\begin{aligned} & u_{i,t} = \theta_i f_t + \varepsilon_{i,t}, \\ & \textit{where } \theta_i \textit{ is non-stochastic and bounded} \\ & f_t \sim i.i.d\left(0, \sigma_f^2\right), \, \varepsilon_{i,t} \sim i.i.d\left(0, \sigma_\varepsilon^2\right), \end{aligned} \tag{35}$$

is weakly correlated so long as $\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^N \theta_i = 0$[11]. *This is not the case, however, using Definition* 1 *since it is straightforward to show that* $\rho_\kappa^{t,t} \nrightarrow 0$ *as* $\kappa \to \infty$ *and therefore* $\sum_{\kappa=0}^{\infty} \rho_\kappa^{t,s}$ *is not bounded. Intuitively, no matter how large* κ *is,* $u_{i,t}$ *and* $u_{i+\kappa,t}$ *remain correlated in a non-negligible way regardless of whether* $\overline{\theta} \to 0$ *or not.*

[10] Of course, this requires a strengthening of the moment restrictions – namely, $E|u_i|^2 < B_u < \infty$, as opposed to – say – $E|u_i|^c < B_u < \infty$ for $c > 1$).

[11] See Pesaran and Tosetti (2007), Theorem 16, page 15.

The following theorem (due to White, 2001, Theorem 3.57) provides a law of large numbers for weakly correlated sequences.

Theorem 5 *Let* $\left\{\upsilon_i^t,\, i \geq 1\right\} = (\upsilon_{1,t}, \upsilon_{2,t}, \upsilon_{3,t}, ...)$ *be a scalar sequence with weakly correlated elements, such that* $E(\upsilon_{i,t}) = 0$ *and* $E\left(\upsilon_{i,t}^2\right) < B_\upsilon < \infty$. *Then*

$$\frac{1}{N}\sum_{i=1}^{N} \upsilon_{i,t} - E(\upsilon_{i,t}) \xrightarrow{p} 0 \tag{36}$$

Proof. It follows directly from Stout (1974, Corollary 2.4.1) and the Kronecker lemma. ■

Theorem 3 shows that so long as (33) holds true, $\upsilon_{i,t}$ is weakly correlated, or asymptotically uncorrelated across i. In turn, according to Theorem 5, the latter implies that the first sample centered moment of υ_i^t will converge in probability to zero. The following corollary provides the extra condition necessary to validate the moment conditions given in (12) under weakly correlated errors:

Corollary 6 *Let* $\left\{\upsilon_i^t,\, i \geq 1\right\}$ *and* $\left\{u_i^t,\, i \geq 1\right\}$ *be two scalar sequences* $(\upsilon_{1,t}, \upsilon_{2,t}, \upsilon_{3,t}, ...)$ *and* $(u_{1,t}, u_{2,t}, u_{3,t}, ...)$ *that satisfy* $||\mathbf{W}||_\infty = o\left(N^{1/2}\right)$ *individually and are, therefore, weakly correlated. The product of these sequences will also satisfy* $||\mathbf{W}||_\infty = o\left(N^{1/2}\right)$ *and will be weakly correlated.*

As a result, we have

$$\begin{aligned} &\text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\sum_{t=s+1}^{T}(y_{i,t-s}\Delta\upsilon_{i,t}) = \\ = &\sum_{t=s+1}^{T}\text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\left[\left(\frac{\alpha_i}{1-\lambda}+\sum_{\tau=0}^{\infty}\lambda^\tau u_{i,t-s-\tau}\right)\Delta u_{i,t}\right] = \\ = &\sum_{t=s+1}^{T}\sum_{\tau=0}^{\infty}\lambda^\tau\text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}u_{i,t-s-\tau}\Delta u_{i,t} = 0 \text{ for } s = 2,...,T-1. \end{aligned} \tag{37}$$

where the last line holds true because both $\{\upsilon_{i,t-s-\tau}, i \geq 1\}$ and $\{\Delta\upsilon_{i,t},\, i \geq 1\} = \{\Delta u_{i,t},\, i \geq 1\}$, are weakly correlated and so their product is also weakly correlated or asymptotically uncorrelated across i, with expected value equal to zero. Thus, the moment conditions used by DIF GMM remain valid.

Remark 7 *Observe that when a weakly correlated process is defined as in Remark* 4, *the sample average over* i *of the product between* $\{\upsilon_{i,t-s-\tau},\, i \geq 1\}$ *and* $\{\Delta\upsilon_{i,t},\, i \geq 1\}$ *does not necessarily converge to its expectation; for instance, for the single-factor process given in* (35) *and assuming that* $\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\theta_i = 0$, *we have* $\frac{1}{N}\sum_{i=1}^{N}u_{i,t-s} - E(u_{i,t-s}) \to 0$ *and* $\frac{1}{N}\sum_{i=1}^{N}\Delta u_{i,t} - E(\Delta u_{i,t}) \to 0$. *However, the sample average* $\frac{1}{N}\sum_{i=1}^{N}u_{i,t-s}\Delta u_{i,t}$ *converges to* $\mu_\theta^2 f_{t-s}\Delta f_t$, *where* $\mu_\theta^2 = plim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\theta_i^2$, *despite the fact that* $E(u_{i,t-s}\Delta u_{i,t}) = 0$ *for* $s = 2,...,t-1$.

Since asymptotic uncorrelatedness encompasses spatial dependence, it follows that DIF GMM is consistent under spatially correlated errors. On the other hand, under factor structure dependence the correlation between υ_{it} and υ_{jt} persists no matter how far apart individuals i and j are. Therefore the law of large numbers provided above breaks down and $\text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N} y_{i,t-s}\upsilon_{i,t} \neq 0$, despite the fact that $E\left(y_{i,t-s}\upsilon_{i,t}\right) = 0$[12].

SYS GMM also remains consistent under weakly or spatial correlated errors because

$$\text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\sum_{t=3}^{T}\left(\Delta y_{i,t-1}\upsilon_{i,t}\right) = \sum_{t=3}^{T}\text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\left[\left(\sum_{\tau=0}^{\infty}\lambda^{\tau}\Delta u_{i,t-1-\tau}\right)\upsilon_{it}\right] = 0. \tag{38}$$

In summary, it has been shown that the dynamic panel GMM estimator does not require cross-sectionally independent errors for consistency − rather, it suffices that, if there is such dependence, this is weak − in the way defined above − at any given point in time. Theorem 3 shows that this holds true under condition (33), which is more general than uniform boundedness of the row and column sums of $\mathbf{W}_N$. The factor structure in the error process violates this condition and therefore the standard GMM estimator is not consistent in this case.

5 Additional Moment Conditions Under Spatial Dependence

Suppose that the errors are spatially correlated but satisfy condition (33). It turns out that not only DIF GMM and SYS GMM are consistent, but also that there is an additional set of moment conditions which becomes relevant in this case. In particular, we consider the basic model given in (3) and for simplicitly we impose a SMA(1) error process − that is, $M = 1$, $\theta_i^1 = \theta$ with $|\theta| < 1$, $\boldsymbol{\xi}^1 = \boldsymbol{\varepsilon}$ and $\mathbf{W}_N$ is given by (8). Hence, the model becomes equal to

$$y_{i,t} = \lambda y_{i,t-1} + \left(\alpha_i + \theta\varepsilon_{j,t} + \varepsilon_{i,t}\right), \quad i = 1, ..., N \text{ and } t = 2, ..., T \tag{39}$$

where $j = i\,(\text{mod}\,N) + 1$[13]. In this case, an interesting result arises, as the following proposition demonstrates:

Proposition 8 *Under Assumptions 1-3, the panel autoregressive model in* (39) *implies that for each individual i there is an additional set of moment conditions that becomes relevant with respect to a different cross section, individual j, both in the first-differenced equations and those in levels. In particular, we have*

Moment Conditions for DIF GMM:

$$\text{plim}_{N\to\infty}\frac{1}{N}\left[\sum_{i=1}^{N}\sum_{t=s+1}^{T} y_{j,t-s}\Delta\upsilon_{i,t}\right] = 0; \text{ for } s = 2, ..., T-1, \tag{40}$$

[12] See also Sarafidis and Robertson (2007).

[13] See also (9). SMA processes of higher order can be accomodated in a similar fashion.

with

$$\text{plim}_{N\to\infty}\frac{1}{N}\left[\sum_{i=1}^{N}\sum_{t=s+1}^{T}y_{j,t-s}\Delta y_{i,t-1}\right]=-\theta\frac{(T-s)}{1+\lambda}\sigma_{\varepsilon}^{2} \tag{41}$$

Moment Conditions for SYS GMM (assuming that (26) is also satisfied):

$$\text{plim}_{N\to\infty}\frac{1}{N}\left[\sum_{i=1}^{N}\sum_{t=3}^{T}\Delta y_{j,t-1}\upsilon_{i,t}\right]=0 \tag{42}$$

with

$$\text{plim}_{N\to\infty}\frac{1}{N}\left[\sum_{i=1}^{N}\sum_{t=3}^{T}\Delta y_{j,t-1}y_{i,t-1}\right]=\theta\frac{(T-2)}{1+\lambda}\sigma_{\varepsilon}^{2} \tag{43}$$

Proof. See Appendix B. ■

Note that the error term of the regression model does not necessarily need to be of a spatial MA form. In fact, it is straightforward to show that these moment conditions are relevant under SAR and SEC errors or under more general spatial processes. Most notably, these moment conditions remain valid under both weakly and strongly correlated errors, as it will be shown in Proposition (9).

In the case of (39), we have

$$\begin{aligned}
E\left[\Delta\boldsymbol{\upsilon}\Delta\boldsymbol{\upsilon}'\right] &= E\left[\left(\theta\left(W_N\otimes I_T\right)\Delta\boldsymbol{\varepsilon}+\Delta\boldsymbol{\varepsilon}\right)\left(\theta\left(W_N\otimes I_T\right)\Delta\boldsymbol{\varepsilon}+\Delta\boldsymbol{\varepsilon}\right)'\right]\\
&= E\left[\left[\theta\left(W_N\otimes I_T\right)+I_{NT}\right]\Delta\boldsymbol{\varepsilon}\Delta\boldsymbol{\varepsilon}'\left[\theta\left(W_N'\otimes I_T\right)+I_{NT}\right]\right]\\
&= \sigma_{\varepsilon}^{2}\left[\theta\left(W_N\otimes I_T\right)+I_{NT}\right]\left(I_N\otimes H_i\right)\left[\theta\left(W_N'\otimes I_T\right)+I_{NT}\right]\\
&= \sigma_{\varepsilon}^{2}\left[\theta\left(W_N\otimes I_T\right)\left(I_N\otimes H_i\right)+I_{NT}\left(I_N\otimes H_i\right)\right]\left[\theta\left(W_N'\otimes I_T\right)+I_{NT}\right]\\
&= \sigma_{\varepsilon}^{2}\theta\left(W_N\otimes H_i\right)\theta\left(W_N'\otimes I_T\right)+\sigma_{\varepsilon}^{2}\theta\left(W_N\otimes H_i\right)\\
&\quad +\sigma_{\varepsilon}^{2}\left(I_N\otimes H_i\right)\theta\left(W_N'\otimes I_T\right)+\sigma_{\varepsilon}^{2}\left(I_N\otimes H_i\right)\\
&= \sigma_{\varepsilon}^{2}\left[\theta^{2}W_NW_N'+\theta\left(W_N+W_N'\right)+I_N\right]\otimes H_i
\end{aligned} \tag{44}$$

and therefore $\left[\theta^{2}W_NW_N'+\delta\left(W_N+W_N'\right)+I_N\right]$ replaces I_N in the expression for the weighting matrix of DIF GMM in (22). A similar point applies to I_N in (30) for SYS GMM. Of course, in practice θ is unknown; one option is to replace θ with an arbitrary value (say $\theta=0.5$) at first stage, and then obtain an estimate of θ by solving the following quadratic equation:

$$\widehat{\theta}_t^{2}r_t(1)-\widehat{\theta}_t+r_t(1)=0 \tag{45}$$

where $r_t(1)=Est.Correlation\left(\widehat{\upsilon}_{i,t},\widehat{\upsilon}_{j,t}\right)$ and $\widehat{\upsilon}_{i,t}$ is the first-stage residual of unit i for $t=2,...,T$. (45) has two solutions for each t, but given that $r_t^{-1}(1)=\widehat{\theta}_t+\widehat{\theta}_t^{-1}$ one root is the reciprocal of the other, which implies that the estimator for θ at time t equals

$$\widehat{\theta}_t=\frac{1-\sqrt{1-4r_t^{2}(1)}}{2r_t(1)} \tag{46}$$

The other solution can be ruled out since it will have an absolute value greater than one, which is not possible given the restriction $|\theta|<1$. A simple average $\widehat{\theta}=\frac{1}{T}\sum\widehat{\theta}_t$ can then be constructed to provide an estimate of $\widehat{\theta}$.

6 Consistent GMM Estimation under both Spatial and Factor Error Structure

The moment conditions analysed in the previous section can be particularly useful in general error processes that include unobserved common factors as well as omitted variables that are spatially correlated. This is because while the standard moment conditions in (12) and (25) are invalidated in this case[14], it turns out that the moment conditions obtained from a different cross section, individual j, are still valid. In particular, consider again the error process of the basic model given in (3) and suppose that while $\mathbf{W}_N^m = N^{-1/2}\mathbf{i}_N\mathbf{i}_N'$ for some m (e.g. $m = 1, ..., M$), there is at least a single $\mathbf{W}_N$ that satisfies condition (33) of Theorem 3. Let this be denoted by $\mathbf{W}_N^{M+1}$ and be equal to (8), although it should be clear by now that any $\mathbf{W}_N$ that satisfies condition (33) or uniform boundedness will do. In this case, the basic model can be rewritten as

$$\begin{aligned} y_{i,t} &= \lambda y_{i,t-1} + \upsilon_{i,t}, \\ \upsilon_{i,t} &= a_i + u_{i,t}, \quad u_{i,t} = \sum_{m=1}^{M} \theta_i^m f_t^m + \varepsilon_{i,t} + \theta\varepsilon_{j,t} \end{aligned} \tag{47}$$

with $f_t^m = N^{-1/2}\left(\sum \varepsilon_{1t}^m + ... + \varepsilon_{Nt}^m\right)$, $j = i \,(\mathrm{mod}\, N) + 1$ and $|\theta| < 1$. A similar error process that is subject to both spatial correlations and common unobserved factors is studied by Pesaran and Tosetti (2007).

Expressing (47) in terms of deviations from time-specific averages and using first-differences yields

$$\Delta\underline{y}_{i,t} = \lambda\Delta\underline{y}_{i,t-1} + \Delta\underline{\upsilon}_{i,t}, \quad \Delta\underline{\upsilon}_{i,t} = \sum_{m=2}^{M} \underline{\theta}_i^m \Delta f_t^m + \Delta\underline{\varepsilon}_{i,t} + \theta\Delta\underline{\varepsilon}_{j,t} \tag{48}$$

The following proposition demonstrates an important result:

Proposition 9 *Under Assumptions 1-4, the panel autoregressive model in* (48) *can be estimated consistently using method of moments estimators that rely on the following moment conditions*

Moment Conditions for DIF GMM:

$$\mathrm{plim}_{N\to\infty}\frac{1}{N}\left[\sum_{i=1}^{N}\sum_{t=s+1}^{T} \underline{y}_{j,t-s}\Delta\underline{\upsilon}_{i,t}\right] = 0; \text{ for } s = 2, ..., T-1, \tag{49}$$

with

$$\mathrm{plim}_{N\to\infty}\frac{1}{N}\left[\sum_{i=1}^{N}\sum_{t=s+1}^{T} \underline{y}_{j,t-s}\Delta\underline{y}_{i,t-1}\right] = -\theta\frac{(T-s)}{1+\lambda}\sigma_\varepsilon^2 \tag{50}$$

[14] The use of other instruments with respect to individual i will not help either, unless these instruments are not functions of (lagged values of) y and certain regularity conditions hold true, such as those in Sarafidis, Yamagata and Robertson (2007).

Moment Conditions for SYS GMM:

$$\text{plim}_{N\to\infty}\frac{1}{N}\left[\sum_{i=1}^{N}\sum_{t=3}^{T}\Delta\underline{y}_{j,t-1}\underline{\upsilon}_{i,t}\right]=0 \tag{51}$$

with

$$\text{plim}_{N\to\infty}\frac{1}{N}\left[\sum_{i=1}^{N}\sum_{t=3}^{T}\Delta\underline{y}_{j,t-1}\underline{y}_{i,t-1}\right]=\theta\frac{(T-2)}{1+\lambda}\sigma_{\varepsilon}^{2} \tag{52}$$

Proof. See Appendix C. ■

The above implies that the model given in (47) can be estimated consistently using a simple IV estimator that employs $\underline{y}_{j,t-2}$ as an instrument for $\Delta\underline{y}_{i,t-1}$, or a first-differenced GMM estimator that instruments $\Delta\underline{y}_{i,t-1}$ by $\underline{y}_{j,t-s}$ for $s=2,3,...$, and a system GMM estimator that uses $\Delta\underline{y}_{j,t-1}$ as an instrument for $\underline{y}_{i,t-1}$ in the levels equations. This is because the correlation between $\underline{y}_{j,t-s}$and $\Delta\underline{y}_{i,t-1}$ (or between $\Delta\underline{y}_{j,t-1}$ and $\underline{y}_{i,t-1}$ in levels) is non-zero while the correlation between $\underline{y}_{j,t-s}$and $\Delta\underline{\upsilon}_{i,t}$ (and $\Delta\underline{y}_{j,t-1}$and $\underline{\upsilon}_{i,t}$ in levels) remains zero. Therefore, defining $\mathbf{Z}_{MM}=\left(Z_1^{MM},...,Z_N^{MM}\right)'$ with $Z_i^{MM}=\left(\underline{y}_{j,1},\underline{y}_{j,2},...,\underline{y}_{j,T-2}\right)'$ as well as the following matrices of instruments

$$\underline{Z}_i^{\dagger}=\begin{bmatrix}\underline{y}_{j,1} & 0 & 0 & \cdots & 0 & 0 & \cdots & 0\\ 0 & \underline{y}_{j,1} & \underline{y}_{j,2} & \cdots & 0 & 0 & \cdots & 0\\ \vdots & \vdots & \vdots & & \vdots & \vdots & & \vdots\\ 0 & 0 & 0 & \cdots & \underline{y}_{j,1} & \underline{y}_{j,2} & \cdots & \underline{y}_{j,T-2}\end{bmatrix};\ \Delta\underline{\upsilon}_i=\begin{bmatrix}\Delta\underline{\upsilon}_{i,3}\\ \Delta\underline{\upsilon}_{i,4}\\ \vdots\\ \Delta\underline{\upsilon}_{i,T}\end{bmatrix}, \tag{53}$$

and

$$\underline{Z}_i^{\dagger sys}=\begin{bmatrix}\underline{Z}_i^{\dagger} & 0 & \cdots & 0\\ 0 & \Delta\underline{y}_{j,2} & & 0\\ . & . & \ddots & .\\ 0 & 0 & \cdots & \Delta\underline{y}_{j,T}\end{bmatrix};\ \underline{\upsilon}_i^{sys}=\begin{bmatrix}\Delta\underline{\upsilon}_i\\ \underline{\upsilon}_i\end{bmatrix}, \tag{54}$$

Proposition 9 implies that the following moment estimators are valid:

$$\widehat{\lambda}_{IV}^{\dagger}=\left(\mathbf{Z}'_{MM}\Delta\underline{\mathbf{y}}_{-1}\right)^{-1}\left(\mathbf{Z}'_{MM}\Delta\underline{\mathbf{y}}\right), \tag{55}$$

$$\widehat{\lambda}_{DIF\ GMM}^{\dagger}=\left[\frac{1}{N}\Delta\underline{\mathbf{y}}'_{-1}\underline{\mathbf{Z}}\left(\widehat{\mathbf{A}}_N^{\dagger}\right)^{-1}\underline{\mathbf{Z}}'\Delta\underline{\mathbf{y}}_{-1}\right]^{-1}\left[\frac{1}{N}\Delta\underline{\mathbf{y}}'_{-1}\underline{\mathbf{Z}}\left(\widehat{\mathbf{A}}_N^{\dagger}\right)^{-1}\underline{\mathbf{Z}}'\Delta\underline{\mathbf{y}}\right] \tag{56}$$

and

$$\widehat{\lambda}_{SYS\ GMM}^{\dagger}=\left[\frac{1}{N}\underline{\mathbf{Y}}'_{-1}\underline{\mathbf{Z}}^{sys}\left(\widehat{\mathbf{A}}_{\mathbf{N,sys}}^{\dagger}\right)^{-1}\underline{\mathbf{Z}}^{sys\prime}\underline{\mathbf{Y}}_{-1}\right]^{-1}\left[\frac{1}{N}\underline{\mathbf{Y}}'_{-1}\underline{\mathbf{Z}}^{sys}\left(\widehat{\mathbf{A}}_{\mathbf{N,sys}}^{\dagger}\right)^{-1}\underline{\mathbf{Z}}^{sys\prime}\underline{\mathbf{Y}}\right]. \tag{57}$$

$\left(\widehat{\mathbf{A}}_N^{\dagger}\right)^{-1}$ is the weighting matrix of the two-step first-differenced GMM estimator, which can be estimated from

$$\widehat{\mathbf{A}}_N^{\dagger} = \underline{\mathbf{Z}}' \left[\left(\Delta \widehat{\underline{\upsilon}} \Delta \widehat{\underline{\upsilon}}' \right) \otimes H_i \right] \underline{\mathbf{Z}} \tag{58}$$

where H_i has been defined in (23) and $\Delta\widehat{\underline{\upsilon}}$ is an $N \times (T-2)$ matrix of residuals, obtained from a first-step first-differenced GMM estimator. Notice that the least-squares estimate of $\Delta\widehat{\underline{\upsilon}}\Delta\widehat{\underline{\upsilon}}'$ in (58) is rank deficient because it is an $N \times N$ matrix and has rank $T-2$. The matrix inside the square brackets of (58) is also rank deficient because it is a square matrix of order $N(T-2)$ and has rank $(T-2)^2$. However, $\widehat{\mathbf{A}}_N^{\dagger}$ is a square $\zeta \times \zeta$ matrix, which has rank equal to $\min\left(\zeta, (T-2)^2\right)$. Therefore, provided that we do not use too many instruments, i.e. $\zeta \leq (T-2)^2$, (58) will be of full rank and the weighting matrix will exist.

$\left(\widehat{\mathbf{A}}_{N,sys}^{\dagger}\right)^{-1}$ is the weighting matrix of the two-step system GMM estimator, which can be estimated from

$$\widehat{\mathbf{A}}_{N,sys}^{\dagger} = \underline{\mathbf{Z}}^{sys\prime} \left(\widehat{\mathbf{Q}}\right) \underline{\mathbf{Z}}^{sys} \tag{59}$$

with $\widehat{\mathbf{Q}}$ being equal to

$$\widehat{\mathbf{Q}} = \begin{bmatrix} \left(\Delta\widehat{\underline{\upsilon}}\Delta\widehat{\underline{\upsilon}}' \right) \otimes (H_i) & \mathbf{0} \\ \mathbf{0} & \left(\widehat{\underline{\upsilon}}\widehat{\underline{\upsilon}}' \right) \otimes (I_{T-2}) \end{bmatrix} \tag{60}$$

7 Properties of GMM Estimators

To investigate the properties of these moment estimators we follow the approach by Blundell and Bond (1998) and we consider the case where $T = 3$, for which there is only a single instrument available for the endogenous regressor, both in the first-differenced equations and those in levels. In this way, DIF GMM and SYS GMM reduce to simple instrumental variable estimators and the corresponding first-stage regressions may help to analyse the 'strength' of the instruments used as a function of the parameters of interest in more general cases.

7.1 Equations in First-Differences

For the equations in first-differences, the first-stage regression is given by

$$\Delta \underline{y}_{i,2} = \pi^d \underline{y}_{j,1} + w_i \tag{61}$$

where w_i is an error term. The ordinary least-squares estimate of π^d, which reflects the strength of the relation between the instrument and the endogenous regressor, is equal to

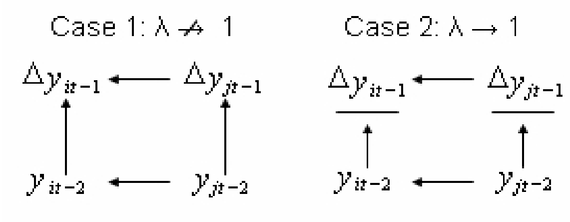
$$\widehat{\pi}^d = \frac{\sum_{i=1}^N \underline{y}_{j,1} \Delta \underline{y}_{i,2}}{\sum_{i=1}^N \underline{y}_{j,1}^2} \tag{62}$$

Using Assumptions 1-4 in model (47) it is straightforward to show that the plim of $\widehat{\pi}^d$ equals

$$
\begin{aligned}
& \operatorname{plim}_{N\to\infty}\left(\widehat{\pi}^d\right) \\
= & \frac{(\lambda-1)\operatorname{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\underline{y}_{j,1}\underline{y}_{i,1}+\operatorname{plim}_{N\to\infty}\frac{1}{N}\sum\underline{y}_{j,1}\left(\underline{y}_{i,2}-\lambda\underline{y}_{i,1}\right)}{\operatorname{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\underline{y}_{j,1}^2} \\
= & (\lambda-1)\,\theta\frac{\sigma_\varepsilon^2}{1-\lambda^2}\left[\frac{\sigma_\alpha^2}{(1-\lambda)^2}+\frac{\sigma_\varepsilon^2}{1-\lambda^2}\left(1+\theta^2\right)+\mathbf{w}_1'\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\mathbf{w}_1\right]^{-1} \\
= & (\lambda-1)\,\theta\left[\frac{\kappa}{\sigma_\alpha^2/\sigma_\varepsilon^2+\left(1+\theta^2\right)\kappa+\left(1-\lambda^2\right)\kappa\left[\mathbf{w}_1'\left(\boldsymbol{\Sigma}_{\boldsymbol{\theta}}/\sigma_\varepsilon^2\right)\mathbf{w}_1\right]}\right] \qquad (63)
\end{aligned}
$$

where $\kappa=\frac{1-\lambda}{1+\lambda}$ and $\mathbf{w}_1=\sum_{s=0}^{\infty}\lambda^s\mathbf{f}_{1-s}$.

Thus, we can see that for fixed T the plim of $\widehat{\pi}^d$ depends on various parameters, namely λ, θ, σ_α^2, $\boldsymbol{\Sigma}_\theta$ and σ_ε^2. For example, as $\lambda\to 1$ the plim of the estimator converges to zero, which implies that the correlation between $\underline{y}_{j,t-2}$ and $\Delta\underline{y}_{i,t-1}$ becomes weak. The intuition behind this is illustrated in the following figure, which shows two cases of λ when the values of σ_α^2, $\boldsymbol{\Sigma}_\theta$, σ_ε^2 and θ are held fixed:



Weak instruments with cross section dependence.

When $\lambda\nrightarrow 1$, $\underline{y}_{j,1}$ is correlated with $\Delta\underline{y}_{i,2}$ and since $Cov\left(\underline{y}_{j,1},\Delta\underline{v}_{i,3}\right)=0$ $\underline{y}_{j,1}$ is a valid instrument. Note that the use of $\underline{y}_{i,1}$ as an instrument is not valid here because $\boldsymbol{\Sigma}_\theta\neq 0$ and therefore $Cov\left(\underline{y}_{i,1},\Delta\underline{v}_{i,3}\right)\neq 0$. On the other hand, as $\lambda\to 1$ the correlation between $\underline{y}_{j,1}$ and $\Delta\underline{y}_{j,2}$ becomes weak; this is because the link between $\Delta\underline{y}_{j,2}$ and $\Delta\underline{y}_{i,2}$ is not effective anymore since $\underline{y}_{j,1}$ is poorly correlated with $\Delta\underline{y}_{j,2}$, while the link between $\underline{y}_{j,1}$ and $\underline{y}_{i,1}$ does not help either because $\underline{y}_{i,1}$ is poorly correlated with $\Delta\underline{y}_{i,2}$.

When there is no variation in the factor loadings across i, $\boldsymbol{\Sigma}_\theta=0$ and the plim of $\widehat{\pi}^d$ remains non-zero but of course in this case $\underline{y}_{i,1}$ is also valid as instrument. On the other hand, for a given non-zero value of θ and $|\lambda|<1$ the plim of the estimator converges to zero as either $\left(\sigma_\alpha^2/\sigma_\varepsilon^2\right)\to\infty$ or $\left(\boldsymbol{\Sigma}_\theta/\sigma_\varepsilon^2\right)\to\infty$. The former result is similar to Blundell and Bond (1998). Interestingly, the same appears to apply for the ratio between $\mathbf{\Sigma}_\theta$ and σ_ε^2. Intuitively, this is because the contribution of the spatial component of the error process in $\Delta\underline{\upsilon}_{i,3}$ (and thereby the correlation between Δy_{i2} and y_{j1}) diminishes with high values of $\mathbf{\Sigma}_\theta$ and increases with high values of σ_ε^2.

7.2 Equations in Levels

For the equations in levels the first-stage regression is given by

$$\underline{y}_{i,2} = \pi^l \Delta \underline{y}_{j,2} + w_i^l \tag{64}$$

and the least squares estimator of π^l equals

$$\widehat{\pi}^l = \frac{\sum_{i=1}^N \Delta \underline{y}_{j,2} \underline{y}_{i,2}}{\sum_{i=1}^N \left(\Delta \underline{y}_{j,2}\right)^2} \tag{65}$$

Using Assumptions 1-4, it is straightforward to show that the plim of $\widehat{\pi}^l$ equals

$$\begin{aligned} & \operatorname{plim}_{N\to\infty}\left(\widehat{\pi}^l\right) \\ = & \frac{\operatorname{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^N \underline{y}_{j,2}\underline{y}_{i,2} - \lambda \operatorname{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^N \underline{y}_{j,1}\underline{y}_{i,1} - \operatorname{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^N \underline{y}_{j,1}\underline{\upsilon}_{i,2}}{\operatorname{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^N \left(\Delta \underline{y}_{j,2}\right)^2} \\ = & \frac{\theta}{2\left(1+\theta^2\right) + (1+\lambda)\left[\Delta \mathbf{w}_2'\left(\mathbf{\Sigma}_{\boldsymbol{\theta}}/\sigma_\varepsilon^2\right)\Delta \mathbf{w}_2\right]} \end{aligned} \tag{66}$$

where $\Delta \mathbf{w}_2 = \sum_{\tau=0}^\infty \lambda^\tau \Delta \mathbf{f}_{2-\tau}$. Here we can see that as $\lambda \to 1$ the above expression converges to

$$\operatorname{plim}_{N\to\infty}\left(\widehat{\pi}^d\right) = \frac{1}{2}\frac{\theta}{\left(1+\theta^2\right) + \Delta \mathbf{w}_2'\left(\mathbf{\Sigma}_{\boldsymbol{\theta}}/\sigma_\varepsilon^2\right)\Delta \mathbf{w}_2} \tag{67}$$

and so $\Delta \underline{y}_{j,2}$ remains informative as an instrument for $\underline{y}_{i,2}$, provided of course that $\theta \neq 0$. In addition, when $\mathbf{\Sigma}_\theta = 0$ the random element in (67) disappears and the expression becomes equal to a constant number $-$ specifically, $\operatorname{plim}_{N\to\infty}\left(\widehat{\pi}^l\right) = \theta/\left[2\left(1+\theta^2\right)\right]$. Similarly to (63), the plim of $\widehat{\pi}^l$ converges to zero as $\left(\mathbf{\Sigma}_{\boldsymbol{\theta}}/\sigma_\varepsilon^2\right) \to \infty$ for the same reason that has been discussed previously $-$ that is, because the contribution of the spatial component of the error process in υ_{i3} diminishes.

8 Small Sample Properties of Moment Estimators

This section investigates the finite-sample performance of the various estimators proposed in this paper using simulated data. The main focus of the analysis lies on the impact of the relative importance of the unobserved factors in the total error process for different values of N, T and λ.

8.1 Monte Carlo Design

The underlying data generating process is given by

$$\begin{aligned} y_{it} &= \lambda y_{it-1} + \alpha_i + u_{it}, \\ u_{it} &= \phi_i f_t + \varepsilon_{it} + \theta \varepsilon_{jt},\ i = 1, 2, ..., N; t = -48, -47, ..., T. \end{aligned} \tag{68}$$

where $\alpha_i \sim iidN\left(0, \sigma_\alpha^2\right)$, $\varepsilon_{it} \sim iidN\left(0, \sigma_\varepsilon^2\right)$, $f_t \sim iidN\left(0, \sigma_f^2\right)$ and $j = i\,(\mathrm{mod}\, N)$. Also, the factor loadings are drawn from

$$\phi_i \sim iidU\left[-0.25, 0.25\right] \tag{69}$$

The performance of GMM estimation depends crucially upon the ratio of the two variance components, a_i and u_{it}, on $var(y_{it})$ as shown in (63). This implies that as the value of λ increases, or the amount of cross-sectional dependence decreases, the impact of α_i on $var\left(y_{it}\right)$ will tend to become larger and thereby comparisons across experiments with different levels of cross section dependence will not be valid. To control this ratio we use the following simple result

$$var\left(y_{it}\right) = var\left[\frac{\alpha_i}{1-\lambda} + \left(\sum_{s=0}^{\infty} \lambda^s u_{it-s}\right)\right] = \frac{\sigma_\alpha^2}{\left(1-\lambda\right)^2} + \frac{\sigma_u^2}{1-\lambda^2} \tag{70}$$

and we set $\sigma_\alpha^2 = \psi\left[\left(1-\lambda\right)/\left(1+\lambda\right)\right]\sigma_u^2$ with $\psi = 1$[15].

In addition to $\sigma_\alpha^2/\sigma_u^2$, the performance of the estimators will depend on the proportion of σ_u^2 attributed to the factor structure in u_{it} $-$ hereafter this proportion is denoted by $\zeta_{(d)}$, $d = 1, ..., 4$. Therefore, noticing that

$$\sigma_u^2 = \left(\mu_\phi\right)^2 \sigma_f^2 + \sigma_\phi^2 \sigma_f^2 + \sigma_\varepsilon^2\left(1+\theta^2\right) \tag{71}$$

and normalising $\sigma_f^2 = 1$, we can produce the following result

$$\sigma_\varepsilon^2 = \frac{\left(1-\zeta_{(d)}\right)}{\zeta_{(d)}} \frac{\left(\mu_\phi\right)^2 + \sigma_\phi^2}{\left(1+\theta^2\right)} \tag{72}$$

Since the values of $\left(\mu_\phi\right)^2$ and σ_ϕ^2 are determined solely by (69) and so they are fixed, normalising $\theta = 0.5$ implies that σ_ε^2 will change only according to $\zeta_{(d)}$. As this ratio increases, the impact of the factor structure in the error process will rise. We choose the following values for $\zeta_{(d)}$:

$$\left\{\begin{array}{cc} \text{Low impact of factor structure on } u_{it}\text{:} & \zeta_{(1)} = 1/3 \\ \text{Medium impact of factor structure on } u_{it}\text{:} & \zeta_{(2)} = 1/2 \\ \text{Medium-to-high impact of factor structure on } u_{it}\text{:} & \zeta_{(3)} = 2/3 \\ \text{High impact of factor structure on } u_{it}\text{:} & \zeta_{(4)} = 3/4 \end{array}\right.$$

[15] See Kiviet (1995) and Bun and Kiviet (2006).

We consider $N = 400$, 800 and $T = 6$, 10, since our focus is T fixed, $N \to \infty$. λ alternates between 0.5, 0.7 and 0.9. The initial value of y_{it} has been set equal to zero but the first 50 observations have been discarded before choosing the sample, so as to ensure that the initial zero values do not have an impact on the results. All experiments are based on 2,000 replications.

8.2 Results

Tables 1-2 report the simulation results in terms of the mean value of $\widehat{\lambda}_r$, where r denotes the r^{th} replication, and RMSE for each of the estimators used in the experiment. FE is the fixed effects estimator, IV is the simple instrumental variables estimator that uses y_{it-2} as an instrument for Δy_{it-1} and DIF and SYS denote the first-differenced and system GMM estimators respectively[16]. The superscript '*' indicates that the corresponding estimator uses instrument(s) with respect to another cross section, unit j.

As expected, the performance of all estimators depends on $\zeta_{(d)}$, the value of λ and the size of T and N. Specifically, as the value of ζ increases for a given value of λ, T and N, the estimators suffer a rise in bias and in RMSE. This is natural because as the relative impact of the factor structure in the total error process increases, the invalidity of the instruments used with respect to unit i itself (such as in IV, DIF and SYS) is magnified. For the estimators that make use of instruments with respect to unit j, the rise in bias and RMSE is also intuitive because as ζ increases, the contribution of the spatial component in the error process − and thereby the correlation between Δy_{it-1} and y_{jt-2} − diminishes.

Having said that, two things are clear from these results; first, IV*, DIF* and SYS* outperform IV, DIF and SYS respectively under all circumstances. Second, the relative performance of IV*, DIF* and SYS* improves with larger values of ζ. This is also intuitive − ultimately, as $\zeta \to 0$ the factor structure in the error process diminishes and the asymptotic bias of IV, DIF and SYS approaches zero. Notice also that in terms of RMSE, SYS* performs better than DIF*, which performs better than IV*, with the relative difference in performance being increased according to the value of λ. As T rises, the performance of the estimators improves without exception.

Finally, it is important to emphasise that as the size of N increases, the bias and RMSE of IV*, DIF* and SYS* decreases considerably. This is not the case for the conventional estimators, IV, DIF and SYS, the performance of which − if anything − deteriorates with larger values of N.

9 Concluding Remarks

Error cross section dependence is an increasingly popular research topic in the analysis of panel data. Despite this fact, the issue has mot attracted much attention in GMM

[16] DIF and SYS are estimated in two steps and they use y_{it-2} and y_{it-3} as instruments for Δy_{it-1} in the first-differenced equations. Furthermore, SYS GMM uses the optimal weighting matrix (when $\sigma_\alpha^2 = 0$), as derived in Windmeijer (2000).

estimation of short dynamic panels, where it is commonly assumed that the regression errors are are independent across i. This paper has shown that, in fact, independence or uncorrelatedness is not necessary for GMM consistency or asymptotic efficiency − rather, it is sufficient that, if there is such correlation in the errors, this is weak in the sense that any two errors that lie sufficiently far apart in the stochastic sequence exhibit very little correlation at any given point in time. If this condition is not satisfied, the errors are said to be strongly correlated. Spatial dependence presents an example of weakly correlated errors while the factor structure dependence provides an example of strongly correlated errors. Therefore, the standard dynamic panel GMM estimators that exist in the literature remain consistent under spatially correlated errors but not so under a factor structure. When the errors are weakly correlated there are additional moment conditions that arise − in particular, instruments with respect to the individual(s) which unit i is correlated with. We demonstrate that these moment conditions can be particularly useful when the errors are subject to both weak and strong correlations, a situation that is likely to arise in practice. The properties of these GMM estimators have been analysed under different circumstances. Simulated experiments have shown that the resulting estimators outperform the conventional ones, in terms of both bias and RMSE. This result is magnified as the impact of the factor structure in the total error process increases. In addition, larger values of N are accompanied by a considerable decrease in bias and RMSE for the estimators put forward in this paper. This is not the case with the conventional estimators, the performance of which is naturally not affected by the size of N.

References

[1] Ahn, S.C and P. Schmidt, (1995) "Efficient Estimations of Models for Dynamic Panel Data", *Journal of Econometrics*, 68, 5-28.

[2] Anderson, T.W. and C. Hsiao (1981) "Estimation of Dynamic Models with Error Components", *Journal of the American Statistical Association*, 76, 598-606.

[3] Arellano, M. (1993) "On the Testing of Correlated Effects with Panel Data", *Journal of Econometrics*, 59, 87-97.

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

[5] Arellano, M. and O. Bover (1995) "Another Look at the Instrumental Variable Estimation of Error-Component Models", *Journal of Econometrics*, 68, 29-51.

[6] Bai, J. (2005) "Panel Data Models with Interactive Fixed Effects", unpublished manuscript.

[7] Baltagi, B., Bresson, G. and A. Pirotte (2007) "Panel Unit Root Tests and Spatial Dependence", *Journal of Applied Econometrics*, 22, 339-360.

[8] Bover, O. and N. Watson (2005) "Are there economies of scale in the demand for money by firms? Some panel data estimates", *Journal of Monetary Economics*, 52, 1569-1589.

[9] Blundell, R. and S. Bond (1998) "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models", *Journal of Econometrics*, 87, 115-143.

[10] Bun, M.J.G. and J.F. Kiviet (2006) "The Effects of Dynamic Feedbacks on LS and MM Estimator Accuracy in Panel Data Models", *Journal of Econometrics* 132, 409-444.

[11] Coakley, J., Fuertes, A. and Smith, R. (2002) "A Principal Components Approach to Cross-Section Dependence in Panels", Unpublished manuscript, Birckbeck College, University of London.

[12] Conley, T.G. (1999) "GMM Estimation with Cross Sectional Dependence", *Journal of Econometrics,* 92, 1-45.

[13] Goldberger, A.S. (1972) "Structural Equation Methods in the Social Sciences", *Econometrica*, 40, 979-1001.

[14] Jöreskog, K.G. and A.S. Goldberger (1975) "Estimation of a model with multiple indicators and multiple causes of a single latent variable" *Journal of the American Statistical Association*, 70, 631-639.

[15] Kapoor, M., Kelejian, H., and I.R. Prucha (2007) "Panel Data Models with Spatially Correlated Error Components", *Journal of Econometrics* 140, 97-130.

[16] Kelejian, H and D. Robinson (1995) "Spatial Correlation: A Suggested Alternative to the Autoregressive Model, in Anselin, L. and R.J. Florax (ed.), *New Directions in Spatial Econometrics*, 75-95, Springer-Verlag, Berlin.

[17] Kiviet, J.F., (1995) "On bias, inconsistency, and efficiency of various estimators in dynamic panel data models", *Journal of Econometrics*, 68, 53-78.

[18] Lee, L. (2007) "GMM and 2SLS Estimation of Mixed Regressive, Spatial Autoregressive Models", *Journal of Econometrics* 137, 489-514.

[19] Moon, H.R. and B. Perron (2004) "Efficient Estimation of the SUR Cointegrating Regression Model and Testing for Purchasing Power Parity", *Econometric Reviews*, Vol. 23, 293-323.

[20] Pesaran, H. (2006) "Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure", *Econometrica*, 74, 967-1012.

[21] Pesaran, H. and E. Tosetti (2007) "Large Panels with Common Factors and Spatial Correlations", unpublished manuscript.

[22] Phillips, P. and D. Sul (2003) “Dynamic Panel Estimation and Homogeneity Testing under Cross Section Dependence”, *The Econometrics Journal*, Vol. 6, 217-259.

[23] Phillips, P. and D. Sul (2007). “Bias in Dynamic Panel Estimation with Fixed Effects, Incidental Trends and Cross Section Dependence”, *Journal of Econometrics*, 137, 162-188.

[24] Robertson, D. and Symons, J. (2000) “Factor Residuals in SUR Regressions: Estimating Panels Allowing for Cross Sectional Correlation”, unpublished manuscript, Faculty of Economics and Politics, University of Cambridge.

[25] Sarafidis, V. Yamagata, T. and D. Robertson (2007) “A Test of Error Cross Section Dependence for a Linear Dynamic Panel Model with Regressors”, unpublished manuscript.

[26] Sarafidis, V. and D. Robertson (2007) “On the Impact of Cross Section Dependence in Short Dynamic Panel Estimation”, unpublished manuscript.

[27] Stetzer, F. (1982) “Specifying Weights in Spatial Forecasting Models: The Results of Some Experiments”, *Environment and Planning A*, 14, 571-584.

[28] Stout, W. (1974) “Almost Sure Convergence”, Academic Press, New York.

[29] White, H. (2001) “Asymptotic Theory for Econometricians”, Academic Press, New York.

[30] Windmeijer, F. (2000) “Efficiency Comparisons for a System GMM Estimator in Dynamic Panel Data Models”, in R.D.H. Heijmans, D.S.G. Pollock and A. Satora (ed.), *Innovations in Multivariate Statistical Analysis*, Dordrecht: Kluwer Academic Publishers.

Appendices

A Proof of Theorem 3

The error process is given by $\upsilon_{i,t} = \alpha_i + \theta_i \sum_{j=1}^{N} w_{i,j}\xi_{j,t} + \varepsilon_{i,t}$, where $E(\upsilon_{i,t}) = 0$. Note that for $E|u_i|^2 < B_u < \infty$, we must have $||\mathbf{W}||_\infty < B_u^{1/2} < \infty$. Hence there are two cases; if the number of non-zero elements in $\mathbf{W}$ is finite, such that $C_i = O(1)$ for all i, where C_i is the number of non-zero elements in row i of $\mathbf{W}$, then $|w_{i,j}|$ can be any real number so long as it is sufficiently bounded. When C_i grows with N, $|w_{i,j}| = O\left(N^{-1/2-\varepsilon}\right)$ for some $\varepsilon \geq 0$.

The correlation coefficient between υ_i^t and υ_{i+k}^s is given by

$$\rho_\kappa^{t,s} = \frac{Cov(\upsilon_{i,t}, \upsilon_{i+k,s})}{\left[Var(\upsilon_{i,t})\, Var(\upsilon_{i+k,s})\right]^{1/2}} = \frac{E(\upsilon_{i,t}\upsilon_{i+k,s})}{\left[E\left(\upsilon_{i,t}^2\right) E\left(\upsilon_{i+\kappa,s}^2\right)\right]^{1/2}}$$
$$= \begin{cases} \dfrac{\theta_i\theta_{i+k}\sigma_\xi^2\left(\sum_{j=1}^N w_{i,j}w_{i+\kappa,j}\right)}{\left[\left(\sigma_\alpha^2+\theta_i^2\sigma_\xi^2\sum_{j=1}^N w_{i,j}^2+\sigma_\varepsilon^2\right)\left(\sigma_\alpha^2+\theta_{i+\kappa}^2\sigma_\xi^2\sum_{j=1}^N w_{i+\kappa,j}^2+\sigma_\varepsilon^2\right)\right]^{1/2}}, & \text{for } t=s \\ 0 \text{ otherwise} & \end{cases}$$

The condition $\sum_{\kappa=0}^{\infty}\rho_\kappa^{t,s} < \infty$ is automatically satisfied when $C_i = O(1)$ for all i, given that the non-zero values of $\mathbf{W}$ are sufficiently bounded. When C_i grows with N, the condition $||\mathbf{W}||_\infty = o\left(N^{1/2}\right)$ implies that $|w_{i,j}| = o\left(N^{-1/2}\right)$ and therefore $|w_{i,j}w_{i+\kappa,j}| = o\left(N^{-1}\right)$. As a result, $\sum_{j=1}^N |w_{i,j}w_{i+\kappa,j}| = o(1)$ and $\left[\left(\sigma_\alpha^2+\theta_i^2\sigma_\xi^2\sum_{j=1}^N w_{i,j}^2+\sigma_\varepsilon^2\right)\left(\sigma_\alpha^2+\theta_{i+\kappa}^2\sigma_\xi^2\sum_{j=1}^N w_{i+\kappa,j}^2+\sigma_\varepsilon^2\right)\right] = O(1)$. Since $\theta_i\theta_{i+k}\sigma_\xi^2 = O(1)$, it follows that $\rho_\kappa^{t,s} = o(1)$ for κ sufficiently large and $t = s$.

B Proof of Proposition 8

Assuming that the $y_{i,t}$ process has started a long time ago, it can be shown that

$$y_{j,t-2} = \frac{\alpha_j}{1-\lambda} + \sum_{s=0}^{\infty}\lambda^s\varepsilon_{j,t-s-2} + \theta\sum_{s=0}^{\infty}\lambda^s\varepsilon_{j',t-s-2} \tag{73}$$

where

$$j' = j \,(\mathrm{mod}\, N) + 1 \tag{74}$$

and $j = i\,(\mathrm{mod}\, N) + 1$. Hence, for DIF GMM we have

$$\begin{aligned}&\mathrm{plim}_{N\to\infty}\frac{1}{N}\left[\sum_{i=1}^{N}\sum_{t=3}^{T} y_{j,t-s}\Delta\upsilon_{i,t}\right] = \\ = &\sum_{t=3}^{T}\mathrm{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\left[\left(\frac{\alpha_j}{1-\lambda}+\sum_{\tau=0}^{\infty}\lambda^\tau\varepsilon_{j,t-\tau-s}+\theta\sum_{\tau=0}^{\infty}\lambda^\tau\varepsilon_{j',t-\tau-s}\right)(\Delta\varepsilon_{i,t}+\theta\Delta\varepsilon_{j,t})\right] = 0,\end{aligned} \tag{75}$$

using 2-5 for $s = 2, ..., T-1$. Furthermore, the covariance between $y_{j,t-2}$ and $\Delta y_{i,t-1}$ is different from zero since

$$\begin{aligned}
& \operatorname{plim}_{N\to\infty} \frac{1}{N}\left[\sum_{i=1}^{N}\sum_{t=s+1}^{T} y_{j,t-s}\Delta y_{i.t-1}\right] = \\
= & \sum_{t=s+1}^{T} \operatorname{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\left[\left(\frac{\alpha_j}{1-\lambda}+\sum_{\tau=0}^{\infty}\lambda^{\tau}\varepsilon_{j,t-\tau-s}+\theta\sum_{\tau=0}^{\infty}\lambda^{\tau}\varepsilon_{j',t-\tau-s}\right)\right. \\
& \left.\cdot\left(\sum_{\tau=0}^{\infty}\lambda^{\tau}\Delta\varepsilon_{i,t-\tau-1}+\theta\sum_{\tau=0}^{\infty}\lambda^{\tau}\Delta\varepsilon_{j,t-\tau-1}\right)\right] \\
= & -\theta\frac{(T-s)}{1+\lambda}\sigma_{\varepsilon}^{2}\neq 0;\ \text{for } s=2,...,T-1. \qquad (76)
\end{aligned}$$

For SYS GMM we have

$$\begin{aligned}
& \operatorname{plim}_{N\to\infty}\frac{1}{N}\left[\sum_{i=1}^{N}\sum_{t=3}^{T}\Delta y_{j,t-1}\upsilon_{i,t}\right] = \\
= & \sum_{t=3}^{T}\operatorname{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\left[\left(\sum_{\tau=0}^{\infty}\lambda^{\tau}\Delta\varepsilon_{j,t-\tau-1}+\theta\sum_{\tau=0}^{\infty}\lambda^{\tau}\Delta\varepsilon_{j',t-\tau-1}\right)(\alpha_i+\varepsilon_{i,t}+\theta\varepsilon_{j,t})\right]=0. \qquad (77)
\end{aligned}$$

Furthermore, the covariance between $\Delta y_{j,t-1}$ and $y_{i,t-1}$ equals

$$\begin{aligned}
& \operatorname{plim}_{N\to\infty}\frac{1}{N}\left[\sum_{i=1}^{N}\sum_{t=3}^{T}\Delta y_{j,t-1}y_{i,t-1}\right] = \\
= & \sum_{t=3}^{T}\operatorname{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\left[\left(\sum_{\tau=0}^{\infty}\lambda^{\tau}\Delta\varepsilon_{j,t-\tau-1}+\theta\sum_{\tau=0}^{\infty}\lambda^{\tau}\Delta\varepsilon_{j',t-\tau-1}\right)\right. \\
& \left.\cdot\left(\sum_{\tau=0}^{\infty}\lambda^{\tau}\varepsilon_{i,t-\tau-1}+\theta\sum_{\tau=0}^{\infty}\lambda^{\tau}\varepsilon_{j,t-\tau-1}\right)\right]=\theta\frac{(T-2)}{1+\lambda}\sigma_{\varepsilon}^{2}\neq 0 \qquad (78)
\end{aligned}$$

C Proof of Proposition 9

Define $\underline{\boldsymbol{\theta}}_i^o = \boldsymbol{\theta}_i - \boldsymbol{\mu}_\theta$ and $\underline{\alpha}_i^o = \alpha_i - \mu_\alpha = \alpha_i$. Under Assumptions 1-4 and we have $E\left[\underline{\boldsymbol{\theta}}_i^o\underline{\alpha}_i^o\right] = \underline{\boldsymbol{\theta}}_i^o E\left(\underline{\alpha}_i^o\right) = 0$ since $\boldsymbol{\theta}_i$ is non-stochastic. Furthemore, $Var\left[\underline{\boldsymbol{\theta}}_i^o\underline{\alpha}_i^o\right] = E\left[\underline{\boldsymbol{\theta}}_i^o\underline{\alpha}_i^o\underline{\alpha}_i^o\underline{\boldsymbol{\theta}}_i^{o\prime}\right] = \underline{\boldsymbol{\theta}}_i^o\underline{\boldsymbol{\theta}}_i^{o\prime}\sigma_\alpha^2$ and $Cov\left[\underline{\boldsymbol{\theta}}_i^o\underline{\alpha}_i^o, \underline{\boldsymbol{\theta}}_j^o\underline{\alpha}_j^o\right] = \underline{\boldsymbol{\theta}}_i^o\underline{\boldsymbol{\theta}}_j^{o\prime}E\left[\underline{\alpha}_i^o\underline{\alpha}_j^o\right] = 0$ for $i \neq j$. Hence, from a Weak Law of Large Numbers we have

$$\frac{1}{N}\sum_{i=1}^{N}\left[\underline{\boldsymbol{\theta}}_i^o\underline{\alpha}_i^o\right]\xrightarrow{p}0$$

Furthermore, following an approach similar to (17) we have

$$\begin{aligned}
& \sqrt{N}\frac{1}{N}\sum_{i=1}^{N}\left[\underline{\boldsymbol{\theta}}_i\underline{\alpha}_i\right]=\sqrt{N}\frac{1}{N}\sum_{i=1}^{N}\left[\underline{\boldsymbol{\theta}}_i^o-\left(\overline{\boldsymbol{\theta}}-\boldsymbol{\mu}_\theta\right)\right]\left[\underline{\alpha}_i^o-\left(\overline{\alpha}-\mu_\alpha\right)\right]= \\
& \sqrt{N}\frac{1}{N}\sum_{i=1}^{N}\underline{\boldsymbol{\theta}}_i^o\underline{\alpha}_i^o-\left(\overline{\alpha}-\mu_\alpha\right)\sqrt{N}\frac{1}{N}\sum_{i=1}^{N}\underline{\boldsymbol{\theta}}_i^o-\left(\overline{\boldsymbol{\theta}}-\boldsymbol{\mu}_\theta\right)\sqrt{N}\frac{1}{N}\sum_{i=1}^{N}\underline{\alpha}_i^o+\sqrt{N}\frac{1}{N}\sum_{i=1}^{N}\left(\overline{\boldsymbol{\theta}}-\boldsymbol{\mu}_\theta\right)\left(\overline{\alpha}-\mu_\alpha\right) \\
= & \sqrt{N}\frac{1}{N}\sum_{i=1}^{N}\underline{\boldsymbol{\theta}}_i^o\underline{\alpha}_i^o+o_p(1) \qquad (79)
\end{aligned}$$

where the last line follows from the fact that $\left(\overline{\boldsymbol{\theta}}-\boldsymbol{\mu}_\theta\right) = O_p\left(N^{-1/2}\right)$, $\left(\overline{\alpha}-\mu_\alpha\right) = O_p\left(N^{-1/2}\right)$, $N^{-1/2}\sum_{i=1}^{N}\underline{\boldsymbol{\theta}}_i^o = O_p(1)$ and $N^{-1/2}\sum_{i=1}^{N}\underline{\alpha}_i^o = O_p(1)$. In the same way we have $E\left[\underline{\boldsymbol{\theta}}_i^o\underline{\varepsilon}_{i,t}^o\right] =$

$\underline{\boldsymbol{\theta}}_i^o E\left(\underline{\varepsilon}_{i,t}^o\right) = 0$, $Var\left[\underline{\boldsymbol{\theta}}_i^o \underline{\varepsilon}_{i,t}^o\right] = \underline{\boldsymbol{\theta}}_i^o \underline{\boldsymbol{\theta}}_i^{o\prime} \sigma_\varepsilon^2$ and $Cov\left[\underline{\boldsymbol{\theta}}_i^o \underline{\varepsilon}_{i,t}^o, \underline{\boldsymbol{\theta}}_j^o \underline{\varepsilon}_{j,s}^o\right] = \underline{\boldsymbol{\theta}}_i^o \underline{\boldsymbol{\theta}}_j^{o\prime} E\left[\underline{\varepsilon}_{i,t}^o, \underline{\varepsilon}_{j,s}^o\right] = 0$ for $i \neq j$ and $\forall$ t, s. As a result, $\frac{1}{N}\sum_{i=1}^N \left[\underline{\boldsymbol{\theta}}_i^o \underline{\varepsilon}_{i,t}^o\right] \xrightarrow{p} 0$ and $\sqrt{N}\frac{1}{N}\sum_{i=1}^N \left[\underline{\boldsymbol{\theta}}_i \underline{\varepsilon}_{i,t}\right] = \sqrt{N}\frac{1}{N}\sum_{i=1}^N \underline{\boldsymbol{\theta}}_i^o \underline{\varepsilon}_{i,t}^o + o_p(1)$. In addition, $\frac{1}{N}\sum_{i=1}^N \left[\underline{\alpha}_i^o \underline{\varepsilon}_{i,t}^o\right] \xrightarrow{p} 0$ and $\sqrt{N}\frac{1}{N}\sum_{i=1}^N \left[\underline{\alpha}_i \underline{\varepsilon}_{i,t}\right] = \sqrt{N}\frac{1}{N}\sum_{i=1}^N \underline{\alpha}_i^o \underline{\varepsilon}_{i,t}^o + o_p(1)$. With these results in mind, the moment conditions given in (49) are equal to

$$\begin{aligned} &\text{plim}_{N\to\infty}\frac{1}{N}\left[\sum_{i=1}^N \sum_{t=s+1}^T \underline{y}_{j,t-s}\Delta\underline{\upsilon}_{i,t}\right] = \\ = \;& \sum_{t=s+1}^T \text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^N \left[\left(\frac{\underline{\alpha}_j}{1-\lambda} + \underline{\boldsymbol{\theta}}_j' \sum_{\tau=0}^\infty \lambda^\tau \mathbf{f}_{t-\tau} + \sum_{\tau=0}^\infty \lambda^\tau \underline{\varepsilon}_{j,t-\tau-s} + \theta \sum_{\tau=0}^\infty \lambda^\tau \underline{\varepsilon}_{j',t-\tau-s}\right)\right. \\ & \left.\cdot\left(\underline{\boldsymbol{\theta}}_i'\Delta\mathbf{f}_t + \Delta\underline{\varepsilon}_{i,t} + \theta\Delta\underline{\varepsilon}_{j,t}\right)\right] = 0;\; \text{for } s = 2, ..., T-1, \end{aligned} \tag{80}$$

since $\text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^N \left[\underline{\boldsymbol{\theta}}_i \underline{\boldsymbol{\theta}}_j'\right] = 0$ and

$$\begin{aligned} &\text{plim}_{N\to\infty}\frac{1}{N}\left[\sum_{i=1}^N \sum_{t=s+1}^T \underline{y}_{j,t-s}\Delta\underline{y}_{i,t-1}\right] = \\ = \;& \sum_{t=s+1}^T \text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^N \left[\left(\frac{\underline{\alpha}_j}{1-\lambda} + \underline{\boldsymbol{\theta}}_j' \sum_{\tau=0}^\infty \lambda^\tau \mathbf{f}_{t-s-\tau} + \sum_{\tau=0}^\infty \lambda^\tau \underline{\varepsilon}_{j,t-s-\tau} + \theta \sum_{\tau=0}^\infty \lambda^\tau \underline{\varepsilon}_{j',t-s-\tau}\right)\right. \\ & \left.\cdot\left(\frac{\underline{\alpha}_i}{1-\lambda} + \underline{\boldsymbol{\theta}}_i' \sum_{\tau=0}^\infty \lambda^\tau \Delta\mathbf{f}_{t-1-\tau} + \sum_{\tau=0}^\infty \lambda^\tau \Delta\underline{\varepsilon}_{i,t-1-\tau} + \theta \sum_{\tau=0}^\infty \lambda^\tau \Delta\underline{\varepsilon}_{j,t-1-\tau}\right)\right] = -\theta\frac{(T-s)}{1+\lambda}\sigma_\varepsilon^2. \end{aligned} \tag{81}$$

For SYS GMM we have

$$\begin{aligned} &\text{plim}_{N\to\infty}\frac{1}{N}\left[\sum_{i=1}^N \sum_{t=3}^T \Delta\underline{y}_{j,t-1}\underline{\upsilon}_{i,t}\right] = \\ = \;& \sum_{t=3}^T \text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^N \left[\left(\underline{\boldsymbol{\theta}}_j' \sum_{\tau=0}^\infty \lambda^\tau \Delta\mathbf{f}_{t-1-\tau} + \sum_{\tau=0}^\infty \lambda^\tau \Delta\underline{\varepsilon}_{j,t-1-\tau} + \theta \sum_{\tau=0}^\infty \lambda^\tau \Delta\underline{\varepsilon}_{j',t-1-\tau}\right)\right. \\ & \left.\cdot\left(\underline{\alpha}_i + \underline{\boldsymbol{\theta}}_i'\mathbf{f}_t + \underline{\varepsilon}_{i,t} + \theta\underline{\varepsilon}_{j,t}\right)\right] = 0, \end{aligned} \tag{82}$$

and

$$\begin{aligned} &\text{plim}_{N\to\infty}\frac{1}{N}\left[\sum_{i=1}^N \sum_{t=3}^T \Delta\underline{y}_{j,t-1}\underline{y}_{i,t-1}\right] = \\ = \;& \sum_{t=3}^T \text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^N \left[\left(\underline{\boldsymbol{\theta}}_j' \sum_{\tau=0}^\infty \lambda^\tau \Delta\mathbf{f}_{t-1-\tau} + \sum_{\tau=0}^\infty \lambda^\tau \Delta\underline{\varepsilon}_{j,t-1-\tau} + \theta \sum_{\tau=0}^\infty \lambda^\tau \Delta\underline{\varepsilon}_{j',t-1-\tau}\right)\right. \\ & \left.\cdot\left(\underline{\boldsymbol{\theta}}_i' \sum_{\tau=0}^\infty \lambda^\tau \mathbf{f}_{t-1-\tau} + \sum_{\tau=0}^\infty \lambda^\tau \underline{\varepsilon}_{i,t-1-\tau} + \theta \sum_{\tau=0}^\infty \lambda^\tau \underline{\varepsilon}_{j,t-1-\tau}\right)\right] = \theta\frac{(T-2)}{1+\lambda}\sigma_\varepsilon^2 \neq 0 \end{aligned} \tag{83}$$

Table 1. Monte Carlo results, $\widehat{\lambda}$ ($RMSE$)

$N = 400$	$\lambda = 0.5$							$\lambda = 0.7$							$\lambda = 0.9$						
	FE	IV	IV*	DIF	DIF*	SYS	SYS*	FE	IV	IV*	DIF	DIF*	SYS	SYS*	FE	IV	IV*	DIF	DIF*	SYS	SYS*
$z = 1/3$, $T = 6$	.161	1.09	.507	.416	.501	.476	.490	.296	.721	.713	.561	.683	.661	.684	.424	7.88	1.01	.560	.788	.844	.876
	(.364)	(25.6)	(.175)	(.226)	(.190)	(.159)	(.109)	(.426)	(1.14)	(.222)	(.309)	(.219)	(.164)	(.112)	(.495)	(307)	(2.92)	(.557)	(.348)	(.165)	(.112)
$z = 1/2$, $T = 6$	.153	.497	.512	.344	.491	.457	.487	.285	.588	.697	.453	.659	.635	.679	.409	1.20	.058	.430	.726	.821	.869
	(.397)	(2.33)	(.300)	(.339)	(.218)	(.219)	(.125)	(.458)	(8.94)	(.971)	(.450)	(.257)	(.223)	(.129)	(.528)	(14.9)	(3.47)	(.693)	(.422)	(.216)	(.127)
$z = 2/3$, $T = 6$	.140	.483	.570	.258	.467	.436	.481	.269	7.43	.739	.342	.616	.611	.669	.389	.134	.736	.351	.649	.797	.858
	(.442)	(3.34)	(2.22)	(.453)	(.251)	(.276)	(.151)	(.504)	(261)	(2.00)	(.574)	(.310)	(.277)	(.155)	(.575)	(21.0)	(7.59)	(.757)	(.487)	(.262)	(.152)
$z = 3/4$, $T = 6$	.132	.565	.574	.212	.404	.423	.475	.258	-3.80	.511	.293	.578	.599	.661	.375	.612	.591	.312	.598	.788	.850
	(.470)	(1.84)	(1.77)	(.507)	(.276)	(.304)	(.171)	(.533)	(215)	(15.3)	(.621)	(.348)	(.303)	(.175)	(.606)	(6.57)	(5.42)	(.785)	(.526)	(.283)	(.171)
$z = 1/3$, $T = 10$	.312	.506	.498	.439	.503	.484	.491	.471	.707	.699	.609	.696	.672	.688	.622	.915	.903	.697	.850	.856	.882
	(.213)	(.207)	(.113)	(.145)	(.107)	(.115)	(.072)	(.247)	(.250)	(.129)	(.176)	(.113)	(.115)	(.072)	(.293)	(2.55)	(.210)	(.320)	(.162)	(.113)	(.070)
$z = 1/2$, $T = 10$	.303	.518	.501	.385	.494	.472	.489	.459	.723	.701	.531	.680	.653	.694	.608	1.23	.913	.591	.814	.835	.878
	(.244)	(.313)	(.160)	(.224)	(.118)	(.165)	(.083)	(.276)	(.663)	(.191)	(.279)	(.130)	(.166)	(.083)	(.318)	(15.2)	(.457)	(.429)	(.207)	(.153)	(.081)
$z = 2/3$, $T = 10$	.291	.533	.534	.320	.477	.457	.485	.443	.718	.705	.447	.652	.634	.678	.589	.570	.935	.515	.759	.816	.870
	(.283)	(.560)	(.524)	(.306)	(.134)	(.212)	(.102)	(.315)	(1.29)	(2.89)	(.374)	(.158)	(.206)	(.102)	(.355)	(16.4)	(13.7)	(.501)	(.267)	(.191)	(.097)
$z = 3/4$, $T = 10$	.283	.832	.497	.287	.464	.449	.481	.433	.773	.560	.409	.631	.625	.673	.576	.970	1.13	.489	.719	.808	.863
	(.307)	(10.4)	(1.58)	(.345)	(.146)	(.235)	(.118)	(.340)	(1.43)	(4.23)	(.414)	(.179)	(.228)	(.118)	(.379)	(13.8)	(15.1)	(.527)	(.305)	(.208)	(.110)

Notes: FE is the fixed effects estimator, IV is the Anderson-Hsiao estimator and DIF and SYS are the first-differenced and system GMM estimators, proposed by Arellano and Bond (1991) and Blundell and Bond (1998) respectively. DIF and SYS are estimated in two steps and they use y_{it-2} and y_{it-3} as instruments for Δy_{it-1} in the first-differenced equations. Furthermore, SYS uses the optimal weighting matrix (when $\sigma_{\alpha}^2 = 0$), as derived in Windmeijer (2000). The superscript '*' indicates that the corresponding estimator uses instrument(s) with respect to another cross section. The data generating process is given by $y_{it} = \lambda y_{it-1} + \alpha_i + \phi_i f_t + \varepsilon_{it} + \theta\varepsilon_{jt}$, $i = 1, 2, ..., N; t = -48, -47, ..., T$ with $y_{i,-49} = 0$ and the initial 50 observations being discarded. $\alpha_i \sim iidN\left(0, \sigma_{\alpha}^2\right)$, $\varepsilon_{it} \sim iidN\left(0, \sigma_{\varepsilon}^2\right)$, $f_t \sim iidN\left(0, \sigma_f^2\right)$ and $\phi_i \sim iidU\left[-0.25, 0.25\right]$. σ_{α}^2 is chosen to ensure that the impact of the two variance components, α_i and u_{it}, on $var\left(y_{it}\right)$ is held constant. σ_f^2 is normalised to the value of (1) and σ_{ε}^2 is set according to (72), such that it changes according to the proportion of σ_u^2 attributed to the factor structure. λ alternates between 0.5, 0.7 and 0.9. All experiments are based on 2,000 replications.

Table 2. Monte Carlo results, $\widehat{\lambda}$ ($RMSE$)

$N=800$	$\lambda=0.5$							$\lambda=0.7$							$\lambda=0.9$						
	FE	IV	IV*	DIF	DIF*	SYS	SYS*	FE	IV	IV*	DIF	DIF*	SYS	SYS*	FE	IV	IV*	DIF	DIF*	SYS	SYS*
$z=1/3$, $T=6$	.162	.534	.501	.427	.508	.478	.496	.295	.750	.703	.569	.701	.663	.693	.424	.968	.885	.548	.867	.846	.888
	(.365)	(.402)	(.120)	(.231)	(.144)	(.159)	(.076)	(.427)	(.795)	(.146)	(.312)	(.164)	(.162)	(.080)	(.495)	(6.07)	(1.94)	(.569)	(.250)	(.163)	(.085)
$z=1/2$, $T=6$	.153	.563	.517	.355	.506	.459	.495	.283	.011	.720	.452	.693	.637	.690	.409	.185	.937	.407	.823	.828	.883
	(.401)	(.717)	(.164)	(.353)	(.162)	(.226)	(.088)	(.462)	(21.2)	(.204)	(.474)	(.187)	(.228)	(.093)	(.527)	(37.8)	(.527)	(.717)	(.303)	(.217)	(.098)
$z=2/3$, $T=6$	.140	.116	.520	.266	.499	.440	.492	.266	.323	.709	.335	.674	.612	.684	.388	.656	.786	.322	.768	.801	.875
	(.450)	(20.1)	(.498)	(.472)	(.190)	(.291)	(.108)	(.512)	(18.9)	(.709)	(.584)	(.232)	(.292)	(.114)	(.578)	(6.69)	(2.90)	(.788)	(.375)	(.276)	(.120)
$z=3/4$, $T=6$	.131	.555	.546	.219	.489	.430	.488	.254	5.75	.771	.281	.653	.600	.679	.374	1.16	.953	.296	.716	.792	.867
	(.481)	(6.64)	(1.55)	(.523)	(.211)	(.323)	(.124)	(.544)	(7.53)	(4.23)	(.638)	(.263)	(.323)	(.131)	(.611)	(13.0)	(4.51)	(.810)	(.430)	(.293)	(.138)
$z=1/3$, $T=10$	.313	.505	.501	.442	.505	.486	.499	.468	.741	.706	.614	.700	.674	.697	.618	1.05	.898	.693	.885	.851	.897
	(.216)	(.199)	(.089)	(.148)	(.067)	(.119)	(.054)	(.252)	(.531)	(.094)	(.179)	(.076)	(.119)	(.047)	(.299)	(2.91)	(.137)	(.326)	(.109)	(.117)	(.049)
$z=1/2$, $T=10$	.305	.532	.501	.391	.501	.474	.498	.455	.719	.714	.534	.699	.654	.698	.601	1.03	.921	.582	.880	.837	.891
	(.249)	(.472)	(.123)	(.229)	(.078)	(.169)	(.056)	(.279)	(.610)	(.153)	(.281)	(.072)	(.169)	(.057)	(.323)	(12.1)	(.236)	(.436)	(.154)	(.159)	(.059)
$z=2/3$, $T=10$	.290	.542	.512	.325	.501	.460	.497	.441	.771	.703	.449	.681	.636	.692	.585	.649	.891	.502	.829	.819	.886
	(.289)	(1.26)	(.211)	(.311)	(.091)	(.219)	(.071)	(.319)	(.994)	(.451)	(.380)	(.101)	(.213)	(.073)	(.358)	(10.5)	(.879)	(.521)	(.211)	(.201)	(.069)
$z=3/4$, $T=10$	.280	.741	.525	.290	.493	.451	.492	.430	.815	.731	.404	.672	.629	.689	.572	.961	.939	.473	.765	.810	.878
	(.313)	(3.41)	(.576)	(.365)	(.101)	(.243)	(.081)	(.347)	(2.39)	(1.74)	(.423)	(.124)	(.232)	(.082)	(.384)	(11.2)	(2.22)	(.541)	(.252)	(.219)	(.079)

Notes: FE is the fixed effects estimator, IV is the Anderson-Hsiao estimator and DIF and SYS are the first-differenced and system GMM estimators, proposed by Arellano and Bond (1991) and Blundell and Bond (1998) respectively. DIF and SYS are estimated in two steps and they use y_{it-2} and y_{it-3} as instruments for Δy_{it-1} in the first-differenced equations. Furthermore, SYS uses the optimal weighting matrix (when $\sigma_\alpha^2 = 0$), as derived in Windmeijer (2000). The superscript '*' indicates that the corresponding estimator uses instrument(s) with respect to another cross section. The data generating process is given by $y_{it} = \lambda y_{it-1} + \alpha_i + \phi_i f_t + \varepsilon_{it} + \theta\varepsilon_{jt}$, $i = 1, 2, ..., N; t = -48, -47, ..., T$ with $y_{i,-49} = 0$ and the initial 50 observations being discarded. $\alpha_i \sim iidN\left(0, \sigma_\alpha^2\right)$, $\varepsilon_{it} \sim iidN\left(0, \sigma_\varepsilon^2\right)$, $f_t \sim iidN\left(0, \sigma_f^2\right)$ and $\phi_i \sim iidU\left[-0.25, 0.25\right]$. σ_α^2 is chosen to ensure that the impact of the two variance components, α_i and u_{it}, on $var\left(y_{it}\right)$ is held constant. σ_f^2 is normalised to the value of (1) and σ_ε^2 is set according to (72), such that it changes according to the proportion of σ_u^2 attributed to the factor structure. λ alternates between 0.5, 0.7 and 0.9. All experiments are based on 2,000 replications.