Open access • Journal Article • DOI:10.1080/07474938.2011.611458

Cross-Sectional Dependence in Panel Data Analysis — Source link

Vasileios Sarafidis, Tom Wansbeek

Institutions: University of Sydney

Published on: 12 Mar 2012 - Econometric Reviews (Taylor & Francis Group)

Cross-sectional Dependence in Panel Data Analysis

Sarafidis, Vasilis and Wansbeek, Tom

The University of Sydney

February 2010

Cross-sectional Dependence in Panel Data Analysis

Vasilis Sarafidis* University of Sydney

Tom Wansbeek† University of Groningen

February 2010

Abstract

This paper provides an overview of the existing literature on panel data models with error cross-sectional dependence. We distinguish between spatial dependence and factor structure dependence and we analyse the implications of weak and strong cross-sectional dependence on the properties of the estimators. We consider estimation under strong and weak exogeneity of the regressors for both T fixed and T large cases. Available tests for error cross-sectional dependence and methods for determining the number of factors are discussed in detail. The finite-sample properties of some estimators and statistics are investigated using Monte Carlo experiments.

Key words: Panel data, Cross-sectional dependence, Spatial dependence, Factor structure, Strong/Weak exogeneity.

JEL Classification: C33; C50.

1 Introduction

The analysis of longitudinal data is common across many fields of research. In econometrics, the topic is invariably called panel data analysis. Over the last fourty years, it has grown into a major subfield of econometrics. Traditionally, the focus has been on panels involving a large number of individual units $i = 1, \ldots, N$, with a few observations over time, $t = 1, \ldots, T$.[1] Often, the data come from surveys where a large group of people or households has been followed over a few years. The National Longitudinal Surveys of Labor Market Experience and the University of Michigan's Panel Study of Income Dynamics are prominent examples. One of the primary reasons for collecting these data has been to overcome aggregation problems that arise with time series data in modelling the behaviour of heterogeneous agents on the basis of the "representative agent" assumption. More recently, considerable interest has also been directed to panels where the cross-sectional and time series dimensions are of similar magnitude. For

*Corresponding author. Faculty of Economics and Business, University of Sydney, NSW 2006, Australia. Tel: +61-2-9036 9120; E-mail: vasilis.sarafidis@sydney.edu.au.

†University of Groningen, P.O.Box 800, 9700 AV Groningen, The Netherlands. Tel: +31-50-363-8339; E-mail: t.j.wansbeek@rug.nl.

[1] An exception to this is the seemingly unrelated regression (SUR) approach due to Zellner (1962); see Section 4.1.

instance, the Penn-World tables cover several countries over relatively long periods and the main focus of study lies in cross-country economic, social and political comparisons.

One major issue that inherently arises in every panel data study with potential implications on parameter estimation and inference is the possibility that the individual units are interdependent. In fact, this notion of 'between group' dependence is familiar in the social sciences since the 1930's, i.e. well before the emergence of panel data econometrics. In specific, Stephan (1934, pg. 165) argues that "in dealing with social data, we know that by virtue of their very social character, persons, groups and their characteristics are interrelated and not independent". Neprash (1934, pg. 168) asserts that "the correlation of spatially distributed variables must be accepted with severe limitations of interpretation. The data involved violate two important conditions of sound application of correlation and sample techniques − namely, the independence of the units of which the traits are measured, and the homogeneity of distribution of the traits within a given area". Fisher in his "Design of Experiments" book (1935, pg. 66) claims that "patches in close proximity are commonly more alike ... than those further apart". Later on, in the field of economic geography Tobler (1970, pg. 236) invoked his 'first law of geography': "everything is related to everything else; but near things are more related than distant things".

Naturally, the issue of how to characterise cross-sectional dependence has attracted considerable attention among researchers over the years. Perhaps the earliest methodology put forward to deal with this issue was the spatial approach. Spatial models were developed primarily for cross-sectional data using a concept of a distance metric, which allowed formulating models with a structure similar to that provided by the time index in time series. The concept of 'economic distance' eventually allowed the use of spatial models in certain economic applications as well, mainly drawn from regional science and urban economics. The increasing availability of panel data during the last decades gave rise to new possibilities in characterising error cross-sectional dependence. A prominent alternative to the spatial approach is the factor structure approach, which assumes that the disturbance term contains a finite number of unobserved factors that influence each individual separately. Initially, the inferential theory for factor models was developed for cases where one dimension was fixed and the other went to infinity. Recently, this theory has been extended for large panels, where both dimensions can go to infinity; see Bai (2003) and Bai and Ng (2002).

In this paper we attempt to provide an overview of some of the recent developments that have been made in the field and link them to earlier related work. Realistically, it is impossible to do justice to the voluminous and still rapidly growing literature of panel data models with error cross-sectional dependence. In what follows, we shall focus on stationary models with a static error structure. Throughout, we try to employ a unified notation. Some of the issues discussed in the paper are also briefly mentioned by Baltagi and Pesaran (2007), which is an introduction to a special issue of the *Journal of Applied Econometrics*, and Hsiao (2007) in a paper reviewing the state of the art and the current issues in panel data analysis. There is a growing literature on dynamic factor models (see e.g. Forni and Lippi 2001, and Forni, Hallin, Lippi and Reichlin,

2000), which is mainly concerned with extraction of common components of economic variables rather than with estimation of structural (regression) parameters; therefore, we do not review this in the present paper.[2] There are also important developments in non-stationary panel data models with error cross-sectional dependence (see e.g. Bai and Ng, 2004, Moon and Perron, 2004, and Phillips and Sul, 2003) for which a succint overview has already been provided by Hurlin and Mignon (2004) and Breitung and Pesaran (2008). The main theoretical results for large dimensional factor analysis using principal components are reviewed in an excellent survey by Bai and Ng (2008).

The set-up of the paper is as follows. The next section describes the spatial and factor structure approaches. Section 3 links these approaches with the concepts of weak and strong cross-sectional dependence. Sections 4 and 5 analyse estimation under strong and weak exogeneity respectively. Section 6 discusses available tests of error cross-sectional dependence and methods for determining the number of factors. We conclude by indicating a number of topics for future research.

In what follows we adopt the conventional mathematics notation where capital letters denote matrices and small letters in bold denote vectors.

2 Spatial Dependence and Factor Structure

Consider the following panel data model:

$$y_{it} = \boldsymbol{\beta}'\mathbf{x}_{it} + \eta_i + \upsilon_{it}, \;\; i = 1, 2, \ldots, N, \, t = 1, 2, \ldots, T, \tag{1}$$

where y_{it} is the observation on the dependent variable for individual i at time t, $\mathbf{x}_{it}$ is a column vector of regressors with dimension K, $\boldsymbol{\beta}$ is the corresponding parameter vector of fixed coefficients, η_i is an individual-specific time-invariant unobserved effect, and υ_{it} is the error component that may be cross-sectionally correlated. The latter would imply that the following is true:

$$\text{Cov}(\upsilon_{it}, \upsilon_{jt}) \neq 0 \text{ for some } t \text{ and some } i \neq j, \tag{2}$$

where the number of possible pairings $(\upsilon_{it}, \upsilon_{jt})$ increases with N.

We view the presence of cross-sectional dependence in the error term as a consequence of model mispecification. In other words, had the model been specified correctly, cross-sectional dependence would had been taken into account and the resulting disturbance would be purely idiosyncratic and uncorrelated across individuals. The advantage of this approach is that it makes the distinction between strongly and weakly exogenous regressors clear. In particular, let υ'_{it} be the disturbance term of a *correctly specified* model. We call the vector $\mathbf{x}_{it}$ strongly exogenous if υ'_{it} is mean-independent of its past, present and future, i.e.

$$E\left(\upsilon'_{it}|\mathbf{x}_{i1}, ..., \mathbf{x}_{iT}\right) = 0, \tag{3}$$

[2] For dynamic factor models see also the special issue in the *Journal of Econometrics*, Vol. 119, 2004.

On the other hand, the vector $\mathbf{x}_{it}$ is weakly exogenous if υ'_{it} is mean-independent of its past and present, so

$$E\left(\upsilon'_{it}|\mathbf{x}_{i1},...,\mathbf{x}_{it}\right)=0. \tag{4}$$

Ignoring cross-sectional dependence may affect the first-order properties (unbiasedness, consistency) of standard panel estimators because even if $\mathbf{x}_{it}$ is strongly, or weakly, exogenous with respect to υ'_{it}, it may not be so with respect to υ_{it}. In addition, even if the first-order properties of these estimators remain unaffected, the presence of error cross-sectional dependence may largely reduce the extent to which they can provide efficiency gains over estimating (1) using, say, OLS for each individual i. In a sense, if all individuals behave similarly there is little gain to be obtained by looking at more than one of them.

Unfortunately, modelling general forms of cross-sectional dependence is not a straightforward task. In specific, contrary to a time series model, where it is natural to specify the correlations between the disturbances to be functions of distance measured by time, in a cross-section there is no such natural ordering of the observations. To deal with this issue, the panel data literature has mainly adopted two different approaches to modelling error cross-sectional dependence, the spatial approach and the factor structure approach.

The former assumes that the structure of cross-sectional dependence is a function of an immutable distance measure, defined according to a pre-specified metric. In economic applications, spatial techniques are adapted using alternative measures of "economic distance" (Conley, 1999), or "policy and social distance" (Conley and Topa, 2002). A number of different spatial processes have been proposed in the literature to model cross-sectional dependence, the most popular of which have been the Spatial Moving Average (SMA), Spatial Auto-Regressive (SAR) and Spatial Error Components (SEC) processes. These can be defined as follows:

$$\begin{aligned}
SMA;\ \upsilon_{it} &= \lambda\sum_{j=1}^{N}w_{ij}\varepsilon_{jt}+\varepsilon_{it},\\
SAR;\ \upsilon_{it} &= \lambda\sum_{j=1}^{N}w_{ij}\upsilon_{jt}+\varepsilon_{it},\\
SEC;\ \upsilon_{it} &= \lambda\sum_{j=1}^{N}w_{ij}\xi_{jt}+\varepsilon_{it},
\end{aligned} \tag{5}$$

where w_{ij} is the i-specific spatial weight attached to individual j, typically determined before estimation, ε_{it} is white noise, and for SEC ξ_{jt} denotes a zero mean random component, uncorrelated with ε_{it} and $\mathbf{x}_{it}$. These spatial models can be estimated using a generalized method of moments (GMM) approach (see e.g. Kapoor, Kelejian and Prucha, 2007; Kelejian and Prucha, 2009), or a method based on maximum likelihood (e.g. Lee, 2004).

The factor structure approach assumes the presence of an unobserved common component in the disturbance which is a linear combination of a fixed number of factors (e.g.

Lawley and Maxwell, 1971; Goldberger, 1972, and Jöreskog and Goldberger, 1975). In this case the error can be written as

$$\upsilon_{it} = \boldsymbol{\lambda}_i' \boldsymbol{\phi}_t + \varepsilon_{it}, \tag{6}$$

where $\boldsymbol{\phi}_t = \left(\phi_{1t}, ..., \phi_{M_0 t}\right)'$ denotes an $M_0 \times 1$ vector of unobserved factors, $\boldsymbol{\lambda}_i = (\lambda_{1i}, ..., \lambda_{M_0 i})'$ is an $M_0 \times 1$ vector of factor loadings[3] and ε_{it} is a purely idiosyncratic component such that $E(\varepsilon_{it}) = 0$ and

$$E(\varepsilon_{it}\varepsilon_{js}) = \begin{cases} \sigma_\varepsilon^2 \text{ for } t = s \text{ and } i = j, \\ 0 \text{ otherwise.} \end{cases}$$

This formulation generates a taxonomy of models depending on whether the $\boldsymbol{\lambda}_i$ and/or $\boldsymbol{\phi}_t$ are correlated with $\mathbf{x}_{it}$ or not. The relative size of N and T is also important. As an example, suppose that (1) combined with (6) is used to model the returns to education, where, as is typical in micro-econometric panels, N is large and T small; in this case the vector of covariates, $\mathbf{x}_{it}$, may include variables like education, experience, and tenure of individual i with the same employer, η_i may capture innate ability (which is constant by definition) and λ_i may reflect time-varying productivity of individual i. Both η_i and λ_i are likely to be correlated with $\mathbf{x}_{it}$.[4] For small T, the factors can be treated as time-specific parameters that reflect how productivity varies over time. Another example can be drawn from the estimation of production and cost functions. For a cost function the vector $\mathbf{x}_{it}$ represents input prices and output, η_i may capture cost efficiency of firm i, ϕ_t may reflect changes in the regulatory regime over time, with λ_i, the impact on firm i, depending on the size of the firm in the market, on financial constraints, technology and other considerations. In this case both ϕ_t and λ_i are likely to be correlated with input prices and output.[5] Depending upon the size of N and T, as well as on the properties of $\mathbf{x}_{it}$, different methods can be used to estimate these models, as we shall see in the following sections.

Sarafidis (2009) shows that all spatial processes can be expressed in the following form:

$$\upsilon_{it} = \left(\boldsymbol{\lambda}_i \odot \mathbf{w}_i\right)' \boldsymbol{\phi}_t + \varepsilon_{it},$$

by setting $M_0 = N$ and imposing appropriate zero restrictions on $\mathbf{w}_i$ and homogeneity restrictions on $\boldsymbol{\lambda}_i$.[6] This may be useful because spatial dependence can be viewed in this case as a special form of factor structure dependence, in which one may think of the unobserved components, $\boldsymbol{\phi}_t$ as shocks, the impact of which is either 'global' (factors) or 'local' (spatially correlated components).

[3] There is large variation in the literature regarding the notation used for factor models. Following Kiviet and Sarafidis (2009), our choice is based on the following reasoning: we use Greek symbols for unobserved variables/parameters and Latin symbols for observed ones. Consequently, we use ε (epsilon) to denote the purely idiosyncratic error component, ϕ (phi) to denote the factors and λ (lamda) to denote their loadings. Similarly, η (eta) is used to denote the individual-specific effect.

[4] The argument here would be that it is the most able and productive individuals who embark on higher education, all other things being equal, such as equal opportunities and so on.

[5] Many other examples are provided by Ahn, Lee and Schmidt (2001) and Bai (2009).

[6] For the factor structure $\mathbf{w}_i = \boldsymbol{\iota}_N$, where $\boldsymbol{\iota}_N$ is a $T \times 1$ column vector of ones.

3 Weak and Strong Cross-sectional Dependence

The spatial approach and the factor structure approach imply different degree of error cross-sectional dependence. However, there is no unique definition of what is 'weak' and what is 'strong' dependence in the literature. In particular, let $\left\{v_i^t,\, i \geq 1\right\}$ be the scalar sequence v_{1t}, v_{2t}, v_{3t},..., and notice that there are T such scalar sequences, for $t = 1, ..., T$. Weak dependence can be defined in the following ways:

Definition 1 *(Chudik, Pesaran and Tosetti, 2009) The double-indexed sequence* $\{v_{it},\, i \geq 1,\, t \geq 1\}$ *is said to be weakly dependent at a given point in time if its weighted average, conditional on the information set available in the previous period,* I_{t-1}*, converges to its expectation in quadratic mean, as* $N \to \infty$ *for all weights that satisfy certain 'granularity conditions'.*[7]

Definition 2 *(Sarafidis, 2009) The double-indexed sequence* $\{v_{it},\, i \geq 1,\, t \geq 1\}$ *is said to be cross-sectionally weakly correlated if, for each* i *and* $j > 0$, $\left\{\gamma_{i,j}^{t,s},\, j \neq i\right\}$ *is absolutely summable, that is,*

$$\sum_{j \neq i} \left|\gamma_{i,j}^{t,s}\right| < \infty, \text{ for all } t \text{ and } s, \tag{7}$$

where $\gamma_{i,j}^{t,s} = Cov(v_{it}, v_{js} | \Upsilon_{i,j})$*, and* $\Upsilon_{i,j}$ *denotes the conditioning set of all time-invariant characteristics of individuals* i *and* j*.*

Neither of these definitions requires the process to be covariance stationary.[8] Furthermore, both definitions imply that spatial dependence is a weak form of dependence. This comes directly from the standard assumption employed in spatial processes that the row and column sums of the weighting matrix $\mathbf{W} = [w_{ij}]$ satisfy a uniform boundedness condition.[9] This can be stated as follows[10]:

$$\sum_{i=1}^{N} |w_{ij}| \leq B_w < \infty \;\forall\; j \text{ and } \sum_{j=1}^{N} |w_{ij}| \leq B_w < \infty \;\forall\; i. \tag{8}$$

It is important to emphasise that spatial dependence in the residuals does not affect the first-order properties (consistency) of standard panel data estimators. In particular, it is straightforward to show that the mean-independence conditions (3) and (4) are preserved when the error term of the mispecified model, v_{it}, follows either one of the three processes in (5). Therefore, the potential gains from modelling spatial dependence arise with respect to estimation efficiency and the validity of inference.

[7] These conditions ensure that the weights are not dominated by a few individuals.

[8] For alternative definitions of weak cross-sectional dependence that require covariance stationarity see Forni and Lippi (2001).

[9] Notice, however, that uniform boundedness is not actually necessary for weak dependence; see Sarafidis (2009).

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

The difference between the two definitions provided above lies mainly in factor structures and it can be illustrated through an example. Consider a single-factor error process

$$\begin{aligned}&\upsilon_{it} = \lambda_i \phi_t + \varepsilon_{it}, \text{ where the following assumptions are made:}\\&E(\phi_t) = E(\varepsilon_{it}) = 0,\ E\left(\phi_t^2\right) = \sigma_\phi^2,\ E\left(\varepsilon_{it}^2\right) = \sigma_\varepsilon^2;\\&E(\phi_t \varepsilon_{it}) = 0,\ E(\phi_t \phi_s) = E(\varepsilon_{ts}\varepsilon_{it}) = 0 \text{ for } s \neq t.\end{aligned} \tag{9}$$

According to Definition 1, the error process is weakly dependent so long as

$$\lim_{N\to\infty} N^{-1} \sum_{i=1}^{N} \lambda_i = 0. \tag{10}$$

This is because $E(\upsilon_{it}) = E(\lambda_i \phi_t + \varepsilon_{it}) = \lambda_i E(\phi_t) + E(\varepsilon_{it}) = 0$ and

$$N^{-1} \sum_{i=1}^{N} \upsilon_{it} = \left[\phi_t N^{-1} \sum_{i=1}^{N} \lambda_i + N^{-1} \sum_{i=1}^{N} \varepsilon_{it}\right] \xrightarrow{p} 0.$$

On the other hand, according to Definition 2 the error process (9) is not weakly dependent because $\gamma_{i,j}^{t,t} = \text{Cov}(\upsilon_{it}, \upsilon_{jt} | \lambda_i, \lambda_j) = \lambda_i \lambda_j \sigma_\phi^2 \neq 0$ and therefore $\sum_{j\neq i} \left|\gamma_{i,j}^{t,t}\right|$ is unbounded. Intuitively, since all individuals are subject to the same shock, ϕ_t, the sum of the absolute conditional covariances between individual disturbances grows with N regardless of whether $N^{-1}\sum_{i=1}^{N} \lambda_i \to 0$ or not. As a result, all factor structures, provided they are non-degenerate, imply strong dependence under Definition 2 but not under Definition 1.

Definition 1 has an encompassing property in the sense that any factor structure with $E\left(\boldsymbol{\lambda}_i' \boldsymbol{\phi}_t\right) = \mathbf{0}$ reduces to a weakly dependent process when the observations are expressed in terms of deviations from time-specific averages. Specifically, suppose that the error term follows process (9). Averaging υ_{it} over all i for each t and subtracting yields

$$\underline{\upsilon}_{it} = \underline{\lambda}_i \phi_t + \underline{\varepsilon}_{it}, \tag{11}$$

where $\underline{\upsilon}_{it} = \upsilon_{it} - \overline{\upsilon}_{.t}$ with $\overline{\upsilon}_{.t} = N^{-1}\sum_{i=1}^{N} \upsilon_{it}$, and so on. Therefore, even if $N^{-1}\sum_{i=1}^{N} \lambda_i \nrightarrow 0$, thus violating (10), we have

$$\lim_{N\to\infty} N^{-1} \sum_{i=1}^{N} \underline{\lambda}_i \to 0 \tag{12}$$

by construction. As a result, in a linear regression model υ_{it} can always be transformed such that it becomes weakly dependent.

Definition 2 has the property that weak dependence is preserved by product. In other words, the product of two (or more) weakly dependent processes is also weakly dependent.[11] As we shall see later on, this is important in panel data models with

[11] See Sarafidis (2009) for a proof.

weakly exogenous regressors. Under Definition 2, weak dependence is not preserved by product. For instance, consider the product between υ_{it} and υ_{is}, as defined in (9). This product then involves $N^{-1}\sum_{i=1}^{N}\lambda_i^2$, which is not converging to zero.

One final remark. Neither Definition 1 implies that weak error cross-sectional dependence cannot affect on first-order properties of standard panel data estimators, nor Definition 2 implies that strong dependence has always an adverse effect on the first-order properties of these estimators. For the former case, one can think of a regression model with a single covariate, x_{it}, and one-factor error structure in which λ_i and ϕ_t are both random with zero mean and x_{it} contains $\lambda_i^*\phi_t$ with $\text{Cov}(\lambda_i, \lambda_i^*) \neq 0$. For the latter case one can think of a single-factor error process in which neither λ_i nor ϕ_t are correlated with $\mathbf{x}_{it}$.

The literature on spatial dependence is rich and is developing rapidly. Notwithstanding, the remainder of this paper focuses mainly on residual factor structures. This is partly because of the generality of this approach relative to spatial dependence, in that it does not require a priori the specification of a distance metric, which may or may not be appropriate in certain economic applications. Furthermore, modelling a factor structure is likely to sweep out the spatial correlations as well (see Pesaran and Tosetti, 2009). Finally, notice that in the spatial dependence case standard panel data estimators can still be used to make robust inferences on the parameters. In particular, one may employ spectral density matrix estimation techniques of the sort popularised in econometrics by Newey and West (1987), valid for large T and fixed N (see Arellano, pg. 19, for details) or the methods of Driscoll and Kraay (1998) and Pesaran and Tosetti (2009), valid for large N and large T.

4 Estimation Under Strict Exogeneity

4.1 The Seemingly Unrelated Regressions Approach

By far the most classic model of error cross-sectional dependence in econometrics is the Seemingly Unrelated Regressions (SUR) approach, due to Zellner (1962).[12] In the form where the same regressors enter the model for all individuals the model is

$$y_{it} = \boldsymbol{\beta}_i'\mathbf{x}_{it} + \eta_i + \upsilon_{it}, \tag{13}$$

where both $\boldsymbol{\beta}_i$ and η_i are treated as fixed. There are two underlying assumptions behind this approach. Firstly, $E(\upsilon_{it}|\mathbf{x}_{it}) = 0$, that is, all regressors remain strongly exogenous in the mispecified model. Therefore, neglecting cross-sectional dependence does not affect the first-order properties of standard panel estimators. Secondly, the asymptotics are fixed N and $T \to \infty$. These assumptions combined imply that the error covariance matrix, $\Sigma = [\sigma_{ij}]$, can be left unrestricted, i.e. there is no need to impose a factor structure in the residuals. The SUR approach leads to a feasible

[12] A thorough review of the large literature accumulated on SUR can be found in Srivastava and Dwivedi (1979) and Srivastava and Giles (1987). A survey of more recent developments is provided by Fiebig (2001) and Moon and Perron (2006).

GLS estimator, in which OLS is used at first-stage for each individual-specific equation to obtain consistent estimates of the parameters, including the $N(N+1)/2$ distinct entries in the error covariance matrix. The resulting estimator of $\boldsymbol{\beta}_i$ is consistent and asymptotically efficient. When T is only slightly greater than N, the estimate of Σ may be ill-conditioned. Kontoghiorghes and Clarke (1995) propose an numerical procedure for estimating a SUR model that avoids the difficulty in directly computing the inverse of the estimated covariance matrix.

When $N > T$, the least-squares estimate of the general, unstructured error covariance matrix, $\widehat{\Sigma}$, with typical entry $T^{-1}\Sigma_{t=1}^{T}\widehat{v}_{it}\widehat{v}_{jt}$, is singular. This implies that the standard SUR estimator is not feasible. Robertson and Symons (2007) propose imposing a factor structure in the residuals, according to (6), and then they estimate the residual covariance matrix using maximum likelihood. Therefore, their method allows SUR estimation of panel models by providing a full-rank estimator of the error covariance matrix when the usual estimate is rank-deficient.

An alternative approach, valid for fixed T, is to impose a factor structure, as in (6), and use a GMM estimator that makes use of the second-order moment restrictions imposed on the covariance matrix of $\Phi\boldsymbol{\lambda}_i+\boldsymbol{\varepsilon}_i$, where $\Phi = (\boldsymbol{\phi}_1, ..., \boldsymbol{\phi}_T)'$ and $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, ..., \varepsilon_{iT})'$; see e.g. Wansbeek and Meijer (2000). This method requires that $\boldsymbol{\lambda}_i$ is random with zero mean, and is uncorrelated with the covariates. Under these assumptions, the model becomes a particular case of a so-called structural equation model (SEM), which can be handled routinely by softwares like LISREL, EQS, AMOS, MX, Mplus, MECOSA, RAMONA, LINCS and PROC CALIS. As Wansbeek and Meijer (2007) indicate, the availability of these programs is not generally known among econometricians, leading sometimes to papers dealing with special cases of a SEM, which in fact are not needed.

4.2 The Principal Components Approach

Coakley, Fuertes and Smith (2002) propose an arguably simpler estimation approach, based on residual principal components analysis. Specifically, they estimate individual OLS regressions of y_{it} on $\mathbf{x}_{it}$ to extract the $M_0 < N$ principal components of $\widehat{\Sigma} = \left[T^{-1}\Sigma_{t=1}^{T}\widehat{v}_{it}\widehat{v}_{jt}\right]$ as proxies for the latent factors. In the second stage these proxies are used as additional regressors with individual-specific coefficients. Similarly to the SUR approach, the estimator requires that the unobserved factors are uncorrelated with $\mathbf{x}_{it}$. Otherwise, the first-stage estimate of Σ is inconsistent, thus invalidating the properties of the estimators.

An alternative estimator based on principal components analysis that does not depend on an initial estimate of $\boldsymbol{\beta}$ is given by

$$\widehat{\boldsymbol{\beta}}_{PC} = \left[\sum_{i=1}^{N}\widetilde{X}_i' M_{\widehat{\Phi}}\widetilde{X}_i\right]^{-1}\sum_{i=1}^{N}\widetilde{X}_i' M_{\widehat{\Phi}}\widetilde{\mathbf{y}}_i, \tag{14}$$

where $\widetilde{X}_i = Q_T X_i$, $Q_T = I_T - T^{-1}\boldsymbol{\iota}_T\boldsymbol{\iota}_T'$, the matrix that transforms the observations in terms of deviations from individual-specific averages to remove η_i, $\boldsymbol{\iota}_T$ is a T column vector

of ones, $X_i = (\mathbf{x}_{i1}, ..., \mathbf{x}_{iT})'$, $\widetilde{\mathbf{y}}_i = Q_T \mathbf{y}_i$, $\mathbf{y}_i = (y_{i1}, ..., y_{iT})'$, $M_{\widehat{\Phi}} = I_T - \widehat{\Phi}\left(\widehat{\Phi}'\widehat{\Phi}\right)^{-1}\widehat{\Phi}'$ and $\widehat{\Phi} = \left(\widehat{\boldsymbol{\phi}}_1, ..., \widehat{\boldsymbol{\phi}}_T\right)'$, a $T \times M_0$ matrix, which is computed as the vector of principal components extracted from the covariates, $\mathbf{x}_{it}$. Intuitively, the idea is to sweep out the factors that are common between the y_{it} and $\mathbf{x}_{it}$ processes by orthogonalising the data prior to estimation. A similar estimator is proposed by Kapetanios and Pesaran (2007) except that $\widehat{\Phi}$ is computed from the vector of principal components extracted from $\mathbf{z}_{it} = (y_{it}, \mathbf{x}'_{it})'$. This can be useful when different factors hit the y and x processes. The rate of convergence of $\widehat{\Phi}$ is $\min\left\{\sqrt{N}, T\right\}$. Therefore for fixed T, $\widehat{\Phi}$ is not consistent, in general, unless the purely idiosyncratic component is serially uncorrelated and homoskedastic (see Bai, 2003).

Bai (2009) proposes an iterative principal components (IPC) estimator such that $\left(\widehat{\boldsymbol{\beta}}_{IPC}, \widehat{\boldsymbol{\phi}}^T\right)$ is the solution to (14) and the following non-linear equation:

$$\left[\frac{1}{NT}\sum_{i=1}^{N}\left(\widetilde{\mathbf{y}}_i - \widetilde{X}_i\widehat{\boldsymbol{\beta}}_{IPC}\right)\left(\widetilde{\mathbf{y}}_i - \widetilde{X}_i\widehat{\boldsymbol{\beta}}_{IPC}\right)'\right]\widehat{\Phi} = \widehat{\Phi}\widehat{V}_{NT}, \tag{15}$$

where $\widehat{V}_{NT}$ is a diagonal matrix that consists of the M_0 largest eigenvalues of the matrix $\frac{1}{NT}\sum_{i=1}^{N}\left(\widetilde{\mathbf{y}}_i - \widetilde{X}_i\widehat{\boldsymbol{\beta}}_{IPC}\right)\left(\widetilde{\mathbf{y}}_i - \widetilde{X}_i\widehat{\boldsymbol{\beta}}_{IPC}\right)'$, arranged in decreasing order. Therefore, given $\widehat{\Phi}$ one can estimate $\boldsymbol{\beta}$ and given $\boldsymbol{\beta}$ one can estimate $\widehat{\Phi}$. The solution can simply be obtained by iteration. The resulting estimator is consistent and asymptotically normal as $(N,T) \overset{jointly}{\to} \infty$, i.e. $\sqrt{NT}\left(\widehat{\boldsymbol{\beta}}_{IPC} - \boldsymbol{\beta}\right) \overset{d}{\to} N\left(\mathbf{0}, \Sigma_{IPC}\right)$, where Σ_{IPC} is the asymptotic variance of $\sqrt{NT}\left(\widehat{\boldsymbol{\beta}}_{IPC} - \boldsymbol{\beta}\right)$. For fixed T the estimator is inconsistent under serial correlation or heteroskedasticity[13]

4.3 The Common Correlated Effects Estimator

In practice, M_0 is most likely to be unknown and so it needs to be estimated.[14] Pesaran (2006) proposes an alternative which does not require estimating the number of latent factors and is valid even when $\boldsymbol{\phi}_t$ is correlated with $\mathbf{x}_{it}$. His 'Common Correlated Effects' (CCE) estimator is given by

$$\widehat{\boldsymbol{\beta}}_{i,CCE} = \left[X_i'\bar{M}_w X_i\right]^{-1} X_i'\bar{M}_w\mathbf{y}_i, \tag{16}$$

where $\bar{M}_w = I_T - \bar{Z}_w\left(\bar{Z}_w'\bar{Z}_w\right)^{-1}\bar{Z}_w'$, and $\bar{Z}_w = \left[(\bar{\mathbf{z}}_{w1}, ...\bar{\mathbf{z}}_{wT})', \boldsymbol{\iota}_T\right]$ is the $T \times (K+2)$ matrix of observations on the weighted cross-sectional averages of the *observed* variables in (13) including a vector of ones, i.e. the typical entry is $\bar{\mathbf{z}}_{wt} = \Sigma_{i=1}^{N} w_i \mathbf{z}_{it}$.[15] The

[13] Notice that if the $\boldsymbol{\lambda}_i$ were known $\widehat{\Phi}$ could be estimated using a cross-sectional regression for each t and the rate of convergence would be $\sqrt{N}$, under arbitrary serial correlation and heteroskedasticity.

[14] The issue of estimating the number of factors is discussed in Section 6.2.

[15] For consistency, it is only required that the chosen weights satisfy, for each i, the condition $\Sigma_{j=1}^{N} w_{ij}^2 \to 0$ as $N \to \infty$. Therefore, an obvious choice is $w_{ij} = N^{-1}$ for all i.

intuition of this method lies in that even if $\boldsymbol{\phi}_t$ is unobserved, it is in the space spanned by the cross-sectional weighted averages of the observed variables. As a result, the projection as in (16) eliminates the factors and hence the inconsistency due to possible correlations that exist between the factors and the regressors.[16] To see this, assume the following general model for the correlation between $\boldsymbol{\phi}_t$ and $\mathbf{x}_{it}$:

$$\mathbf{x}_{it} = \Lambda_i^{*\prime}\boldsymbol{\phi}_t + \boldsymbol{\eta}_i^* + \boldsymbol{\upsilon}_{it}^*, \tag{17}$$

where Λ_i^* is the $M_0 \times K$ matrix of factor loadings of the covariates, $\boldsymbol{\upsilon}_{it}^*$ is the $K \times 1$ vector of the specific disturbances of $\mathbf{x}_{it}$ such that $E\left(\boldsymbol{\upsilon}_{it}^* | \Lambda_i^{*\prime}\boldsymbol{\phi}_t, \boldsymbol{\eta}_i^*\right) = 0$. Combining (13) with (17) yields

$$\begin{aligned} \underset{(K+1)\times 1}{\mathbf{z}_{it}} &= \begin{pmatrix} y_{it} \\ \mathbf{x}_{it} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\lambda}_i' + \boldsymbol{\beta}_i'\Lambda_i^* \\ \Lambda_i^{*\prime} \end{pmatrix} \boldsymbol{\phi}_t + \begin{pmatrix} \eta_i + \boldsymbol{\beta}_i'\boldsymbol{\eta}_i^* \\ \boldsymbol{\eta}_i^* \end{pmatrix} + \begin{pmatrix} \upsilon_{it} + \boldsymbol{\beta}_i'\boldsymbol{\upsilon}_{it}^* \\ \boldsymbol{\upsilon}_{it}^* \end{pmatrix} \\ &= \underset{(K+1)\times M}{\Lambda_i'} \underset{M\times 1}{\boldsymbol{\phi}_t} + \underset{(K+1)\times 1}{\boldsymbol{\eta}_i} + \underset{(K+1)\times 1}{\boldsymbol{\upsilon}_{it}}. \end{aligned} \tag{18}$$

Setting $w_{ij} = N^{-1}$ for all i and averaging over i gives

$$\mathbf{z}_t = \bar{\Lambda}'\boldsymbol{\phi}_t + \overline{\boldsymbol{\eta}} + \bar{\boldsymbol{\upsilon}}_t.$$

Assuming that

$$Rank\left(\bar{\Lambda}\right) = M_0 \leq K + 1 \text{ for all } N \tag{19}$$

and using the result that $\bar{\boldsymbol{\upsilon}}_t \xrightarrow{p} \mathbf{0}$ as $N \to \infty$ for each t, we have

$$\boldsymbol{\phi}_t - \left[\bar{\Lambda}\bar{\Lambda}'\right]^{-1}\bar{\Lambda}\left(\bar{\mathbf{z}}_t - \overline{\boldsymbol{\eta}}\right) \xrightarrow{p} 0 \text{ as } N \to \infty. \tag{20}$$

Thus employing the Frisch-Waugh theorem, (20) suggests using $\overline{y}_t$, $\overline{\mathbf{x}}_t$ and $\boldsymbol{\iota}_T$ as observable proxies for $\boldsymbol{\phi}_t$.[17]

Efficiency gains from pooling the observations over the cross-sectional dimension can be achieved when the individual slope coefficients are the same, i.e. $\boldsymbol{\beta}_i = \boldsymbol{\beta}$. Setting $w_{ij} = N^{-1}$ for all i yields the following pooled CCE estimator:

$$\widehat{\boldsymbol{\beta}}_{PCCE} = \left[\sum_{i=1}^{N} X_i'\bar{M}X_i\right]^{-1} \sum_{i=1}^{N} X_i'\bar{M}\mathbf{y}_i, \tag{21}$$

where $\bar{M}{=}I_T - \bar{Z}\left(\bar{Z}'\bar{Z}\right)^{-1}\bar{Z}'$ with $\bar{Z} = \left[\left(\bar{\mathbf{z}}_1, ...\bar{\mathbf{z}}_T\right)', \boldsymbol{\iota}_T\right]$ and $\bar{\mathbf{z}}_t = N^{-1}\Sigma_{j=1}^{N}\mathbf{z}_{jt}$. $\widehat{\boldsymbol{\beta}}_{PCCE}$ is asymptotically (large N) unbiased for $\boldsymbol{\beta}$, and as $(N,T) \overset{jointly}{\to} \infty$,

$$\sqrt{NT}\left(\widehat{\boldsymbol{\beta}}_{PCCE} - \boldsymbol{\beta}\right) \xrightarrow{d} N\left(\mathbf{0}, \Sigma_{PCCE}\right), \tag{22}$$

[16] A similar projection is proposed by Mundlak (1978) with the difference being that Pesaran's approach includes the cross-sectional mean of the dependent variable as well. Mundlak's projection will not work if the regressors are correlated with the factors.

[17] The scaling does not affect $\boldsymbol{\phi}_t$.

where Σ_{PCCE} is the asymptotic variance of $\sqrt{NT}\left(\widehat{\boldsymbol{\beta}}_{PCCE}-\boldsymbol{\beta}\right)$. For fixed T the distribution of $\widehat{\boldsymbol{\beta}}_{PCCE}$ is non-standard because it depends on nuisance parameters. The method of bootstrapping could be used to obtain standard errors for $\widehat{\boldsymbol{\beta}}_{PCCE}$ in this case although this is still a matter of research. Kapetanios, Pesaran, and Yamagata (2009) have extended the results of Pesaran (2006) by allowing unobserved common factors to follow unit root processes.

The CCE estimator is attractive because it is computationally very simple. Furthermore, the estimator has the additional advantage that it does not require specifying the number of factors, which is necessary if the latent factors are estimated using maximum likelihood or an approach based on principal components analysis. On the other hand, it is clear from (20) that the rank condition (19) might be crucial for the estimator. This will be violated if the number of unobserved factors is larger than $K+1$ or if, for example, the average of the factor loadings in the y_{it} and $\mathbf{x}_{it}$ equations tends to a zero vector, in which case $Rank\left(\bar{\Lambda}\right)<M_0$.[18] When the rank condition is violated, the CCE estimator requires, for consistency, that the factor loadings satisfy a random coefficients type assumption − specifically that $\boldsymbol{\lambda}_i$ and Λ_i^* are mutually independent and also independent from $\boldsymbol{\phi}_t$. Notice that under such assumption, the fixed effects estimator remains unbiased and consistent even if $\boldsymbol{\phi}_t$ is correlated with $\mathbf{x}_{it}$, provided that the observations are expressed in terms of deviations from time-specific averages.[19] This is because

$$\begin{aligned} & E\left(\underline{\upsilon}_{it}|\underline{\mathbf{x}}_{i1},...,\underline{\mathbf{x}}_{iT}\right)=E\left(\underline{\boldsymbol{\lambda}}_i'\boldsymbol{\phi}_t+\underline{\varepsilon}_{it}|\underline{\mathbf{x}}_{i1},...,\underline{\mathbf{x}}_{iT}\right)=E\left(\underline{\boldsymbol{\lambda}}_i'\boldsymbol{\phi}_t|\underline{\mathbf{x}}_{i1},...,\underline{\mathbf{x}}_{iT}\right) \\ = \quad & E\left(\underline{\boldsymbol{\lambda}}_i'|\underline{\mathbf{x}}_{i1},...,\underline{\mathbf{x}}_{iT}\right)E\left(\boldsymbol{\phi}_t|\underline{\mathbf{x}}_{i1},...,\underline{\mathbf{x}}_{iT}\right)=E\left(\underline{\boldsymbol{\lambda}}_i'\right)E\left(\boldsymbol{\phi}_t|\underline{\mathbf{x}}_{i1},...,\underline{\mathbf{x}}_{iT}\right)=\mathbf{0}, \end{aligned} \tag{23}$$

where the second equality holds under strong exogeneity of the covariates with respect to the idiosyncratic error, the third equality holds because $\boldsymbol{\lambda}_i$ and $\boldsymbol{\phi}_t$ are mutually independent and the fourth equality because $\boldsymbol{\lambda}_i$ and Λ_i^* are mutually independent. This result implies that even if $\boldsymbol{\phi}_t$ is correlated with $\mathbf{x}_{it}$ it is still possible to obtain consistent estimates of the parameters using the SUR approach of Robertson and Symons (2007) and the residual principal components estimator of Coakley, Fuertes and Smith (2002), provided that the first-stage estimated error covariance matrix, $\widehat{\Sigma}$, is based on the two-way fixed effects regression.

4.4 A Monte Carlo Study

4.4.1 Design

We investigate the finite sample performance of the above estimators using a limited Monte Carlo study. The underlying data generating process is given by

$$\begin{aligned} y_{it} &= \beta x_{it}+\omega_{it},\ \omega_{it}=\eta_i+\upsilon_{it},\ \ \upsilon_{it}=\lambda_i^1\phi_t^1+\lambda_i^2\phi_t^2+\varepsilon_{it}, \\ x_{it} &= \lambda_i^{*1}\phi_t^1+\lambda_i^{*2}\phi_t^2+\eta_i+\varepsilon_{it}^*,\ \ i=1,...,N,\ t=1,...,T, \end{aligned} \tag{24}$$

[18] The latter implies weak cross-sectional dependence under Definition 1. See e.g. example (9).

[19] Essentially this is the two-way error component fixed effects estimator. See e.g. Hsiao (2003, section 3.6) and Baltagi (2008, chapter 3).

where $\eta_i \sim i.i.d.N\left(0,\sigma_\eta^2\right)$, $\varepsilon_{it} \sim i.i.d.N\left(0,\sigma_\varepsilon^2\right)$ and $\varepsilon_{it}^* \sim i.i.d.N\left(0,1\right)$.[20] Furthemore, $\phi_t^m \sim i.i.d.N\left(0,1\right)$, $\lambda_i^m \sim i.i.d.N\left(\mu_{\lambda_m},\sigma_{\lambda_m}^2\right)$ for $m=1,2$, and

$$\lambda_i^{*m} = \sigma_{\lambda_m^*}[\rho_{\lambda_m}\lambda_i^m/\sigma_{\lambda_m} + (1-\rho_{\lambda_m}^2)^{1/2}\lambda_i^{**m}] = \rho_{\lambda_m}^*\lambda_i^m + \rho_{\lambda_m}^{**}\lambda_i^{**m}, \tag{25}$$

where $\rho_{\lambda_m} = E[(\lambda_i^m - \mu_{\lambda_m})(\lambda_i^{*m} - \mu_{\lambda_m^*})]/(\sigma_{\lambda_m}\sigma_{\lambda_m^*})$, $\rho_{\lambda_m}^* = \sigma_{\lambda_m^*}\rho_\lambda/\sigma_{\lambda_m}$, $\rho_{\lambda_m}^{**} = \sigma_{\lambda_m^*}(1-\rho_{\lambda_m}^2)^{1/2}$ and $\lambda_i^{**m} \sim i.i.d.N(\mu_{\lambda_m^{**}},1)$. Hence, λ_i^{*m} is a weighted sum of the two mutually independent random components, λ_i^m and λ_i^{**m}, which are weighted such that $Var(\lambda_i^{*m}) = \sigma_{\lambda_m^*}^2$ and $E[(\lambda_i^m - \mu_{\lambda_m})(\lambda_i^{*m} - \mu_{\lambda_m^*})] = \rho_{\lambda_m}\sigma_{\lambda_m}\sigma_{\lambda_m^*}$, as required. We set $\rho_{\lambda_m} = \rho_\lambda$ and $\sigma_{\lambda_m^*}^2 = \sigma_{\lambda_m}^2$ for $m=1,2$. We also set $\beta = 2$, $N = 100$, $T = 50$. 2,000 replications are performed.

Following Kiviet and Sarafidis (2009) we choose values for the simulation parameters on the basis of (i) Φ_1, the fraction of the 'structured' noise, υ_{it}, over the total noise, ω_{it}, (hence just excluding the idiosyncratic disturbance noise), i.e.

$$\Phi_1 \equiv \frac{\sigma_\eta^2 + \sum_{m=1}^2 \sigma_{\lambda_m}^2 + \sum_{m=1}^2 \mu_{\lambda_m}^2}{\sigma_\varepsilon^2 + \sigma_\eta^2 + \sum_{m=1}^2 \sigma_{\lambda_m}^2 + \sum_{m=1}^2 \mu_{\lambda_m}^2}; \tag{26}$$

(ii) Φ_2, which is the fraction of the factor noise over all structured noise, i.e.

$$\Phi_2 \equiv \frac{\sum_{m=1}^2 \sigma_{\lambda_m}^2 + \sum_{m=1}^2 \mu_{\lambda_m}^2}{\sigma_\eta^2 + \sum_{m=1}^2 \sigma_{\lambda_m}^2 + \sum_{m=1}^2 \mu_{\lambda_m}^2}; \tag{27}$$

and (iii) Φ_3, which reflects the closeness of the factor structure to an ordinary time effect (which it is when $\sigma_{\lambda_m}^2 = 0$ for $m=1,2$), i.e.

$$\Phi_3 \equiv \frac{\sum_{m=1}^2 \mu_{\lambda_m}^2}{\sum_{m=1}^2 \sigma_{\lambda_m}^2 + \sum_{m=1}^2 \mu_{\lambda_m}^2}. \tag{28}$$

Normalising $\sigma_\varepsilon^2 = 1$ implies that these three fractions parameterise completely σ_η^2, $\sum_{m=1}^2 \sigma_{\lambda_m}^2$ and $\sum_{m=1}^2 \mu_{\lambda_m}^2$. In particular, it is straightforward to show that $\sigma_\eta^2 = \frac{1-(\Phi_1+\Phi_2)+\Phi_1\Phi_2}{(\Phi_2\Phi_3)^2}$, $\sum_{m=1}^2 \sigma_{\lambda_m}^2 = \frac{1-(\Phi_1+\Phi_3)+\Phi_1\Phi_3}{\Phi_2(\Phi_3)^2}$ and $\sum_{m=1}^2 \mu_{\lambda_m}^2 = \frac{\Phi_1\Phi_2\Phi_3}{1-\Phi_1}$.[21] We set $\Phi_1 = 0.85$, $\Phi_2 = 0.80$ and $\Phi_3 = 0.80$. Further, we set $\sigma_{\lambda_1}^2\left(=\sigma_{\lambda_1^*}^2\right) = \sigma_{\lambda 2}^2\left(=\sigma_{\lambda_2^*}^2\right) = 2^{-1}\sum_{m=1}^2 \sigma_{\lambda_m}^2$, $\mu_{\lambda_1} = \left(\sum_{m=1}^2 \mu_{\lambda_m}^2 - 1\right)^{1/2}$ and $\mu_{\lambda_2} = 1$ We consider Case I in which

[20] We also performed the experiments with $\upsilon_{it}^* = u_{it} + \theta u_{it-1}$, $\theta = 0.5$, $u_{it} \sim i.i.d.N\left(0,1\right)$. However, the results were very similar and therefore they are not reported in the paper. They are available from the authors upon request.

[21] An alternative design, which is common practice in the literature, would be to choose the mean and variance of the error components such that the average error cross-sectional correlation, ρ_υ, equals a specific value. However, as noted in Kiviet and Sarafidis (2009), this is problematic because a particular value of the average error cross-sectional correlation can be obtained at a multitude of combinations of parameter values. On the contrary, reporting the values of these ratios enhance the transparency of the design.

the rank condition (19) is satisfied and Case II in which the rank condition is violated. The former sets $\mu_{\lambda_1^{**}} = \mu_{\lambda_1}\left(\frac{1-\rho_\lambda}{\left(1-\rho_\lambda^2\right)^{1/2}}\right)\left(\frac{1}{\sigma_{\lambda_1^*}}\right)$ and $\mu_{\lambda_2^{**}} = -\mu_{\lambda_2}\left(\frac{1-\rho_\lambda}{\left(1-\rho_\lambda^2\right)^{1/2}}\right)\left(\frac{1}{\sigma_{\lambda_2^*}}\right)$. This implies that $E\left(\lambda_i^{*1}\right) = \mu_{\lambda_1}$ but $E\left(\lambda_i^{*2}\right) = -\mu_{\lambda_2} \neq \mu_{\lambda_2}$. The latter sets $\mu_{\lambda_m^{**}} = \mu_{\lambda_m}\left(\frac{1-\rho_\lambda}{\left(1-\rho_\lambda^2\right)^{1/2}}\right)\left(\frac{1}{\sigma_{\lambda_m^*}}\right)$, which implies that $E\left(\lambda_i^{*m}\right) = \mu_{\lambda_m}$ for $m = 1, 2$. These two cases yield the following expectation for the matrix of factor loadings:

$$\text{Case I} : E\begin{bmatrix} \lambda_i^1 & \lambda_i^{*1} \\ \lambda_i^2 & \lambda_i^{*2} \end{bmatrix} \approx \begin{bmatrix} 1.347 & 1.347 \\ 1 & -1 \end{bmatrix},$$

and

$$\text{Case II} : E\begin{bmatrix} \lambda_i^1 & \lambda_i^{*1} \\ \lambda_i^2 & \lambda_i^{*2} \end{bmatrix} \approx \begin{bmatrix} 1.347 & 1.347 \\ 1 & 1 \end{bmatrix}.$$

We also consider two sub-cases for the correlation between λ_i and λ_i^* − specifically, $\rho_\lambda \in \{0, 0.5\}$, which generates Case I(I)a and Case I(I)b respectively.

4.4.2 Results

Table 1 reports bias, expressed as a percentage, and root mean square error (RMSE) for all estimators.[22] FE and $TWFE$ denote the one-way and two-way error component fixed effects estimators respectively, FE-PC and $TWFE$-PC denote the principal components estimator proposed by Coakley, Fuertes and Smith (2002), based on FE and $TWFE$ residuals; and $PCCE$ and PC denote (21) and the iterative version of (14) respectively. Firstly, we can see that FE exhibits a large bias in all cases. This is because the within transformation does not eliminate the factor structure and the ϕ_ts are correlated with x_{it} given (24). $TWFE$ performs, perhaps surprisingly, very well in terms of both bias and RMSE so long as the factor loadings of y and x are uncorrelated − namely, $\rho_\lambda = 0$. As expected, the estimator is not affected by whether the rank condition is satisfied or not. However, when $\rho_\lambda = 0.5$ both bias and RMSE of $TWFE$ increase substantially because (23) does no longer hold true. The performance of FE-PC and $TWFE$-PC is naturally affected by the properties of the residuals they use at first-stage. Hence, FE-PC is biased in all circumstances, while $TWFE$-PC appears to perform very well when $TWFE$ also does well. Of course in this case $TWFE$-PC outperforms $TWFE$ in terms of variance and RMSE since $TWFE$-PC augments the $TWFE$ model by including estimates of the M_0 principal components in the set of regressors. $PCCE$ performs best when the rank condition is satisfied. In fact, in this case $PCCE$ outperforms PC, which is remarkable given that M_0 is assumed to be known. When the rank condition is violated the estimator seems to do well for $\rho_\lambda = 0$, although it is outperformed by $TWFE$-PC and PC in this case. When $\rho_\lambda = 0.5$ the performance

[22] Specifically, we report bias in terms of $100\left(\overline{\beta}_c - \beta\right)/\beta$, where $\overline{\beta}$ is the average estimate over all replications of β, obtained using method c. Since $\beta = 1$, the entries represent essentially bias multiplied by one hundred.

of $PCCE$ deteriorates substantially. On the other hand, while PC is not affected by the rank condition, it is affected by the value of ρ_λ. Therefore, for $\rho_\lambda = 0.5$ PC outperforms $PCCE$ only when the rank condition is violated. In this case $TWFE$-PC does best.

Table 1. Bias in % and RMSE of Estimators

FE	$TWFE$	FE-PC	$TWFE$-PC	$PCCE$	PC
Case Ia: Rank condition satisfied, $\rho_\lambda = 0$.					
13.7	.000	3.67	.027	.027	.063
(.290)	(.034)	(.089)	(.029)	(.013)	(.036)
Case Ib: Rank condition satisfied, $\rho_\lambda = 0.5$.					
32.0	10.5	4.01	4.72	.271	4.04
(.643)	(.213)	(.087)	(.107)	(.014)	(.094)
Case IIa: Rank condition violated, $\rho_\lambda = 0$.					
31.7	.000	2.43	.051	.051	.067
(.625)	(.034)	(.051)	(.023)	(.027)	(.023)
Case IIb: Rank condition violated, $\rho_\lambda = 05$.					
35.1	10.5	4.48	3.57	5.95	4.06
(.704)	(.213)	(.092)	(.083)	(.125)	(.093)

In conclusion, we can see that possible non-zero correlations between the factor loadings of y and x can have an adverse effect on the estimators. Whether this issue applies or not in practice depends on the specific application of course. As an example, suppose that (1) represents a cost function, where y_{it} denotes cost, $\mathbf{x}_{it}$ denotes a vector of input prices and output, and ϕ_t denotes an oil price shock that hits the industry as a whole at time t; in this case, it is natural to think that the factor loadings of input prices will be (highly) correlated with the factor loadings of cost. Similarly, one may think of examples where the factor loadings would be mutually uncorrelated.

As mentioned above, PCCE and PC are valid under N and T both large. When T is small it may be more natural to employ a fixed effects treatment of $\boldsymbol{\lambda}_i$ and use one of the methods described in Sections 5.3 and 5.4.

5 Estimation Under Weak Exogeneity

Economic behaviour is intrinsically dynamic. For example, as a result of the force of habit, individual agents may change their consumption and investment patterns with a lapse of time. Similarly, technological and institutional reasons may prevent firms from switching between optimal levels of capital and labor instantaneously. Imperfect knowledge and uncertainty may also contribute to persistence, or a delayed response to shocks by decision makers. In most cases future expectations can play an important role in decision making and expectational errors may imply some degree of dependence between a subset of the regressors and lagged disturbances, leading to weak exogeneity as in (4).

5.1 Asymptotic Properties of Least Squares Estimators

Often the weakly exogenous regressor takes the form of a lagged dependent variable. Since this variable is by construction correlated with the individual effect, η_i, estimation of the dynamic panel data model is not straightforward and indeed it has spawned a vast literature, which is still growing. In its simplest form, where the lagged dependent variable is the only regressor, the model is

$$y_{it} = \alpha y_{it-1} + \eta_i + \upsilon_{it}, \quad |\alpha| < 1,\, i = 1, 2, \ldots, N,\, t = 1, 2, \ldots, T. \tag{29}$$

Since strong exogeneity is violated in (29), standard least-squares-based estimators that rely on the elimination of the individual effect yield inconsistent parameter estimates even if there is no error cross-sectional dependence. Two such estimators are the fixed effects and first-differenced estimators, which converge to the following limiting values[23]:

$$\begin{aligned} \text{plim}_{N\to\infty}\hat{\alpha}_{FE} &= \alpha - \xi_N(\alpha, T)\,\xi_D(\alpha, T)^{-1}, \text{ and} & (30)\\ \text{plim}_{N\to\infty}\hat{\alpha}_{FD} &= \frac{\alpha - 1}{2}, & (31) \end{aligned}$$

where $\xi_A(\alpha, T) = [T(1-\alpha)]^{-1}\left[T - \left(1-\alpha^T\right)(1-\alpha)^{-1}\right]$ and $\xi_B(\alpha, T)^{-1} = \left[(T-1)\left(1-\alpha^2\right)^{-1}\right]$ $\left[1 - 2\alpha\left[(1-\alpha)(T-1)\right]^{-1}\left(1 - \left(1-\alpha^T\right)[T(1-\alpha)]^{-1}\right)\right]$. It follows that both $\hat{\alpha}_{FE}$ and $\hat{\alpha}_{FD}$ are inconsistent for fixed T as $N \to \infty$. For $T \to \infty$, $\hat{\alpha}_{FE}$ is consistent but $\hat{\alpha}_{FD}$ is not, unless $\text{Var}(\eta_i) = 0$. One way to obtain consistent parameter estimates is to start from (30) or (31); since both estimators converge into functions of α (and T) alone, it is possible to solve in terms of α and obtain $\sqrt{N}$-consistent estimates of the autoregressive parameter. In the case of (30) the solution requires a numerical approach due to the fact that ξ_A and ξ_B are highly nonlinear. However, (31) involves a linear function, making the construction of a consistent, or "bias-corrected", estimator trivial[24]

$$\hat{\alpha}_{BCFD} = 2\hat{\alpha}_{FD} + 1. \tag{32}$$

Phillips and Sul (2007) analyse the properties of $\hat{\alpha}_{FE}$ under error cross-sectional dependence. They show that the estimator converges, for fixed T, to

$$\text{plim}_{N\to\infty}\hat{\alpha}_{FE} = \alpha - \left[\sigma_\varepsilon^2\xi_A(\alpha, T) + \psi_{AT}^{(1)}\right]\left[\sigma_\varepsilon^2\xi_B(\alpha, T) + \psi_{BT}^{(1)}\right]^{-1}, \tag{33}$$

where $\xi_A(\alpha, T)$ and $\xi_B(\alpha, T)$ are defined in (30), and

$$\begin{aligned} \psi_{AT}^{(1)} &= \sum_{t=1}^{T}\left(\boldsymbol{\phi}_t - \overline{\boldsymbol{\phi}}\right)'\left[\Sigma_\lambda + \boldsymbol{\mu}_\lambda\boldsymbol{\mu}_\lambda'\right](\mathbf{w}_{t-1} - \overline{\mathbf{w}}_{,-1}), \text{ and} \\ \psi_{BT}^{(1)} &= \sum_{t=1}^{T}(\mathbf{w}_{t-1} - \overline{\mathbf{w}}_{,-1})'\left[\Sigma_\lambda + \boldsymbol{\mu}_\lambda\boldsymbol{\mu}_\lambda'\right](\mathbf{w}_{t-1} - \overline{\mathbf{w}}_{,-1}), \end{aligned} \tag{34}$$

[23] See Nickell (1981) and Phillips and Sul (2007).

[24] See Chowdhury (1987). It is worth mentioning that bias-corrected estimators of this type have not found many applications in the literature since it is not straightforward to generalise them into models that include weakly exogenous regressors other than the lagged dependent variable.

where $\mathbf{w}_{t-1} = \sum_{\tau=0}^{\infty} \alpha^{\tau} \boldsymbol{\phi}_{t-1-\tau}$, $\overline{\mathbf{w}}_{,-1} = T^{-1} \sum_{t=1}^{T} \mathbf{w}_{t-1}$, $E(\boldsymbol{\lambda}_i) = \boldsymbol{\mu}_{\lambda}$ and $E(\boldsymbol{\lambda}_i - \boldsymbol{\mu}_{\lambda})(\boldsymbol{\lambda}_i - \boldsymbol{\mu}_{\lambda})' = \Sigma_{\lambda}$. Therefore, cross-sectional dependence adds an extra source of bias; for $\boldsymbol{\lambda}_i = 0$ (33) reduces to (30). It is worth noting that contrary to (30), the probability limit in (33) depends on nuisance parameters, in particular $\boldsymbol{\phi}_t$. According to Phillips and Sul (2007), this may explain the substantial variability observed in dynamic panel estimates when there is cross-sectional dependence, even in situations where N is large.

It is worth mentioning that time-specific demeaning of the observations will not remove the source of bias that is due to the factor structure from (33), even if the factor loadings satisfy a random coefficients type assumption. Instead, it is straightfoward to show that the only difference in the plim of $\widehat{\alpha}_{FE}$ is that (34) changes to

$$\begin{aligned} \widetilde{\psi}_{AT}^{(1)} &= \sum_{t=1}^{T} \left(\boldsymbol{\phi}_t - \overline{\boldsymbol{\phi}}\right)' \Sigma_{\lambda} \left(\mathbf{w}_{t-1} - \overline{\mathbf{w}}_{,-1}\right), \text{ and} \\ \widetilde{\psi}_{BT}^{(1)} &= \sum_{t=1}^{T} \left(\mathbf{w}_{t-1} - \overline{\mathbf{w}}_{,-1}\right)' \Sigma_{\lambda} \left(\mathbf{w}_{t-1} - \overline{\mathbf{w}}_{,-1}\right). \end{aligned} \tag{35}$$

This is contrary to the case of strong exogeneity, in which the two-way error component fixed effects estimator is consistent under a random coefficients type assumption for the factor loadings, even if the latent factors are correlated with the regressors.

Unfortunately, under weak exogeneity the methods discussed in the previous section, based on maximum likelihood and principal components analysis, will not generally yield consistent parameter estimates either. For instance, the PC estimator may be thought of as a two-stage process, whereby Φ is purged from the model by multiplying through by the projection M_{ϕ} and then $\widehat{\alpha}$ is obtained by (non-)linear regression. However this procedure transforms the residuals of the model to $M_{\phi}\widetilde{\boldsymbol{\upsilon}}_i$. Each entry of this vector is a linear combination of the elements of the whole time-series $\widetilde{\boldsymbol{\upsilon}}_i$ and thus it is not orthogonal to the corresponding entry in $\widetilde{X}_i$. This means that writing (14) as

$$\widehat{\boldsymbol{\beta}}_{PC} = \boldsymbol{\beta} + \left[N^{-1} \sum_{i}^{N} \widetilde{X}_i' M_{\widehat{\phi}} \widetilde{X}_i \right]^{-1} N^{-1} \sum_{i}^{N} \widetilde{X}_i' M_{\widehat{\phi}} \widetilde{\boldsymbol{\upsilon}}_i, \tag{36}$$

the second term on the right-hand side will not have zero probability limit as $N \to \infty$ for fixed T, or even $(N,T) \overset{jointly}{\to} \infty$. The same issue arises with the CCE estimator, which is equivalent to a (stacked) linear regression of $\mathbf{y}_i$ on $\bar{M}X_i$ where in this case $\bar{M}$ is the projection that removes $\overline{\mathbf{y}}_t$ and $\overline{\mathbf{x}}_t$, and as such the transformed residuals are not orthogonal to the regressor unless the latter is strongly exogenous.

The properties of the bias-corrected FD estimator (32) under error cross-sectional dependence are investigated by Hayakawa (2007). The author shows that

$$\text{plim}_{N\to\infty} \hat{\alpha}_{BCFD} = 2\alpha + 1 + \frac{2T^{-1}\psi_{AT}^{(2)} - 2\sigma_{\varepsilon}^2}{T^{-1}\psi_{BT}^{(2)} + \frac{2}{1+\alpha}\sigma_{\varepsilon}^2}, \tag{37}$$

where

$$\psi_{AT}^{(2)}=\sum_{t=1}^{T}\Delta\boldsymbol{\phi}_t{}'\left[\Sigma_\lambda+\boldsymbol{\mu}_\lambda\boldsymbol{\mu}_\lambda'\right]\Delta\mathbf{w}_{t-1}\text{ and }\psi_{BT}^{(2)}=\sum_{t=1}^{T}\Delta\mathbf{w}_{t-1}'\left[\Sigma_\lambda+\boldsymbol{\mu}_\lambda\boldsymbol{\mu}_\lambda'\right]\Delta\mathbf{w}_{t-1},\quad(38)$$

where $\Delta\mathbf{w}_{t-1}=\sum_{\tau=0}^{\infty}\alpha^\tau\Delta\boldsymbol{\phi}_{t-1-\tau}$. Therefore, in this case cross-sectional dependence turns an otherwise consistent estimator inconsistent, for fixed T. On the other hand, for $T\to\infty$ we have that $\text{plim}_{N\to\infty}\hat{\alpha}_{BCFD}=\alpha$, regardless of whether N is fixed or $N\to\infty$.

5.2 Asymptotic Properties of IV and GMM Estimators

By far the most popular approach for estimating dynamic panel data models is the method of instrumental variables and the Generalised Method of Moments (see Anderson and Hsiao, 1981, Holtz-Eakin, Newey and Rosen, 1988, and Arellano and Bond, 1991). The point of departure is the simple observation that

$$E\left[y_{it-s}\Delta\upsilon_{it}\right]=0;\text{ for }t=2,...,T\text{ and }2\leq s\leq t.\quad(39)$$

Replacing the expectation by the sample average and minimising a weighted quadratic distance function with respect to α produces a consistent estimator. Subsequent extensions augment the standard first-differenced GMM estimator with additional moment conditions implied either by the same basic assumptions (Ahn and Schmidt, 1995) or by additional assumptions regarding the initial conditions (Arellano and Bover, 1995, and Blundell and Bond, 1998). The latter allows one to combine the equations in first-differences with the equations in levels, constructing a 'system' GMM estimator.

Sarafidis and Robertson (2009) analyse the behavior of the standard IV estimator under error cross-sectional dependence. They show that the estimator has the following probability limit:

$$\text{plim}_{N\to\infty}\left(\hat{\alpha}_{IV}-\alpha\right)=\frac{\text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\sum_{t=2}^{T}y_{it-2}\Delta\upsilon_{it}}{\text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\sum_{t=2}^{T}y_{it-2}\Delta y_{it-1}}=\psi_{AT}^{(3)}\left[\psi_{BT}^{(3)}-\frac{(T-1)}{1+\alpha}\sigma_\varepsilon^2\right]^{-1},\quad(40)$$

where

$$\psi_{AT}^{(3)}=\sum_{t=2}^{T}\Delta\mathbf{f}_t'\left[\Sigma_\lambda+\boldsymbol{\mu}_\lambda\boldsymbol{\mu}_\lambda'\right]\mathbf{w}_{t-2}\text{ and }\psi_{BT}^{(3)}=\sum_{t=2}^{T}\Delta\mathbf{w}_{t-1}'\left[\Sigma_\lambda+\boldsymbol{\mu}_\lambda\boldsymbol{\mu}_\lambda'\right]\mathbf{w}_{t-2}.\quad(41)$$

Therefore, similarly to the result for BCFD, cross-sectional dependence renders an otherwise consistent estimator inconsistent for fixed T.[25] Essentially, this is because the numerator in (40) converges to the population moment condition, conditional on $\{\boldsymbol{\phi}_\tau\}_{-\infty}^{t}$, which is $E\left[\left(y_{it-2}\Delta\upsilon_{it}\right)|\left\{\boldsymbol{\phi}_\tau\right\}_{-\infty}^{t}\right]\neq0$, even if the unconditional expectation $E\left[\left(y_{it-2}\Delta\upsilon_{it}\right)\right]=0$. A direct by-product of this result is that all standard GMM estimators that make use of lagged values of the dependent variable as instruments for

[25] Notice, however, that for large T the bias diminishes because $\frac{1}{T}\psi_{AT}^{(3)}=o_p\left(1\right)$.

the endogenous regressor are inconsistent. This holds true for any lag length of the instruments used. For instance, (39) becomes

$$E\left(y_{it-s}\Delta\upsilon_{it}|\left\{\boldsymbol{\phi}_{\tau}\right\}_{-\infty}^{t}\right)=\Delta\boldsymbol{\phi}_{t}^{\prime}\left[\Sigma_{\lambda}+\boldsymbol{\mu}_{\lambda}\boldsymbol{\mu}_{\lambda}^{\prime}\right]\mathbf{w}_{t-s}\neq0;\text{ for }t=2,...,T\text{ and }2\leq s\leq t-1. \tag{42}$$

A similar result applies for system GMM. In general, the asymptotic bias of these estimators will depend on the particular transformation employed, the number of instruments used and the choice of the weighting matrix. It is worth emphasizing that not all forms of cross-sectional dependence are detrimental to GMM. Sarafidis (2009) focuses on the conditions required on the cross-sectional dimension of the error process for the standard dynamic panel GMM estimator to remain consistent. He demonstrates that, if there is cross-sectional dependence in the errors, it suffices that this is weak (under Definition 2).

Notice that for the single-factor case, the asymptotic bias of $\hat{\alpha}_{IV}$ reduces to

$$\text{plim}_{N\rightarrow\infty}\left(\hat{\alpha}_{IV}-\alpha\right)=\frac{\left(\sigma_{\lambda}^{2}+\mu_{\lambda}^{2}\right)\kappa_{1}}{\left(\sigma_{\lambda}^{2}+\mu_{\lambda}^{2}\right)\kappa_{2}-\frac{(T-1)}{1+\alpha}\sigma_{\varepsilon}^{2}}, \tag{43}$$

where $\kappa_{1}=\sum_{t=2}^{T}w_{t-2}\Delta\phi_{t}$ and $\kappa_{2}=\sum_{t=2}^{T}\Delta w_{t-1}w_{t-2}=\sum_{t=2}^{T}w_{t-1}w_{t-2}-\sum_{t=2}^{T}\left(w_{t-2}\right)^{2}$. Sarafidis and Robertson (2009) demonstrate that $\hat{\alpha}_{IV}$ is biased downwards in this case.

When the observations are expressed in terms of deviations from time-specific averages, the asymptotic bias of the IV estimator is

$$\text{plim}_{N\rightarrow\infty}\left(\widetilde{\alpha}_{IV}-\alpha\right)=\frac{\text{plim}_{N\rightarrow\infty}\frac{1}{N}\sum_{i=1}^{N}\sum_{t=2}^{T}\underline{y}_{it-2}\Delta\underline{\upsilon}_{it}}{\text{plim}_{N\rightarrow\infty}\frac{1}{N}\sum_{i=1}^{N}\sum_{t=2}^{T}\underline{y}_{it-2}\Delta\underline{y}_{it-1}}=\overset{\sim}{\psi}_{AT}^{(3)}\left[\overset{\sim}{\psi}_{BT}^{(3)}-\frac{(T-2)}{1+\lambda}\sigma_{\varepsilon}^{2}\right]^{-1}, \tag{44}$$

where $\underline{y}_{it}=y_{it}-\overline{y}_{t}$, $\overset{\sim}{\psi}_{AT}^{(3)}=\sum_{t=2}^{T}\Delta\mathbf{f}_{t}^{\prime}\Sigma_{\lambda}\mathbf{w}_{t-2}$ and $\overset{\sim}{\psi}_{BT}^{(3)}=\sum_{t=2}^{T}\Delta\mathbf{w}_{t-1}^{\prime}\Sigma_{\lambda}\mathbf{w}_{t-2}$, while for $M_{0}=1$ (44) reduces to

$$\text{plim}_{N\rightarrow\infty}\left(\widetilde{\alpha}_{IV}-\alpha\right)=\frac{\sigma_{\lambda}^{2}\kappa_{1}}{\sigma_{\lambda}^{2}\kappa_{2}-\frac{(T-1)}{1+\alpha}\sigma_{\varepsilon}^{2}}. \tag{45}$$

Notice that for $\sigma_{\lambda}^{2}=0$ (i.e. factor loadings have zero variance) the bias in $\widetilde{\alpha}_{IV}$ disappears. Sarafidis and Robertson demonstrate that unless $\mu_{\lambda}=0$ the asymptotic bias of $\widetilde{\alpha}_{IV}$ is, in general, going to be smaller than $\hat{\alpha}_{IV}$. Intuitively, this is because time-specific demeaning reduces the impact of the factor structure (by removing the mean value of $\boldsymbol{\lambda}_{i}$), which is the reason for the asymptotic bias of the IV estimator. Simulation results confirm these findings and provide a formal justification to the practice of including common time effects in the context of a short dynamic panel data model with large N and T fixed.

Although time-specific demeaning may reduce the impact of cross-sectional dependence (provided that $\boldsymbol{\mu}_{\lambda}\neq\mathbf{0}$) it will not eliminate it, unless $\Sigma_{\lambda}=0$. Sarafidis, Yamagata and Robertson (2009) show that one can still obtain fixed-T, $\sqrt{N}$-consistent estimates

of the parameters within the GMM framework by using instruments with respect to the subset of regressors that are strongly exogenous (if any), provided that they remain so in the mispecified model. Strong exogeneity of a subset of $\mathbf{x}_{it}$ will be maintained in the mispecified model if their factor loadings are either zero, or mutually uncorrelated with the factor loadings involved in the y process. Empirically, this can be determined using Sargan's (1958) or Hansen's (1982) overidentification restrictions test statistic.

Sarafidis (2009) demonstrates that under weakly correlated errors (using Definition 2), an additional, non-redundant, set of moment conditions arises for each individual i − specifically, instruments with respect to the individual(s) with which unit i is weakly correlated. This set of instruments can be used to obtain consistent estimates of the parameters in situations where the error structure is subject to both weak and strong correlations.[26] For instance, consider (29) and let

$$\upsilon_{it} = \boldsymbol{\lambda}_i' \boldsymbol{\phi}_t + \varepsilon_{it} + \theta \varepsilon_{jt}. \tag{46}$$

Hence the composite error, υ_{it}, is subject to a multi-factor structure and the purely idiosyncratic component, ε_{it}, is spatially correlated and follows an MA(1) process.[27] In other words, mispecification of the model results in both global (factors) and local (spatial) correlations. Transforming in terms of deviations from time-specific averages yields

$$\underline{y}_{it} = \alpha \underline{y}_{it-1} + \underline{\eta}_i + \underline{\boldsymbol{\lambda}}_i' \boldsymbol{\phi}_t + \underline{\varepsilon}_{i,t} + \theta \underline{\varepsilon}_{j,t}. \tag{47}$$

As mentioned above, the moment conditions with respect to lagged values of y are invalidated under the multi-factor structure (see e.g. (42)). However, it turns out that

$$E\left(\underline{y}_{j,t-s} \Delta \underline{\upsilon}_{i,t} | \left\{\boldsymbol{\phi}_\tau\right\}_{-\infty}^{t}\right) = 0; \text{for } t = 2, ..., T \text{ and } 2 \leq s \leq t-1. \tag{48}$$

The required assumption for the above result is that the factor loadings are cross-sectionally uncorrelated, i.e. $E\left(\underline{\boldsymbol{\lambda}}_i \underline{\boldsymbol{\lambda}}_j' | \Phi\right)$. A similar expression to (48) (mutatis mutandis) applies for system GMM. The resulting GMM approach is attractive because it does not require strongly exogenous regressors under the mispecified model, although it does require the specification of a weighting matrix, which may or may not be appropriate in certain economic applications.

5.3 Estimation Using Quasi-Differencing Approaches

A different approach, valid for fixed T, which does not require strongly exogenous regressors or spatially correlated errors, is to treat $\boldsymbol{\phi}_t$ as a vector of fixed parameters, specific to each time period, and the vector $\boldsymbol{\lambda}_i$ as random variables which are correlated with the covariates but uncorrelated with the purely idiosyncratic component, ε_{it}. Essentially, this is the usual fixed effects assumption extended to the factor case, although

[26] This structure is also studied by Pesaran and Tosetti (2009) and Chudick, Pesaran and Tosetti (2009).

[27] Spatial dependence is only a device here; in fact, ordering of the observations is not necessary to obtain the results.

the within transformation cannot eliminate the common factors in this case. To this end, a number of different transformations, all based on quasi-differencing, have been proposed to eliminate the factors from the model and estimate the structural parameters using the generalised method of moments.

An early application of this approach has been considered by Wansbeek and Knaap (1999) who imposed $M_0 = 1$ and $\phi_t = t$. So instead of an arbitrary sequence of time fixed effects $\phi_1, ..., \phi_T$, entering the model multiplicatively, there is a linear trend with individual-specific coefficients. After taking first-differences η_i drops out of the model. The linear trend becomes a constant, which disappears after taking first-differences again. Double-differencing may eliminate much of the variation of the data and the issue of weak instruments might arise, cf. Bekker (1994), also discussed by Wansbeek and Knaap (1999).

A generalization of the model above is given by Nauges and Thomas (2003), employing a transformation proposed by Holtz-Eakin, Newey and Rosen (1988). In particular, they consider (29) with a single-factor error structure, i.e. $\upsilon_{it} = \lambda_i \phi_t + \varepsilon_{it}$, and T fixed. They use first-differencing to eliminate the individual effects, which yields

$$\Delta y_{it} = \alpha \Delta y_{it-1} + \Delta \upsilon_{it}, \;\; \Delta \upsilon_{it} = \lambda_i \Delta \phi_t + \Delta \varepsilon_{it}. \tag{49}$$

Define $\varrho_t = \Delta\phi_t / \Delta\phi_{t-1}$; lagging (49), multiplying by ϱ_t and subtracting yields

$$\begin{aligned} \Delta y_{it} - \varrho_t \Delta y_{it-1} &= \alpha \left(\Delta y_{it-1} - \varrho_t \Delta y_{it-2} \right) + \lambda_i \left[\Delta \phi_t - \varrho_t \Delta \phi_{t-1} \right] + \left(\Delta \varepsilon_{it} - \varrho_t \Delta \varepsilon_{it-1} \right) \\ &= \alpha \left(\Delta y_{it-1} - \varrho_t \Delta y_{it-2} \right) + \left(\Delta \varepsilon_{it} - \varrho_t \Delta \varepsilon_{it-1} \right). \end{aligned} \tag{50}$$

Notice that appropriate lagged values of the dependent variable will be uncorrelated with the transformed error term, leading to a GMM estimator based on Arellano-Bond type of moment conditions. Assuming ε_{it} is serially uncorrelated, this set of moment conditions is

$$E\left[y_{it-s} \left(\Delta \varepsilon_{it} - \varrho_t \Delta \varepsilon_{it-1} \right) \right] = 0; \text{ for } t = 3, ..., T \text{ and } 3 \leq s \leq t. \tag{51}$$

The main difference with (39) is that the moment conditions above are non-linear because the time-specific nuisance parameters, ϱ_t, need to be estimated jointly with the structural parameter, α. The results from their Monte Carlo study are mixed; while the proposed GMM estimator exhibits, in general, smaller bias compared to the standard first-differenced GMM estimator, it also has larger variance to the extent that it is outperformed in terms of RMSE.[28]

Ahn, Lee and Schmidt (2006) consider a model with a multi-factor error structure and weakly/strongly exogenous regressors. They use a different transformation, based on multi-quasi-differencing, and propose a GMM estimator applied on the multi-quasi-differenced model. To see how this method works, assume, without loss of generality, that $M_0 = 2$ and consider the following model:

$$y_{it} = \alpha y_{it-1} + \upsilon_{it}, \;\; \upsilon_{it} = \lambda_i^1 \phi_t^1 + \lambda_i^2 \phi_t^2 + \varepsilon_{it}, \tag{52}$$

[28] An alternative transformation for the single-factor model, based on quasi-differencing as well, is provided by Ahn, Lee and Schmidt (2001).

where ε_{it} is serially uncorrelated. Identification of this factor model requires M_0^2 [$= 4$] restrictions. Typically, $M_0 (M_0 + 1)/2$ restrictions arise by normalising

$$\sum_{t=1}^{T} \phi_t^m \phi_t^n = \begin{cases} 1 \text{ for } m = n, \\ 0 \text{ otherwise.} \end{cases} \tag{53}$$

Additional $M_0 (M_0 - 1)/2$ restrictions are usually obtained by requiring that the factor loadings are mutually uncorrelated. Since $M_0 = 2$, this would yield one extra restriction in the present case. Alternatively, one can impose M_0^2 restrictions solely on the factors, which are treated as parameters. This is the approach followed by Ahn, Lee and Schmidt. In particular, they normalise $\phi_T^1 = 1$, $\phi_{T-1}^1 = 0$, $\phi_T^2 = 0$, $\phi_{T-1}^2 = 1$. In this case model (52) becomes, for periods $T - 1$ and T, respectively,

$$y_{iT-1} = \gamma y_{iT-2} + \lambda_i^2 + \varepsilon_{iT-1}, \tag{54}$$

and

$$y_{iT} = \alpha y_{iT-1} + \lambda_i^1 + \varepsilon_{iT}. \tag{55}$$

Multiplying (54) and (55) by ϕ_t^1 and ϕ_t^2 respectively and subtracting from (52) yields

$$\begin{aligned} & y_{it} - \phi_t^1 y_{iT} - \phi_t^2 y_{iT-1} \\ = \; & \alpha \left(y_{it-1} - \phi_t^1 y_{iT-1} - \phi_t^2 y_{iT-2} \right) + \left(\varepsilon_{it} - \phi_t^1 \varepsilon_{iT} - \phi_t^2 \varepsilon_{iT-1} \right). \end{aligned} \tag{56}$$

This suggests the following $(T - M_0)(T - M_0 + 1)/2$ non-linear moment conditions:

$$E\left[y_{it-s} \left(\varepsilon_{it} - \phi_t^1 \varepsilon_{iT} - \phi_t^2 \varepsilon_{iT-1} \right) \right] = 0; \text{ for } t = 2, ..., T \text{ and } 2 \leq s \leq t, \tag{57}$$

which lead to joint estimation of the structural parameter, α, and the $(T - 2) \times 2$ nuisance parameters. In a compact form, for any fixed number of factors one may write the model as

$$\mathbf{y}_i = \alpha \mathbf{y}_{i,-1} + \boldsymbol{\Xi} \lambda_i + \boldsymbol{\varepsilon}_i, \tag{58}$$

where $\Xi = (\Phi_u', -I_M)$ and Φ_u is the $(T - M_0 - 1) \times M_0$ matrix of unrestricted parameters with typical entry $\boldsymbol{\phi}_t^m$ for $t = 2, ..., T - M_0$ and $m = 1, ..., M_0$. The transformation that makes the error orthogonal to the factors amounts to pre-multiplying (58) by $\boldsymbol{\Xi}_1'$, where $\boldsymbol{\Xi}_1 = (I_{T-M}, \Phi_u)'$, since $\boldsymbol{\Xi}_1' \boldsymbol{\Xi} = 0$ by construction.

5.4 Alternative Approaches

An alternative approach to quasi-differencing involves introducing explicitly a new set of parameters, which, under the fixed effects assumption, represent the unobserved covariances between the covariates and the common factor component of the disturbance. This is the method followed by Robertson, Sarafidis and Symons (2010) and Bai (2010). The former is based on the generalised method of moments and the latter on the method of maximum likelihood. It is instructive to illustrate these methods using model (52) and letting M_0 free. Let $E(\boldsymbol{\lambda}_i y_{is}) = \boldsymbol{\gamma}_s$ for any t and s. Notice that under no serial

correlation in ε_{it}, we have $E(\varepsilon_{it} y_{is}) = 0$ for any $s \leq t-1$. Therefore, the following $T(T+1)/2$ centered moment conditions exist:

$$E\left[y_{is} v_{it} - \boldsymbol{\gamma}_s' \boldsymbol{\phi}_t\right] = 0, \text{ for } t = 1, ..., T \text{ and } s \leq t-1. \tag{59}$$

The above expression is similar to a moment condition like $E(X_i - \mu) = 0$, except that the former is non-linear because some of the parameters enter multiplicatively. Writing the model in vector form, we have

$$\mathbf{y}_i = \alpha \mathbf{y}_{i,-1} + \boldsymbol{v}_i, \;\; \boldsymbol{v}_i = \Phi \boldsymbol{\lambda}_i + \boldsymbol{\varepsilon}_i, \tag{60}$$

where $\mathbf{y}_i = (y_{i1}, ..., y_{iT})'$, $\mathbf{y}_{i,-1} = (y_{i0}, ..., y_{iT-1})'$, $\Phi = (\boldsymbol{\phi}_1, ...\boldsymbol{\phi}_T)'$ is a $T \times M_0$ matrix, and $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, ..., \varepsilon_{iT})'$. Define the matrix of instruments as follows:

$$\underset{T \times T(T+1)/2}{Z_i} = \begin{bmatrix} y_{i0} & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & y_{i0} & y_{i1} & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & & & & \\ & & & & & & & \\ 0 & 0 & 0 & ... & y_{i0} & y_{i1} & ... & y_{iT-1} \end{bmatrix}. \tag{61}$$

We have

$$E\left[Z_i' \mathbf{u}_i - S(I_T \otimes \Gamma) \boldsymbol{\phi}^T\right] = \mathbf{0}, \tag{62}$$

where S is a selector matrix of order $T(T+1)/2 \times T^2$ that consists of 0s and 1s, with a single 1 in each row[29], $\Gamma = \left(\boldsymbol{\gamma}_0, ...\boldsymbol{\gamma}_{T-1}\right)'$ is a $T \times M_0$ matrix, and $\boldsymbol{\phi}^T = \left(\boldsymbol{\phi}_1', \boldsymbol{\phi}_2', ..., \boldsymbol{\phi}_T'\right)'$ is a $TM_0 \times 1$ vector.

Replacing the population moments with their sample averages yields

$$N^{-1} \sum_{i=1}^{N} \left[Z_i' \mathbf{u}_i - S(I_T \otimes \Gamma) \boldsymbol{\phi}^T\right] = \mathbf{0}. \tag{63}$$

Defining $\boldsymbol{\theta} = \left(\alpha, \Gamma, \boldsymbol{\phi}^T\right)$ the GMM estimator is

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\theta} \left(N^{-1} \sum_{i=1}^{N} \left[Z_i' \mathbf{u}_i - S(I_T \otimes \Gamma) \boldsymbol{\phi}^T\right]\right)' A_N \left(N^{-1} \sum_{i=1}^{N} \left[Z_i' \mathbf{u}_i - S(I_T \otimes \Gamma) \boldsymbol{\phi}^T\right]\right), \tag{64}$$

where A_N is a non-negative definite weight matrix.

Robertson, Sarafidis and Symons (2010) call estimators in this class Factor Instrumental Variable (FIV) estimators. They note that in most practical circumstances a set of linear restrictions can be demonstrated to hold among the parameters, namely the matrix Γ. These can be obtained by writing the model as

$$\mathbf{z}_{it}'(1-\alpha) = \boldsymbol{\lambda}_i' \phi_t + \varepsilon_{it}, \tag{65}$$

[29] The number of rows of S corresponds to the number of moment conditions available and the number of columns corresponds to the number of regressors (1 at present) times the number of time periods available squared.

where $\mathbf{z}_{it} = (y_{it}, y_{it-1})'$, and then multiplying through by $\boldsymbol{\lambda}_i$ and taking expectations:

$$E\left(\boldsymbol{\lambda}_i \mathbf{z}_{it}'\right)(1-\alpha) = \Sigma_\lambda \phi_t, \tag{66}$$

where $\Sigma_\lambda = E\left(\boldsymbol{\lambda}_i \boldsymbol{\lambda}_i'\right)$. The key point here is that the elements in $E\left(\boldsymbol{\lambda}_i \mathbf{z}_{it}'\right)$ include terms in various of the $\boldsymbol{\gamma}_s$ because the instrument set includes elements of $\mathbf{z}_{it}$, so the left-hand side of (66) is a linear function of the entries in Γ. For example, for the single-factor model the linear restrictions take the form

$$\gamma_{s+1} = \alpha\gamma_s + \sigma_\lambda^2 \phi_{s+1},\ s = 0, ..., T-1,$$

where $\sigma_\lambda^2 = E\left(\lambda_i^2\right)$. For $s = T-1$, $\gamma_{s+1} = \gamma_T = E\left(y_{iT}\lambda_i\right)$, which can be regarded as a constant to be estimated. Robertson, Sarafidis and Symons (2010) call the GMM estimator that exploits these restrictions FIVR (restricted FIV), in contrast to the estimator obtained when these restrictions are not imposed, FIVU (unrestricted FIV). They show that FIVU is asymptotically equivalent to the quasi-differenced GMM estimator of Ahn, Lee and Schmidt (2006), while FIVR is asymptotically more efficient.

Bai (2010) proposes controlling the correlations between the regressors and the common factor component using the method of Chamberlain (1982). In particular, consider the linear projection of y_{i0} on $\boldsymbol{\lambda}_i$:

$$y_{i0} = \delta_0^* + \boldsymbol{\lambda}_i' \boldsymbol{\phi}_0^* + \varepsilon_0^*, \tag{67}$$

which can be derived using the reduced form of y_{i0},

$$y_{i0} = \frac{1}{1-\alpha}\delta_0 + \boldsymbol{\lambda}_i' \sum_{j=0}^{\infty} \alpha^j \boldsymbol{\phi}_{-j} + \sum_{j=0}^{\infty} \alpha^j \varepsilon_{i,-j}, \tag{68}$$

with $\delta_0^* = \delta_0/(1-\alpha)$, $\boldsymbol{\phi}_0^* = \sum_{j=0}^{\infty} \alpha^j \boldsymbol{\phi}_{-j}$ and $\varepsilon_0^* = \sum_{j=0}^{\infty} \alpha^j \varepsilon_{i,-j}$.

This implies a system of $T+1$ equations

$$A\mathbf{y}_i^+ = \delta^+ + \Phi^+ \boldsymbol{\lambda}_i + \boldsymbol{\varepsilon}_i^+, \tag{69}$$

where $\mathbf{y}_i^+ = (y_{i0}, y_{i1}, ..., y_{iT})'$, $\delta^+ = (\delta_0^*, 0, ..., 0)'$, $\Phi^+ = (\boldsymbol{\phi}_0^*, \boldsymbol{\phi}_1, ..., \boldsymbol{\phi}_T)'$, $\boldsymbol{\varepsilon}_i^+ = (\varepsilon_{i0}^*, \varepsilon_{i1}, ..., \varepsilon_{iT})'$ and

$$A = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ -\alpha & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & -\alpha & 1 \end{bmatrix}.$$

Let $\Omega^+ = \Phi^+ \Sigma_\lambda \Phi^{+\prime} + \Sigma_{\varepsilon^+}$, the covariance of $\Phi^+ \boldsymbol{\lambda}_i + \boldsymbol{\varepsilon}_i^+$, and $\boldsymbol{\upsilon}_i^+ = A\mathbf{y}_i^+ - \delta^+$. The log-likelihood function for $\mathbf{y}_i^+$ is

$$-\frac{N}{2}\ln\left|\Omega^+\right| - \frac{1}{2}\sum_{i=1}^{N} \boldsymbol{\upsilon}_i^{+\prime}\left(\Omega^+\right)^{-1}\boldsymbol{\upsilon}_i^+. \tag{70}$$

Notice that since A is a lower-triangular matrix, $\det(A) = 1$ and therefore the Jacobian term does not enter into the likelihood. Bai suggests estimating (70) using a quasi-maximum likelihood (QML) procedure based on the ECM (expectation and conditional maximization) algorithm.

When the model includes covariates, $\mathbf{x}_{it}$, the reduced form of y_{i0} can be written as

$$y_{i0} = \delta_0^* + \boldsymbol{\lambda}_i' \boldsymbol{\phi}_0^* + \mathbf{w}_i' \boldsymbol{\psi}_0 + \varepsilon_0^*, \tag{71}$$

where $\mathbf{w}_i = vec(\mathbf{x}_i')$ and $\boldsymbol{\psi}_0$ is loosely speaking the linear projection of $\mathbf{x}_i$ on $\boldsymbol{\lambda}_i$. Hence the residual is $\boldsymbol{\upsilon}_i^+ = A\mathbf{y}_i^+ - \delta^+ - \Psi \mathbf{w}_i$, with $\Psi = \left[\boldsymbol{\psi}_0' \; I_T \otimes \boldsymbol{\beta}'\right]'$, and the likelihood function is identical to (70).

The attractive feature of the FIVR and QML estimators is that they can both allow a fixed effects specification as a special case, in which one of the factors is constant over time. Furthermore, they are valid under strongly and weakly exogenous regressors and permit unit roots. In this way, these estimators generalise the classical error components formulation for a wide range of panel data models. Moreover, FIV estimators share the traditional advantage of method of moments estimators in that they exploit only a set of orthogonality conditions and make no use of subsidiary assumptions such as homoskedasticity or other assumed distributional properties of the error process. One difference between FIV and QML is that in the former approach, the factors, the loadings of which are uncorrelated with the regressors, will enter into the residuals of the model, thus resulting in fewer parameters to be estimated. QML will estimate all factors, which can also be desirable if these factors have a structural significance.

When T is large, treating $\boldsymbol{\phi}_t$ fixed leads to an incidental parameters problem so the methods described above are not appropriate. One way to proceed is to treat $\boldsymbol{\phi}_t$ as random and use the panel feasible generalised median unbiased (PFGMU) estimator proposed by Phillips and Sul (2003). This involves using the residuals obtained from a first-stage panel median unbiased estimator to construct an invertible estimate of the error covariance matrix by means of a method of moments procedure, estimating the regression model using a feasible generalised FE (FGFE) estimator and subsequently calculating PFGMU using the median function of FGFE. Alternative methods for projecting out estimates of the factor loadings have been proposed by Moon and Perron (2004) and Bai and Ng (2004). All these methods are valid for large T only, and it is not straightforward to generalise them into models that include weakly exogenous regressors other than the lagged dependent variable.

6 Testing for Cross-Sectional Dependence

6.1 Available Tests for the Presence of Error Cross-Sectional Dependence

The null hypothesis of interest is

$$H_0 : \text{Cov}(\upsilon_{it}, \upsilon_{jt}) = 0 \text{ for all } t \text{ and all } i \neq j, \tag{72}$$

vs alternative hypothesis (2). Several tests for error cross-sectional dependence have been proposed in the literature. Perhaps the most widely known test is the LM statistic by Breusch and Pagan (1980). The basic idea of this kind of tests is to substitute in the score vector the parameter estimates obtained from the restricted model under the null hypothesis and check whether the null vector is sufficiently close to zero. It turns out that under the null the test statistic can be based on the residuals from individual-specific OLS regressions. Let

$$\hat{\rho}_{ij} = \hat{\rho}_{ji} = \frac{\sum_{t=1}^{T} \widehat{v}_{it}\widehat{v}_{jt}}{\left(\sum_{t=1}^{T} \widehat{v}_{it}^2\right)^{1/2}\left(\sum_{t=1}^{T} \widehat{v}_{jt}^2\right)^{1/2}}. \tag{73}$$

Then under H_0, as $T \to \infty$ for fixed N, we have

$$LM = T\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} \hat{\rho}_{ij}^2 \xrightarrow{d} \chi^2_{N(N-1)/2}, \tag{74}$$

where the number of degrees of freedom equals the number of distinct off-diagonal elements of the error covariance matrix.

As noted by Pesaran (2004) and Pesaran, Ullah and Yamagata (2008), the LM statistic (74) is likely to have poor size properties when N is large − clearly, an empirically relevant situation. Pesaran (2004) shows that when both N and T are large, (74) can be modified in a straightforward way. In particular under H_0, for any given pair $i \neq j$, we have

$$T\hat{\rho}_{ij}^2 \xrightarrow{d} \chi_1^2, \tag{75}$$

for $T \to \infty$. Therefore, since the $\hat{\rho}_{ij}^2$ are asymptotically uncorrelated, the following scaled version of the LM statistic can be considered:

$$LM_2 = \sqrt{\frac{1}{N(N-1)}}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\left(T\hat{\rho}_{ij}^2 - 1\right) \xrightarrow{d} N(0,1), \tag{76}$$

for $T \to \infty$ and $N \to \infty$ sequentially.

For fixed T both LM and LM_2 statistics are likely to exhibit substantial size distortions (Pesaran, 2004). This is mainly due to the fact that $E(T\hat{\rho}_{ij}^2 - 1)$ will not be correctly centered at zero when T is small and with large N the incorrect centering of the statistics is likely to be accentuated, resulting potentially in large size distortions.

Friedman (1937) proposed a non-parametric test, appropriate for large N and fixed T, based on Spearman's rank correlation coefficient. The latter can be thought of as the regular product-moment correlation coefficient except that it is computed from ranks. In particular, under the null, as $N \to \infty$ for T fixed we have

$$FR = (T-1)\left[(N-1)R_{AVE} + 1\right] \xrightarrow{d} \chi^2_{T-1},$$

where

$$R_{AVE} = \frac{1}{N(N-1)/2}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} \widehat{r}_{ij} \tag{77}$$

and $\widehat{r}_{ij}$ denotes Spearman's rank correlation coefficient given by

$$r_{ij} = r_{ji} = \frac{\sum_{t=1}^{T} \left(r_{i,t} - (T+1/2)\right)\left(r_{j,t} - (T+1/2)\right)}{\sum_{t=1}^{T} \left(r_{i,t} - (T+1/2)\right)^2}. \tag{78}$$

A closely related test is developed by Pesaran (2004). He proposes a simple alternative, based on regular product-moment correlation coefficients, which has exactly mean zero for fixed values of either N or T,

$$CD = \sqrt{\frac{2T}{N(N-1)}} \left(\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \hat{\rho}_{ij} \right) \overset{d}{\rightarrow} N(0,1). \tag{79}$$

Pesaran shows that the above statistic is valid under a wide class of panel data models, including heterogeneous models, dynamic models and regression models with multiple structural breaks, provided that the unconditional means of y_{it} and x_{it} are time-invariant and their innovations are symmetrically distributed. Chen, Gao and Li (2009) extend this method by developing a nonparametric counterpart of the CD statistic for testing error cross-sectional dependence in nonparametric models.

Both the CD and FR statistics share a common weakness in that they may lack power to detect the alternative hypothesis under which the sign of the elements of the error covariance matrix is alternating − that is, there are positive and negative correlations in the residuals. This can arise if, for example, cross-sectional dependence is characterised by a factor model with zero mean factor loadings. Notice that the same problem might arise even if the factor loadings have mean different from zero; one such instance is when time-specific dummies are included in the regression model to capture possible common variations in the dependent variable. In fact, this practice is not uncommon for fixed T and amounts to transforming the observations in terms of time-specific averages. Thus, suppose that the disturbance follows a single-factor process, as in (9), in which case time-specific demeaning yields process (11). Observe that

$$\text{Cov}\left(\underline{\upsilon}_{it}, \underline{\upsilon}_{jt}\right) = E\left(\underline{\lambda}_i\right) E\left(\underline{\lambda}_j\right) = 0. \tag{80}$$

Therefore the CD and $RAVE$ statistics will be centered around zero, which implies that the power of the tests will not increase with N and therefore they may be inconsistent.

Frees (1995) proposes a test statistic that is not subject to this problem and is valid for fixed T, large N. Specifically, define

$$Q = b_1(T)\left(\chi_1^2 - (T-1)\right) + b_2(T)\left(\chi_2^2 - T(T-3)/2,\right) \tag{81}$$

where χ_1^2 and χ_2^2 are independent chi-square distributed variables with $T-1$ and $T(T-3)/2$ degrees of freedom respectively and

$$b_1(T) = \frac{4(T+2)}{5(T-1)^2(T+1)},\ b_2(T) = \frac{2(5T+6)}{5T(T-1)(T+1)}. \tag{82}$$

Also, let

$$R^2_{AVE} = \frac{2}{N(N-1)/2} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \widehat{r}^2_{ij}. \tag{83}$$

Frees (1995) shows that $FRE = N\left[R^2_{AVE} - (T-1)^{-1}\right]$ follows asymptotically a Q distribution for $N \to \infty$, T fixed. Therefore, the null is rejected if R^2_{AVE} is larger than $(T-1)^{-1} + Q_q/N$, where Q_q is an appropriate quantile from the Q distribution.[30]

Pesaran, Ullah and Yamagata (2008) argue that the FRE statistic tends to behave similarly to the uncorrected version of the LM statistic for large N when the model involves more than one explanatory variable (intercept). They propose a bias-adjusted version of the LM test that makes use of the exact mean and variance of the LM statistic and is valid under strongly exogenous regressors and normal errors. This is defined as

$$LM_3 = \sqrt{\frac{2}{N(N-1)}} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \left[\frac{(T-K)\hat{\rho}^2_{ij} - \mu_{T_{ij}}}{\sigma^2_{T_{ij}}}\right] \xrightarrow{d} N(0,1), \tag{84}$$

where

$$\mu_{T_{ij}} = E\left[(T-K)\hat{\rho}^2_{ij}\right] = \frac{1}{T-K} tr(M_i M_j), \tag{85}$$

with $M_i = \left(I_T - X_i (X_i'X_i)^{-1} X_i'\right)$, $M_j = \left(I_T - X_j \left(X_j'X_j\right)^{-1} X_j'\right)$ and

$$\sigma^2_{T_{ij}} = var\left[(T-K)\hat{\rho}^2_{ij}\right] = \left[tr(M_i M_j)\right]^2 \alpha_{1T} + 2tr\left[(M_i M_j)\right]^2 \alpha_{2T}, \tag{86}$$

with $\alpha_{1T} = \alpha_{2T} - (T-K)^{-2}$ and

$$\alpha_{2T} = 3\left[(T-K-8)(T-K+2)+24\right]^2 \left[(T-K+2)(T-K-2)(T-K-4)\right]^{-2}. \tag{87}$$

Notice that the test statistic is feasible only when $T > K+8$, and it has exactly mean zero regardless of the value of T, unlike the LM statistic. On the other hand, unless T is large the covariance between $\sqrt{T-K}\hat{\rho}^2_{ij}$ and $\sqrt{T-K}\hat{\rho}^2_{ij'}$, for any $j \neq j'$, is different from zero even under the normality assumption. Therefore (84) is valid under the sequential asymptotic $T \to \infty$ first and then $N \to \infty$. Simulation evidence provided by Pesaran, Ullah and Yamagata (2008) indicate that the test has good size and power for $T \geq 20$.

Sarafidis, Robertson and Yamagata (2009) propose a testing procedure that does not require normality and is valid for fixed T, large N panel data models with weakly exogenous regressors. Their testing procedure is based on Sargan's difference-test statistic for overidentifying restrictions. In particular, consider the following model

$$\underline{\mathbf{y}}_i = \underline{X}_{w,i}\boldsymbol{\beta}_w + \underline{X}_{s,i}\boldsymbol{\beta}_s + \underline{\eta}_i \iota_T + \underline{\boldsymbol{v}}_i, \;\; \underline{\boldsymbol{v}}_i = \boldsymbol{\phi}^T \boldsymbol{\lambda}_i + \boldsymbol{\varepsilon}_i, \tag{88}$$

[30] de Hoyos and Sarafidis (2006) show how to perform all these tests in Stata using the command *xtcsd*; see http://ideas.repec.org/c/boc/bocode/s456736.html.

where $\underline{\mathbf{y}}_i$ is a $T \times 1$ vector of stacked time series observations expressed in terms of deviations from time-specific averages, and similarly for the remaining variables, while $\underline{X}_{w,i}$ and $\underline{X}_{s,i}$ are $T \times K_w$ and $T \times K_s$ matrices of weakly and strongly exogenous regressors respectively. The null hypothesis of interest is

$$H_0 : var(\boldsymbol{\lambda}_i) = \Sigma_\lambda = \mathbf{0} \tag{89}$$

against the alternative

$$H_1 : \Sigma_\lambda \neq \mathbf{0}, \tag{90}$$

as opposed to (72) and (2). The aim of the test is to examine whether there is cross-sectional dependence left out in the errors after time-specific demeaning takes place.[31] Let $\underline{X}_{\widetilde{s},i}$ be the $T \times K_{\widetilde{s}}$ matrix of regressors that remain strongly exogenous in the mispecified model. Thus, $\underline{X}_{\widetilde{s},i}$ is a subset of $\underline{X}_{s,i}$ and includes covariates, the factor loadings of which are either zero (so these covariates are not hit by the factors) or mutually uncorrelated with $\boldsymbol{\lambda}_i$.[32] Furthermore, let $\underline{Z}_i$ be the matrix of instrumental variables that makes use of the full set of moment conditions, while $\underline{Z}_{\widetilde{s},i}$ be the corresponding matrix that makes use of the moment conditions that arise with respect to $\underline{X}_{\widetilde{s},i}$ only. Sargan's (1958) or Hansen's (1982) test of overidentifying restrictions based on the full set of moment conditions is given by

$$S_F = N^{-1} \left(\sum_{i=1}^{N} \widehat{\underline{\boldsymbol{v}}}_i^{*\prime} \underline{Z}_i \right) \dot{\widehat{\Omega}}^{-1} \left(\sum_{i=1}^{N} \underline{Z}_i' \widehat{\underline{\boldsymbol{v}}}_i^{*\prime} \right), \tag{91}$$

where $\widehat{\underline{\boldsymbol{v}}}_i^*$ is the residual vector obtained from the following two-stage linear GMM estimator of $\boldsymbol{\beta} = \left(\boldsymbol{\beta}_w', \boldsymbol{\beta}_s\right)'$ with the general form

$$\ddot{\widehat{\boldsymbol{\beta}}}_F = \left(\sum_{i=1}^{N} \underline{W}_i^{*\prime} \underline{Z}_i \dot{\widehat{\Omega}}_F^{-1} \sum_{i=1}^{N} \underline{Z}_i' \underline{W}_i^* \right)^{-1} \sum_{i=1}^{N} \underline{W}_i^{*\prime} \underline{Z}_i \dot{\widehat{\Omega}}_F^{-1} \sum_{i=1}^{N} \underline{Z}_i' \underline{\mathbf{y}}_i^*, \tag{92}$$

where $\underline{\mathbf{y}}_i^*$ and $\underline{W}_i^*$ denote some transformation[33] of $\underline{\mathbf{y}}_i$ and $\underline{W}_i = \left(\underline{X}_{w,i} \vdots \underline{X}_{s,i} \right)$ respectively and $\dot{\widehat{\Omega}}_F$ is the estimated weight matrix obtained from a first-stage GMM estimator. Similarly, Sargan's/Hansen's test of overidentifying restrictions based on the subset of moment conditions with respect to $\underline{X}_{\widetilde{s},i}$ is given by

$$S_R = N^{-1} \left(\sum_{i=1}^{N} \widetilde{\underline{\boldsymbol{v}}}_i^{*\prime} \underline{Z}_{\widetilde{s},i} \right) \dot{\widetilde{\Omega}}^{-1} \left(\sum_{i=1}^{N} \underline{Z}_{\widetilde{s},i}' \widetilde{\underline{\boldsymbol{v}}}_i^{*\prime} \right), \tag{93}$$

[31] The authors phrase H_0 and H_1 as 'homogeneous' and 'heterogeneous' cross-sectional dependence respectively.

[32] Membership in the subset $\underline{X}_{\widetilde{s},i}$ is testable using Sargan's/Hansen's test for overidentifying restrictions.

[33] For example, first-differences, orthogonal deviations and so on.

where $\widetilde{\underline{\boldsymbol{\upsilon}}}_i^*$ is the residual obtained from the two-stage linear GMM estimator of the following general form

$$\widehat{\ddot{\boldsymbol{\beta}}}_R = \left(\sum_{i=1}^{N} \underline{W}_i^{*\prime} \underline{Z}_{\widetilde{s},i} \widehat{\dot{\Omega}}_R^{-1} \sum_{i=1}^{N} \underline{Z}_{\widetilde{s},i}^{\prime} \underline{W}_i^{*} \right)^{-1} \sum_{i=1}^{N} \underline{W}_i^{*\prime} \underline{Z}_{\widetilde{s},i} \widehat{\dot{\Omega}}_F^{-1} \sum_{i=1}^{N} \underline{Z}_{\widetilde{s},i}^{\prime} \underline{\mathbf{y}}_i^{*}, \tag{94}$$

with similar definitions applying as before (mutatis mutandis) and we assume that the number of columns of $\underline{Z}_{\widetilde{s},i}$ is larger than $\underline{W}_i^*$. Under the null hypothesis as $N \to \infty$ for fixed T,

$$D_{SYR} = (S_F - S_R) \xrightarrow{d} \chi^2_{h_d}, \tag{95}$$

where h_d is the difference between the number of columns of $\underline{Z}_i$ and $\underline{Z}_{\widetilde{s},i}$.

The D_{SYR} statistic is very general since it can be performed using alternative GMM estimators which are not necessarily asymptotically efficient. On the other hand, it requires $h_d > K_w + K_s$ valid moment restrictions under the alternative. This will be violated if, for example, the (non-zero) factor loadings of the covariates included in $\underline{X}_{\widetilde{s},i}$ are correlated with $\boldsymbol{\lambda}_i$, or if $\boldsymbol{\beta}_s = 0$ (all regressors are weakly exogenous in the correctly specified model).[34] Yamagata (2008) proposes testing for error cross-sectional dependence using a joint serial correlation test applied after estimating the model using the first-differenced GMM estimator (Arellano and Bond, 1991). Essentially the procedure involves an examination of the joint significance of estimates of second and up to pth-order (first-differenced) error serial correlations. The intuition of the test lies in that error cross-sectional dependence is also likely to show up as serial correlation in the residuals. To see this, consider the single-factor error process (9) and let $\lambda_i \sim i.i.d.\left(0, \sigma_\lambda^2\right)$. Applying time-specific demeaning and taking expectations, conditional upon ϕ_t, yields

$$E\left[\Delta\widehat{\underline{\upsilon}}_{it} \Delta\widehat{\underline{\upsilon}}_{t+s}\right] = E\left[\left(\underline{\lambda}_i \Delta\phi_t + \Delta\underline{\varepsilon}_{it}\right)\left(\underline{\lambda}_i \Delta\phi_{t+s} + \Delta\underline{\varepsilon}_{it+s}\right)\right] = \Delta\phi_t \Delta\phi_{t+s} \sigma_\lambda^2 \neq 0. \tag{96}$$

Notice that the magnitude of $E\left[\Delta\widehat{\underline{\upsilon}}_{it} \Delta\widehat{\underline{\upsilon}}_{t+s}\right]$ does not necessarily decrease as s increases for a given t. Therefore, the null hypothesis of interest becomes

$$H_0 : E\left[\Delta\widehat{\underline{\upsilon}}_{it} \Delta\widehat{\underline{\upsilon}}_{t+s}\right] = 0 \text{ jointly for } s = 2, 3, ..., p\,[\leq T-2], \tag{97}$$

against the alternative

$$H_1 : E\left[\Delta\widehat{\underline{\upsilon}}_{it} \Delta\widehat{\underline{\upsilon}}_{t+s}\right] \neq 0 \text{ for some } s, \tag{98}$$

and $t = 2, 3, ..., T-s$. Under the null hypothesis, as $N \to \infty$ for fixed T, the joint statistic for second up to pth-order serial correlation is

$$m^2_{(2,p)} = \boldsymbol{\iota}_N^{\prime} H \left(G^{\prime} G\right)^{-1} H^{\prime} \boldsymbol{\iota}_N \xrightarrow{d} \chi^2_{(p-1)}, \tag{99}$$

where $H = (\boldsymbol{\nu}_1, ..., \boldsymbol{\nu}_N)^{\prime}$, $\boldsymbol{\nu}_i = (\nu_{i2}, ..., \nu_{ip})^{\prime}$, $\nu_{is} = \sum_{t=2}^{T-s} \Delta\widehat{\underline{\upsilon}}_{it} \Delta\widehat{\underline{\upsilon}}_{t+s}$, $G = (\mathbf{g}_1, ..., \mathbf{g}_N)^{\prime}$, $\mathbf{g}_i = (g_{i2}, ..., g_{ip})^{\prime}$, $g_{is} = \nu_{is} - \psi_{Ns}^{\prime} Q_N^{-1} A_N^{\prime} \widehat{\ddot{\Omega}}^{-1} \underline{Z}_i^{\prime} \Delta\widehat{\underline{\boldsymbol{\upsilon}}}_i$, $\psi_{Ns} = N^{-1} \sum_{i=1}^{N} \left(\sum_{t=2}^{T-s} \Delta\widehat{\underline{\upsilon}}_{it} \Delta\underline{\mathbf{w}}_{it+s} \right)$,

[34] In this case, the null hypothesis could be addressed using a simple overidentifying restrictions test.

$Q_N = A_N' \hat{\ddot{\Omega}} A_N$, $A_N = \left(\mathbf{N}^{-1} \sum_i^N \underline{Z}_i' \Delta \underline{W}_i\right) \hat{\ddot{\Omega}} \left(N^{-1} \sum_i^N \Delta \underline{W}_i' \underline{Z}_i\right)$ and $\hat{\ddot{\Omega}}$ is the estimated weighting matrix obtained from the two-stage first-differenced GMM estimator. It would be interesting to extend this approach to alternative models and estimation methods but we do not have any results as yet.

6.2 Determining the Number of Factors

Once the null hypothesis of no (heterogeneous) error cross-sectional dependence is rejected, an important issue comes into play for all estimators allowing for a multi-factor error structure except the CCE estimator; this is how to determine the appropriate number of factors. The simplest way to decide upon this is to use the 'Kaiser criterion', which retains all those factors associated with eigenvalues that are above average, or equivalently greater than one for standardised data. Intuitively, this is because the chosen factor must extract as much variation as the original variable. However, in practice this criterion is often found to be too conservative.

Bai and Ng (2002) propose determining the number of factors by minimising certain model selection information criterion functions. In particular, consider the $T \times K$ matrix of observed variables $X_i = (\mathbf{x}_{i1}, ..., \mathbf{x}_{iT})'$ and let

$$W_i = \boldsymbol{\phi}_{(M_0)}^T \boldsymbol{\lambda}_i + \boldsymbol{\varepsilon}_i, \tag{100}$$

where $\boldsymbol{\phi}_{(M_0)}^T$ contains T observations for each of the $M_0 \times 1$ largest principal components of the covariance matrix of W_i. The main task is to estimate M_0. Define

$$V\left(M, \widehat{\boldsymbol{\phi}}_{(M)}^T\right) \equiv \frac{1}{NT} \sum_i^N \left(W_i' W_i - W_i' P_{\boldsymbol{\phi}_{(M)}^T} W_i\right), \tag{101}$$

where $P_{\boldsymbol{\phi}_{(M)}^T}$ is the projection of W_i onto the column space defined by $\boldsymbol{\phi}_{(M)}^T$, for any $M \leq M_{\max}$, where is $M_{\max}$ is the maximum possible value of M_0 . Bai and Ng (2002) estimate M_0 as the solution to either one of the following minimisation problems:

$$\widehat{M_1} = \arg \min_{M \leq M_{\max}} \left[\ln V\left(M, \widehat{\boldsymbol{\phi}}_M^T\right) + M \left(\frac{N+T}{NT}\right) \ln\left(\frac{NT}{N+T}\right)\right], \tag{102}$$

$$\widehat{M_2} = \arg \min_{M \leq M_{\max}} \left[\ln V\left(M, \widehat{\boldsymbol{\phi}}^T\right) + M \left(\frac{N+T}{NT}\right) \ln C_{NT}^2\right], \tag{103}$$

and

$$\widehat{M_3} = \arg \min_{M \leq M_{\max}} \left[\ln V\left(M, \widehat{\boldsymbol{\phi}}^T\right) + M \left(\frac{\ln C_{NT}^2}{C_{NT}^2}\right)\right], \tag{104}$$

where $C_{NT}^2 = \min(N, T)$. The authors demonstrate that (102)-(104) are asymptotically equivalent and they estimate the true number of factors consistently as $\min(N, T) \to \infty$, i.e. $\widehat{M_j} \xrightarrow{p} M_0$ for $j = 1, 2, 3$. In finite samples the performance of the above information criteria will be different. Using simulated data, Bai and Ng show that $\widehat{M_1}$ and $\widehat{M_2}$ are

more robust than $\widehat{M_3}$ when either N or T is fairly small and they perform well so long as $min\,(N,T) \geq 40$. Otherwise, these criteria may not work well, leading to too many factors being estimated.

Kapetanios (2009) proposes a different method to determine the appropriate number of factors. This is based on the result that the largest eigenvalue of the sample covariance matrix of the data converges almost surely to $(1+\sqrt{c})^2$, where $c = \lim_{N,T\to\infty} \frac{N}{T}$, which implies that if there is no factor structure in the data, the maximum eigenvalue of the sample covariance matrix should not exceed $(1+\sqrt{c})^2$ almost surely, in large samples. Therefore, the method starts essentially be checking whether the factor structure is supported by the data at all, using as threshold the value $(1+\sqrt{c})^2 + d$, where $d > 0$ is chosen a priori. Kapetanios suggests choosing for d the mean eigenvalue of the covariance matrix (for standardised data this equals to 1). Hence, if the maximum eigenvalue of the covariance matrix exceeds this threshold, the maximum principal component is obtained and the data are orthogonalised from a regression on the first principal component. Next, the maximum eigenvalue of the resulting covariance matrix is compared against $(1+\sqrt{c})^2 + d$ and the process is repeated until the maximum eigenvalue of the resulting covariance matrix does not exceed the threshold value. Using simulated data, Kapetanios shows that in a majority of circumstances of empirical interest this method outperforms the information criteria (102)-(104).

The method proposed by Kapetanios requires that the idiosyncratic errors of the approximate factor model are i.i.d. Onatski (2007) develops a similar estimator that makes less stringent assumptions on the serial correlation and heteroskedasticity pattern of the idiosyncratic errors. His method is based on the mirror image of Kapetanios' argument, i.e. for data characterised by M_0 latent common factors, the largest M_0 eigenvalues of the covariance matrix of the data grow with N, while the rest of the eigenvalues are bounded. Hence, the Onatski estimator equals the number of eigenvalues greater than a threshold value:

$$\widehat{M_4} = \arg \max_{M \leq M_{\max}} \left[M | \mu_M > (1+\delta)\, c_1 \right], \tag{105}$$

where μ_M denotes the M^{th} largest eigenvalue of the sample covariance matrix of W_i, δ is a parameter to be chosen a priori and $c_1 = \vartheta \mu_{M_{\max}+1} + (1-\vartheta)\,\vartheta \mu_{2M_{\max}+1}$, with $\vartheta = 2^{2/3}\left(2^{2/3}-1\right)^{-1}$, is a threshold obtained from the empirical distribution of the eigenvalues to distinguish the diverging ones from the bounded ones. Under the assumption that the idiosyncratic errors of the approximate factor model are either serially uncorrelated, or cross-sectionally independent (but not both), the above estimator is shown to be consistent.

Ahn and Horenstein (2008) argue that the above methods can be somewhat generous in penalizing large M. Another potential problem in using these methods is that they all require a choice of $M_{\max}$. In large samples this is certainly not an issue provided that $M_{\max} > M_0$. However, in finite samples the estimate of M_0 could be sensitive to the choice of $M_{\max}$. To this end, Ahn and Horenstein propose estimating the number of factors by maximising the ratio of two adjacent eigenvalues, or the ratio of their growth

rate. In particular, we have

$$\widehat{M}_5 = \arg \max_{M \leq M_{\max}} \left[\mu_M / \mu_{M+1} \right], \tag{106}$$

and

$$\widehat{M}_6 = \arg \max_{M \leq M_{\max}} \left[\frac{\ln \left(\mu_M^* \right)}{\ln \left(\mu_{M+1}^* \right)} \right], \tag{107}$$

where $\mu_M^* = \left(\sum_{j=M}^{T} \mu_j \right) / \left(\sum_{j=M+1}^{T} \mu_j \right)$ and $M \leq M_{\max}$. Using simulated data they show that the proposed estimators outperform the existing ones even in samples with small N and T unless the signal-to-noise ratio of the model is too small.[35]

Notice that the set-up in all the above methods is such that the factors are extracted from observed variables. Therefore, it is not clear what the properties of these methods are when the factors are extracted from estimated residuals, which is precisely what is of main interest in this paper. We explore this issue via Monte Carlo experiments.

6.3 A Monte Carlo Study

6.3.1 Design

The underlying data generating process is given by

$$\begin{aligned} y_{it} &= \beta_1 x_{1it} + \beta_2 x_{2it} + \omega_{it}, \, \omega_{it} = \eta_i + \upsilon_{it}, \;\; \upsilon_{it} = \sum_{m=1}^{M} \lambda_i^m \phi_t^m + \varepsilon_{it}, \\ x_{1it} &= \sum_{m=1}^{M} \lambda_i^{*m} \phi_t^m + \varepsilon_{it}^*, \;\; i = 1, ..., N, \, t = 1, ..., T, \end{aligned} \tag{108}$$

which is similar to Section 4.4 except that we add an extra regressor, x_{2it}. To examine the impact of strict/weak exogeneity on the properties of the tests we set (i) $x_{2it} = \eta_i + \varpi_{it}$, $\varpi_{it} \sim i.i.d.N\left(\mu_{x_2}, \sigma_{x_2}^2 \right)$ and (ii) $x_{2it} = y_{it-1}$. In the former case we specify the parameters such that the signal-to-noise ratio depends solely on the slope coefficients, β_1 and β_2. In particular, define $y_{it}^* = y_{it} - \eta_i$ such that

$$y_{it}^* = \beta_1 x_{1it} + \beta_2 x_{2it} + \upsilon_{it}, \tag{109}$$

and let the signal-to-noise ratio be denoted by $\zeta = \sigma_s^2 / \sigma_\upsilon^2$, where σ_s^2 is the variance of the signal and σ_υ^2 is the total error variance. The signal variance equals

$$\sigma_s^2 = var\left(\beta_1 x_{1it}^* + \beta_2 x_{2it}^2 \right) = var\left(y_{it}^* - \upsilon_{it} \right) = var\left(y_{it}^* \right) + var\left(\upsilon_{it} \right) - 2cov\left(y_{it}^*, \upsilon_{it} \right). \tag{110}$$

[35] The issue of determining the number of factors in the dynamic factor model case is analysed, among others, by Amengual and Watson (2007) and Hallin and Liska (2007).

We consider each of the terms in (110) sequentially. We have

$$\begin{aligned} var\left(y_{it}^{*}\right) &= \beta_1^2 var\left(x_{1it}\right)+\beta_2^2 var\left(x_{2it}\right)+var\left(\upsilon_{it}\right)+2\beta_1 cov\left(x_{1it},\upsilon_{it}\right)= \\ &= \beta_1^2\left[\sum_{m=1}^{M}\sigma_{\lambda_m^*}^2+\sum_{m=1}^{M}\mu_{\lambda_m^*}^2+\sigma_{\varepsilon^*}^2\right]+\beta_2^2\sigma_{x_2}^2 \\ &\quad +\left[\sum_{m=1}^{M}\sigma_{\lambda_m}^2+\sum_{m=1}^{M}\mu_{\lambda_m}^2+\sigma_{\varepsilon}^2\right]+2\beta_1\sum_{m=1}^{M}\rho_{\lambda_m}\sigma_{\lambda_m}\sigma_{\lambda_m^*}, \end{aligned} \tag{111}$$

$$var\left(\upsilon_{it}\right)=\sum_{m=1}^{M}\sigma_{\lambda_m}^2+\sum_{m=1}^{M}\mu_{\lambda_m}^2+\sigma_{\varepsilon}^2, \tag{112}$$

and

$$\begin{aligned} -2cov\left(y_{it}^{*},\upsilon_{it}\right) &= -2cov\left(\beta_1 x_{1it}+\beta_2 x_{2it}^2+\upsilon_{it},\upsilon_{it}\right)=-2\left[\beta_1 cov\left(x_{1it},\upsilon_{it}\right)+var\left(\upsilon_{it}\right)\right] \\ &= -2\beta_1\sum_{m=1}^{M}\rho_{\lambda_m}\sigma_{\lambda_m}\sigma_{\lambda_m^*}+\left[\sum_{m=1}^{M}\sigma_{\lambda_m}^2+\sum_{m=1}^{M}\mu_{\lambda_m}^2+\sigma_{\varepsilon}^2\right]. \end{aligned} \tag{113}$$

Setting

$$\begin{aligned} \sigma_{\varepsilon}^2 &= \sigma_{\varepsilon^*}^2; \\ \sigma_{x_2}^2 &= \left[\sum_{m=1}^{M}\sigma_{\lambda_m^*}^2+\sum_{m=1}^{M}\mu_{\lambda_m^*}^2+\sigma_{\varepsilon^*}^2\right]; \text{ and} \\ \mu_{\lambda_m} &= \mu_{\lambda_m^*},\ \sigma_{\lambda_m^*}^2=\sigma_{\lambda_m}^2,\ \rho_{\lambda_m}=\rho_{\lambda} \text{ for } m=1,..,M, \end{aligned} \tag{114}$$

and combining (110)-(113) yields

$$\sigma_s^2=\left(\beta_1^2+\beta_2^2\right)\left[\sum_{m=1}^{M}\sigma_{\lambda_m}^2+\sum_{m=1}^{M}\mu_{\lambda_m}^2+\sigma_{\varepsilon}^2\right].$$

Therefore,

$$\zeta=\sigma_s^2/\sigma_{\upsilon}^2=\beta_1^2+\beta_2^2. \tag{115}$$

In the case of weak exogeneity (so $x_{2it}=y_{it-1}$) we define $y_{it}^{*}=y_{it}-\frac{\eta_i}{1-\beta_2}$ such that

$$y_{it}^{*}=\frac{\beta_1}{1-\beta_2}x_{1it}+\frac{1}{1-\beta_2}\upsilon_{it}. \tag{116}$$

The variance of the signal equals

$$\begin{aligned} \sigma_s^2 &= var\left(y_{it}^{*}-\upsilon_{it}\right)=var\left(y_{it}^{*}\right)+var\left(\upsilon_{it}\right)-2cov\left(y_{it}^{*},\upsilon_{it}\right) \\ &= \left(\frac{\beta_1}{1-\beta_2}\right)^2 var\left(x_{1it}\right)+\left(\frac{1}{1-\beta_2}\right)^2 var\left(\upsilon_{it}\right)+2\frac{\beta_1}{\left(1-\beta_2\right)^2}cov\left(x_{1it},\upsilon_{it}\right)+var\left(\upsilon_{it}\right) \\ &\quad -\frac{2}{1-\beta_2}\left[\beta_1 cov\left(x_{1it},\upsilon_{it}\right)+var\left(\upsilon_{it}\right)\right]. \end{aligned} \tag{117}$$

Using (114) and imposing $\rho_\lambda = 0$ it is straightforward to show that

$$\zeta = \sigma_s^2/\sigma_v^2 = \left(\beta_1^2 + \beta_2^2\right) / \left(1 - \beta_2\right)^2 . \tag{118}$$

To examine the impact of the signal-to-noise ratio on the performance of the statistics we set $\zeta \in \{1, 5\}$ and we select $\beta_k = \sqrt{\zeta/2}$ for $k = 1, 2$. As in Section 4.4 we choose values for the remaining parameters subject to the three fractions Φ_1, Φ_2, and Φ_3. Normalising $\sigma_\varepsilon^2 = 1$ implies that these fractions parameterise completely σ_η^2, $\sum_{m=1}^2 \sigma_{\lambda_m}^2$ and $\sum_{m=1}^2 \mu_{\lambda_m}^2$. To simplify things we let $\sigma_{\lambda_m}^2 = M^{-1} \sum_{m=1}^M \sigma_{\lambda_m}^2$, $\mu_{\lambda_m} = \left(M^{-1} \sum_{m=1}^M \mu_{\lambda_m}^2\right)^{1/2}$ for all m. Further, we fix $\Phi_2 = 0.9$ and we set $\Phi_1 \in \{0.5, 0.9\}$ and $\Phi_3 \in \{0.5, 0.8\}$ to examine, respectively, the impact of the relative size of the purely idiosyncratic error component over factor noise and the closeness of the factor structure to an ordinary time effect. We expect the performance of the statistics to deteriorate with high values of Φ_1 and Φ_3. Notice, however, that as Φ_1 approaches 1 the impact of the factor structure on the properties of the estimates of the structural parameters is likely to become smaller. Furthermore, when $\Phi_3 = 1$ the multi-factor structure degenerates to a single individual-invariant effect which can be accounted for using time-specific dummy variables. Finally, it is also worth pointing out that consistent estimation of the structural parameters only requires that $\widehat{M} \geq M_0$. Therefore, the cost of underestimating the number of factors is greater than estimating more factors.

We consider $M \in \{1, 3\}$ and $N = 100$, $T \in \{10, 50, 100\}$. As before, we perform 2,000 replications. The starting value of y, y_{i0}, is drawn from a stationary process. All statistics are calculated using the residuals obtained by OLS for each individual. Prior to computation of the eigenvectors each i-specific residual vector is standardised to have unit variance.

6.3.2 Results

Table 2 reports the results in terms of the frequency of the statistics to select the true number of factors, M_0. If the statistic selects an incorrect number of factors with higher frequency than M_0, then we report both frequencies, as well as the value of $\widehat{M} \neq M_0$ in brackets. For example, '.000 (0; 1.00)' means that the frequency of selecting M_0 is zero and the statistic has selected $M = 0$ with frequency 1. The $\widehat{M}_j$ refer to the corresponding statistics defined in (102)-(107). Following Onatski (2007) we choose $\delta \in \left\{0, \max\left(N^{-1/2}, T^{-1/2}\right), \max\left(N^{-2/3}, T^{-2/3}\right)\right\}$. Therefore, $\widehat{M}_4$ contains three cases, $\widehat{M}_{4_{(1)}}$, $\widehat{M}_{4_{(2)}}$ and $\widehat{M}_{4_{(3)}}$ that correspond to each of these different values of δ respectively.

As we can see, the performance of the statistics varies across different experiments and depends crucially upon the size of M_0 (the smaller the better), T (the larger the better) and the values of Φ_1 and Φ_2 (the smaller the better). For $T = 100$, most statistics perform well even if the factors are extracted from residuals rather than observed variables, unless Φ_1 and Φ_2 are both close to unity and $M_0 = 1$. In this case all statistics heavily underestimate M_0 although $\widehat{M}_{4_{(1)}}$ and $\widehat{M}_{4_{(3)}}$ do less so than the others. This finding is not surprising because most of the noise is idiosyncratic in this case. $\widehat{M}_3$

outperforms $\widehat{M}_1$ and $\widehat{M}_2$, unless $\Phi_1 = \Phi_2 = 0.5$, in which case it compares less favorably. $\widehat{M}_{4_{(1)}}$ does relatively poorly in most circumstances but $\widehat{M}_{4_{(2)}}$ and $\widehat{M}_{4_{(3)}}$ perform quite well even for large values of either Φ_1 or Φ_2, although they are both sensitive to small values of T. $\widehat{M}_5$ appears to perform well even for $T = 10$ and it underestimates M_0 mostly when Φ_1 is large. Similar results have been obtained for $\zeta = 5$ and $\rho_\lambda = 0.5$ and therefore it appears that these two parameters are not crucial for the performance of the statistics. Furthermore, we have reached similar conclusions for the case where $x_{2it} = y_{it-1}$ but to save space we do not report these here.[36]

In summary, we may argue that some of the statistics considered here, based on residuals rather than observed variables, perform reasonably well (especially $\widehat{M}_{4_{(2)}}$, $\widehat{M}_{4_{(3)}}$ and $\widehat{M}_5$) under both strong and weak exogeneity, unless a large proportion of the variation in total noise is due to the purely idiosyncratic component, or there is little variation in the factor loadings, or T is small. The first case might be less of a problem in practice because the impact of the factor structure can be small while the second case can be accounted for quite effectively using time dummies. Determining the number of factors under small T is certainly an issue that requires further research.

To this end, Ahn, Lee and Schmidt (2006) propose determining the number of factors using a sequential method, based on GMM and Sargan's (1958) or Hansen's (1982) test statistic. Their method is appropriate for fixed T. The intuition of this approach is that if $\widehat{M} < M_0$, this is likely to show up as a significant overidentifying restrictions test statistic. Therefore, one may start by testing the null $M_0 = 0$ against the alternative $M_0 > 0$. Then if the null is rejected, one can move to test the null $M_0 = 1$ against the alternative $M_0 > 1$ and so one until the null hypothesis is not rejected. Naturally, the significance level used for this sequential method needs to be appropriately adjusted. This approach is valid under strongly and weakly exogenous regressors. In the former case, it will identify only the factors whose factor loadings are correlated with the regressors. An alternative approach can be based on the joint serial correlation test (Yamagata, 2008) combined with a GMM estimator that allows for factor residuals in the same sequential manner, but we have no results as yet. A further possibility under fixed T is to construct a criterion based on a likelihood ratio test statistic. Lawley and Maxwell (1971, section 2.6) provide details for the case of extracting latent factors from observed variables, although the case of extracting factors from regression residuals remains unexplored in the literature. We do expect the issue of determining the number of factors in fixed T cases to attract more attention in the near future.

[36] The results are available from the authors upon request.

Table 2 Performance of statistics for selecting the number of factors, $\zeta = 1$.

T	M_0	Φ_1	Φ_3	$\widehat{M}_1$	$\widehat{M}_2$	$\widehat{M}_3$	$\widehat{M}_{4_{(1)}}$	$\widehat{M}_{4_{(2)}}$	$\widehat{M}_{4_{(3)}}$	$\widehat{M}_5$	$\widehat{M}_6$
100	1	0.5	0.5	1.00	1.00	.995	.314 (2; .45)	.885	.613	1.00	1.00
50	1	0.5	0.5	1.00	1.00	1.00	.174 (2; .43)	.835	.562	1.00	1.00
10	1	0.5	0.5	.003 (6; .97)	.054 (6; .70)	.000 (0; 1.0)	.156 (0; .84)	.002 (0; .99)	.006 (0; .99)	.904	.000 (7; 1.0)
100	1	0.9	0.5	1.00	.994	1.00	.351 (2; .45)	.913	.672	1.00	1.00
50	1	0.9	0.5	.850	.609	.999	.208 (2; .45)	.880	.602	1.00	1.00
10	1	0.9	0.5	.000 (6; .70)	.001 (0; .71)	.000 (6; .98)	.001 (0; .99)	.000 (0; 1.0)	.000 (0; 1.0)	.218 (7; .25)	.000 (7; 1.0)
100	1	0.5	0.8	1.00	1.00	.999	.301 (2; .48)	.902	.628	1.00	1.00
50	1	0.5	0.8	1.00	1.00	1.00	.184 (2; .43)	.850	.554	1.00	1.00
10	1	0.5	0.8	.007 (6; .93)	.026 (6; .63)	.000 (6; 1.0)	.074 (0; .93)	.000 (0; 1.0)	.001 (0; 1.0)	.849	.000 (7; 1.0)
100	1	0.9	0.8	.448 (0; .55)	.055 (0; .95)	1.00	.368 (2; .42)	.907	.682	1.00	1.00
50	1	0.9	0.8	.023 (0; .97)	.002 (0; .99)	.573	.201 (2; .45)	.856	.590	.979	.992
10	1	0.9	0.8	.000 (6; .88)	.000 (6; .61)	.000 (6; .99)	.000 (0; 1.0)	.000 (0; 1.0)	.000 (0; 1.0)	.147 (7; .30)	.000 (7; 1.0)
100	3	0.5	0.5	1.00	1.00	.907	.726	.980	.907	1.00	.119 (1; .72)
50	3	0.5	0.5	.998	.998	1.00	.649	.998	.914	.999	.118 (1; .67)
10	3	0.5	0.5	.000 (6; .99)	.000 (6; .98)	.000 (6; 1.0)	.002 (0; .53)	.000 (0; .97)	.000 (0; .90)	.340	.000 (7; 1.0)
100	3	0.9	0.5	.238 (0; .55)	.012 (0; .75)	.999	.813	.995	.995	.785	.000 (0; .99)
50	3	0.9	0.5	.017 (1; .68)	.000 (1; .60)	.619	.720	.938	.926	.449	.002 (1; .96)
10	3	0.9	0.5	.000 (6; 1.0)	.000 (6; .50)	.000 (6; .99)	.000 (0; .60)	.000 (0; 1.0)	.000 (0; 1.0)	.074 (1; .34)	.000 (7; 1.0)
100	3	0.5	0.8	.014 (1; .72)	.000 (1; .98)	.997	.796	.988	.944	.746	.000 (1; 1.0)
50	3	0.5	0.8	.001 (1; .96)	.000 (1; 1.0)	.237 (2; .54)	.701	.770	.841	.449	.000 (1; 1.0)
10	3	0.5	0.8	.000 (6; .97)	.000 (6; .71)	.000 (6; 1.0)	.000 (0; .95)	.000 (0; 1.0)	.000 (0; 1.0)	.094 (1; .90)	.002 (7; 1.0)
100	3	0.9	0.8	.000 (0; .55)	.000 (0; .96)	.000 (1; .98)	.296 (2; .52)	.017 (1; .68)	.108 (2; .51)	.000 (1; 1.0)	.000 (1; 1.0)
50	3	0.9	0.8	.000 (0; .96)	.000 (0; .99)	.000 (1; .59)	.359 (2; .37)	.018 (1; .78)	.096 (1; .50)	.005 (1; .97)	.002 (1; .99)
10	3	0.9	0.8	.000 (6; .89)	.000 (6; .62)	.000 (6; .99)	.000 (0; 1.0)	.000 (0; 1.0)	.000 (0; 1.0)	.085 (7; .31)	.000 (7; 1.0)

7 Current Challenges and Future Directions

There have been several major advances in the theoretical literature of panel data analysis with error cross-sectional dependence over the last ten years. Methods developed for dealing with fixed- and large-T cases, strongly and weakly exogenous regressors, non-stationary panels and testing for non-zero correlations across individuals have all helped to (re)address more effectively the issue of cross-sectional dependence and ultimately that of unobserved heterogeneity. Notwithstanding, there is still an abundance of non-trivial problems that require research attention. For instance, the literature is mute on dealing with cross-sectional dependence in non-linear panel data models, in which case it is typically assumed, for identification purposes rather than descriptive accuracy, that all observations are independent across individuals. Testing for cross-sectional dependence in non-linear models is not straightforward either, although some progress has been made by Hsiao, Pesaran and Pick (2009). There is a large range of other models that await possible extensions of the existing methods, such as panel VARs with a multi-factor error structure, systems of simultaneous equations and models with heterogeneous coefficients.

Finally, there is yet a relatively small empirical literature that deals with cross-sectional dependence in practice. It will be useful, as well as interesting, to see the extent to which economic applications can benefit from theoretical advances in the field.

References

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

[2] Ahn, S. C. and Horenstein, A. 2008. Eigenvalue ratio test for the number of factors. Mimeo.

[3] Ahn, S. C., Y. H. Lee and P. Schmidt. 2006. GMM estimation of linear panel data models with time-varying individual effects. Journal of Econometrics, 101, 219-255.

[4] Ahn, S. C., Y. H. Lee and P. Schmidt. 2006. Panel Data Models with Multiple Time-Varying Individual Effects. Mimeo.

[5] Amengual, D. and Watson, M. W. 2007. Consistent estimation of the number of dynamic factors in a large N and T panels. Journal of Business & Economic Statistics, 25(1), 91-96.

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

[7] Arellano, M. 2003. Panel Data Econometrics. Oxford University Press, Oxford.

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

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

[10] Bai, J. 2009. Panel Data Models with Interactive Fixed Effects. Econometrica, 77, 1229-1279.

[11] Bai, J. 2010. Likelihood approach to small T dynamic panel models with interactive effects. Mimeo.

[12] Bai, J. and Ng, S. 2002. Determining the Number of Factors in Approximate Factor Models. Econometrica, 70, 191-22.

[13] Bai, J. and Ng, S. 2002. A PANIC Attack on Unit Roots and Cointegration. Econometrica, 72(4), 1127-1177.

[14] Baltagi, B. 2008. Econometric Analysis of Panel Data, 4th ed. John Willey & Sons, West Sussex.

[15] Baltagi B. H. and Pesaran M. H. 2007. Heterogeneity and cross section dependence in panel data models: theory and applications - Introduction. Journal of Applied Econometrics, 22(2), 229-232.

[16] Bekker, P.A. 1994. Alternative Approximations to the Distributions of Instrumental Variable Estimators. Econometrica, 62, 657-681.

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

[18] Breitung, J. and Pesaran, M.H. 2008. Unit Roots and Cointegration in Panels, in L. Matyas and Sevestre P. (eds.) The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice, Kluwer Academic Publishers.

[19] Breusch, T. and A. Pagan. 1980. The Lagrange multiplier test and its application to model specification in econometrics. Review of Economic Studies 47, 239-253.

[20] Chen, J., Gao, J. and Li, D. 2009. A New Diagnostic Test for Cross–Section Independence in Nonparametric Panel Data Models. Mimeo.

[21] Chudik, A. Pesaran, M. H. and Tosetti, E. 2009. Weak and Strong Cross Section Dependence and Estimation of Large Panels. Mimeo.

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

[23] Conley, T.G. 1999. GMM Estimation with Cross Sectional Dependence. Journal of Econometrics, 92,1-45.

[24] Conley, T.G., and Topa, G. 2002. Socio-economic Distance and Spatial Patterns in Unemployment. Journal of Applied Econometrics, 17, 303-327.

[25] de Hoyos, R. E. and Sarafidis, V. 2006. Testing for Cross-sectional Dependence in Panel Data Models. The Stata Journal 6(4): 482-496.

[26] Driscoll, J.C., and Kraay, A.C. 1998. Consistent Covariance Matrix Estimation with Spatially Dependent Data. The Review of Economics and Statistics, 80, 549-560.

[27] Fiebig, D. G. 2001. Seemingly Unrelated Regression, in Baltagi, B. eds, A Companion to Theoretical Econometrics, Backwell Publishers, 101-121.

[28] Fisher, R.A. 1935. The Design of Experiments. Oliver and Boyd, Edinburgh.

[29] Forni, M. and Lippi, M. 2001. The Generalized Dynamic Factor Model: Representation Theory. Econometric Theory 17, 1113-1141.

[30] Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2000). The generalized factor model: identification and estimation. The Review of Economics and Statistics, 82, 540-554.

[31] Frees, E. W. 1995. Assessing Cross-sectional Correlation in Panel Data. Journal of Econometrics, 69, 393–414.

[32] Friedman, M. 1937. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. Journal of the American Statistical Association 32, 675–701.

[33] Goldberger, A. 1972. Structural equation methods in the social sciences. Econometrica 40 (6), 979–1001.

[34] Hallin, M. and Liška, R. 2007. Determining the number of factors in the general dynamic factor model. Journal of the American Statistical Association, 102(478), 603-617.

[35] Hansen, L. P. 1982. Large Sample Properties of Generalized Method of Moments Estimators, Econometrica, 50, 1029-1054.

[36] Hayakawa, K. 2009. Bias Corrected Estimation of Dynamic Panel Data Models with Interactive Fixed Effects. Mimeo.

[37] Holtz-Eakin D, Newey W. and Rosen H. 1988. Estimating Vector Autoregressions with Panel Data. Econometrica, 56, 1371-1395.

[38] Hurlin, C., Mignon, V. 2004. Second generation panel unit root tests. Mimeo.

[39] Hsiao, C. Analysis of Panel Data. 2nd ed. Cambridge University Press, Cambridge.

[40] Hsiao, C. 2007. Panel Data Analysis - Advantages and Challenges. TEST. Vol. 16, pp. 1-22.

[41] Hsiao, C., Pesaran, M. H. and Pick, A. 2009. Diagnostic Tests of Cross Section Independence for Nonlinear Panel Data Models. Mimeo.

[42] Jöreskog, K. G. and Goldberger, A. S. 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.

[43] Kapetanios, G. An Alternative Method for Determining the Number of Factors in Factor Models with Large Data Sets. Kapetanios G. Journal of Business and Economic Statistics, forthcoming.

[44] Kapetanios, G., and Pesaran, M. H. 2007. Small Sample Properties of Cross Section Augmented Estimators for Panel Data Models with Residual Multi-factor Structures; with M. H. Pesaran. In The Refinement of Econometric Estimation and Test Pro-cedures: Finite Sample and Asymptotic Analysis, Garry Phillips and Elias Tzavalis (eds.), Cambridge University Press, Cambridge.

[45] Kapetanios, G., Pesaran, M. H. and Yamagata, T. 2009. Panels with Nonstationary Multifactor Error Structures. Mimeo.

[46] Kapoor, M., Kelejian, H. and Prucha, I. 2007. Panel Data Models with Spatially Correlated Error Components. Journal of Econometrics, 140, 97–130.

[47] Kelejian, H. and Prucha, I. 2010. "Specification and Estimation of Spatial Autoregressive Models with Autoregressive and Heteroskedastic Disturbances. Journal of Econometrics, forthcoming.

[48] Kiviet, J. and Sarafidis, V. 2000. Cross-sectional Correlation in Panel Data Relationships. Mimeo.

[49] Kontoghiorghes, E. J. and Clarke, M. R. B. 1995. An alternative approach for the numerical solution of seemingly unrelated regression equations models. Computational Statistics & Data Analysis, 19(4), 369-377.

[50] Lee, L. F. 2004. Asymptotic Distributions of Quasi-Maximum Likelihood Estimators for Spatial Autoregressive Models. Econometrica, 72, 1899–1925.

[51] Lee, L. F. 2007. GMM and 2SLS estimation of mixed regressive, spatial autoregressive models. Journal of Econometrics, 137, 489–514.

[52] Lawley, D.N. and Maxwell A.E. 1971. Factor Analysis as a Statistical Method. Butterworth, London.

[53] Moon, R. G. and Perron, B. 2004. Efficient Estimation of the SUR Cointegrating Regression Model and Testing for Purchasing Power Parity. Econometric Reviews, 23, 293-323.

[54] Moon, H. R. and Perron, B. 2006. Seemingly Unrelated Regressions. Mimeo.

[55] Mundlak, Y. 1978. On the pooling of time series and cross section data. Econometrica, 46, 69-85.

[56] Nauges, C. and Thomas, A. 2003. Consistent estimation of dynamic panel data models with time-varying individual effects. Annales d'Economie et de Statistique, 70, 53-74.

[57] Neprash, J.A. 1934. Some Problems in the Correlation of Spatially Distributed Variables. Journal of the American Statistical Association, 29, 167-168.

[58] Newey, W. and West, K. 1987. A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. Econometrica, 55(3), 703-708.

[59] Nickell, S. 1981. Biases in Dynamic Models with Fixed Effects. Econometrica, 49, 1417-1426.

[60] Onatski, A. 2007. A formal statistical test for the number of factors in the approximate factor models. Mimeo

[61] Pesaran, M. H. 2004. General diagnostic tests for cross section dependence in panels. University of Cambridge, Faculty of Economics, Cambridge Working Papers in Economics No. 0435.

[62] Pesaran, M. H. and Tosetti, E. 2009. Large panels with common factors and spatial correlations. Mimeo.

[63] Pesaran, M. H., A. Ullah, and Yamagata, T. 2008. A bias-adjusted test of error cross section dependence. The Econometrics Journal, 11, 105-127.

[64] Phillips, P. and Sul, D. 2003. Dynamic Panel Estimation and Homogeneity Testing under cross-sectional Dependence. Econometrics Journal 6, 217-259.

[65] Phillips, P. and Sul, D. 2007. Bias in Dynamic Panel Estimation with Fixed Effects, Incidental Trends and cross-sectional Dependence. Journal of Econometrics 137, 162-188.

[66] Robertson, D. and Symons. J. 2007. Maximum Likelihood Factor Analysis with Rank Deficient Sample Covariance Matrices. Journal of Multivariate Analysis, 98(4), 813-828.

[67] Robertson, D., V. Sarafidis, and J. Symons (2010). IV Estimation of Panels with Factor Residuals. mimeo.

[68] Sarafidis, V. 2009. GMM Estimation of Short Dynamic Panel Data Models with Error Cross-sectional Dependence. Mimeo.

[69] Sarafidis, V. and Robertson, D. 2009. On the Impact of Error Cross-sectional Dependence in Short Dynamic Panel Estimation. The Econometrics Journal, 12(1), 62-81.

[70] Sarafidis, V., Yamagata, T. and Robertson, D. 2009. A Test of Cross Section Dependence for a Linear Dynamic Panel Model with Regressors. Journal of Econometrics, 148(2), 149-161.

[71] Sargan, J.D. 1958. The Estimation of Economic Relationships Using Instrumental Variables. Econometrica, 26, 393-495.

[72] Stephan, F.F. 1934. Sampling Errors and Interpretations of Social Data Ordered in Time and Space. Journal of the American Statistical Association, 29, 165-166.

[73] Srivastava, V. K. and Dwivedi, T. D. 1979. Estimation of seemingly unrelated regression equations – a brief survey. Journal of Econometrics, 10, 15-32.

[74] Srivastava. S. and Giles, D. 1987. Seemingly Unrelated Regression Equations Models. Marcel Dekker, New York.

[75] Tobler, W. 1970. A Computer Movie Simulating Urban Growth in the Detroit Region. Economic Geography, 46, 234-240.

[76] Yamagata, T. 2008. A Joint Serial Correlation Test for Linear Panel Data Models. Journal of Econometrics 146, 13-145.

[77] Wansbeek, T., and Knaap, T. 1999. Estimating a Dynamic Panel Data Model with Heterogenous Trends. Annales d'Economie et de Statistique, 55-56, 331-349.

[78] Wansbeek, T., and E. Meijer. 2000. Measurement Error and Latent Variables in Econometrics. Amsterdam, Elsevier.

[79] Wansbeek, T., and E. Meijer. 2007. Comments on; Panel data Analysis - Advantages and Challenges. TEST. Vol. 16, pp. 33-36.

[80] Zellner, A. 1962. An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias. Journal of the American Statistical Association, 57, 348-368.