Large Panel Data Models with Cross-Sectional Dependence: A Survey*

Alexander Chudik†

Federal Reserve Bank of Dallas and CAFE

M. Hashem Pesaran‡

University of Southern California, CAFE, USA, and Trinity College, Cambridge, UK

9 August 2013

Abstract

This paper provides an overview of the recent literature on estimation and inference in large panel data models with cross-sectional dependence. It reviews panel data models with strictly exogenous regressors as well as dynamic models with weakly exogenous regressors. The paper begins with a review of the concepts of weak and strong cross-sectional dependence, and discusses the exponent of cross-sectional dependence that characterizes the different degrees of cross-sectional dependence. It considers a number of alternative estimators for static and dynamic panel data models, distinguishing between factor and spatial models of cross-sectional dependence. The paper also provides an overview of tests of independence and weak cross-sectional dependence.

Keywords: Large panels, weak and strong cross-sectional dependence, factor structure, spatial dependence, tests of cross-sectional dependence.
JEL Classification: C31, C33.

*We are grateful to an anonymous referee, Cheng Chou, Ron Smith, Vanessa Smith, Wei Xie, Takashi Yamagata and Qiankun Zhou for helpful comments. Pesaran acknowledges financial support from ESRC grant no. ES/I031626/1.

†Federal Reserve Bank of Dallas, 2200 N. Pearl Street, Dallas, Texas. E-mail: alexander.chudik@dal.frb.org. The views expressed in this paper are those of the authors and do not necessarily reflect those of the Federal Reserve Bank of Dallas or the Federal Reserve System.

‡Department of Economics, University of Southern California, 3620 South Vermont Ave, Los Angeles, California 90089, USA. Email: pesaran@usc.edu; http://www.econ.cam.ac.uk/faculty/pesaran/

1 Introduction

This paper reviews econometric methods for large linear panel data models subject to error cross-sectional dependence. Early panel data literature assumed cross-sectionally independent errors and homogeneous slopes. Heterogeneity across units was confined to unit-specific intercepts, treated as fixed or random (see, e.g. the survey by Chamberlain (1984)). Dependence of errors was only considered in spatial models, but not in standard panels. However, with an increasing availability of data (across countries, regions, or industries), the panel literature moved from predominantly micro panels, where the cross dimension (N) is large and the time series dimension (T) is small, to models with both N and T large, and it has been recognized that, even after conditioning on unit-specific regressors, individual units, in general, need not be cross-sectionally independent.

Ignoring cross-sectional dependence of errors can have serious consequences, and the presence of some form of cross-sectional correlation of errors in panel data applications in economics is likely to be the rule rather than the exception. Cross correlations of errors could be due to omitted common effects, spatial effects, or could arise as a result of interactions within socioeconomic networks. Conventional panel estimators such as fixed or random effects can result in misleading inference and even inconsistent estimators, depending on the extent of cross-sectional dependence and on whether the source generating the cross-sectional dependence (such as an unobserved common shock) is correlated with regressors (Phillips and Sul (2003); Andrews (2005); Phillips and Sul (2007); Sarafidis and Robertson (2009)). Correlation across units in panels may also have serious drawbacks on commonly used panel unit root tests, since several of the existing tests assume independence. As a result, when applied to cross-sectionally dependent panels, such unit root tests can have substantial size distortions (O'Connell (1998)). If, however, the extent of cross-sectional dependence of errors is sufficiently weak, or limited to a sufficiently small number of cross-sectional units, then its consequences on conventional estimators will be negligible. Furthermore, the consistency of conventional estimators can be affected only when the source of cross-sectional dependence is correlated with regressors. The problem of testing for the extent of cross-sectional correlation of panel residuals and modelling the cross-sectional dependence of errors are therefore important issues.

In the case of panel data models where the cross section dimension is short and the time series dimension is long, the standard approach to cross-sectional dependence is to consider the equations from different cross-sectional units as a system of seemingly unrelated regression equations (SURE), and then estimate it by Generalized Least Squares techniques (see Zellner (1962)). This approach assumes that the source generating cross-sectional dependence is not correlated with regressors and this assumption is required for the consistency of the SURE estimator. If the time series dimension is not sufficiently large, and in particular if $N > T$, the SURE approach is not feasible.

Currently, there are two main strands in the literature for dealing with error cross-sectional dependence in panels where N is large, namely the spatial econometric and the residual multifactor approaches. The spatial econometric approach assumes that the structure of cross-sectional correlation is related to location and distance among units, defined according to a pre-specified metric given by a 'connection or spatial' matrix that characterizes the pattern of spatial dependence according to pre-specified rules. Hence, cross-sectional correlation is represented by means of a spatial process, which explicitly relates each unit to its neighbors (see Whittle (1954), Moran (1948), Cliff and Ord (1973 and 1981), Anselin (1988 and 2001), Haining (2003, Chapter 7), and the recent survey by Lee and Yu (2013)). This approach, however, typically does not allow for slope heterogeneity across the units and requires a priori knowledge of the weight matrix.

The residual multifactor approach assumes that the cross dependence can be characterized by a small number of unobserved common factors, possibly due to economy-wide shocks that affect all units albeit with different intensities. Geweke (1977) and Sargent and Sims (1977) introduced dynamic factor models, which have more recently been generalized to allow for weak cross-sectional dependence by Forni and Lippi (2001), Forni et al. (2000) and Forni et al. (2004). This approach does not require any prior knowledge regarding the ordering of individual cross section units.

The main focus of this paper is on estimation and inference in the case of large N and T panel data models with a common factor error structure. We provide a synthesis of the alternative approaches proposed in the literature (such principal components and common correlated effects approaches), with particular focus on key assumptions and their consequences from the practitioners' view point. In particular, we discuss robustness of estimators to cross-sectional dependence of errors, consequences of coefficient heterogeneity, discuss panels with strictly or weakly exogenous regressors, including panels with a lagged dependent variable, and highlight how to test for residual cross-sectional dependence.

The outline of the paper is as follows: an overview of the different types of cross-sectional dependence is provided in Section 2. The analysis of cross-sectional dependence using a factor error structure is presented in Section 3. A review of estimation and inference in the case of large panels with a multifactor error structure and strictly exogenous regressors is provided in Section 4, and its extension to models with lagged dependent variables and/or weakly exogenous regressors is given in Section 5. A review of tests of error cross-sectional dependence in static and dynamics panels is presented in Section 6. Section 7 discusses application of common correlated effects estimators and tests of error cross-sectional dependence to unbalanced panels, and the final section concludes.

2 Types of Cross-Sectional Dependence

A better understanding of the extent and nature of cross-sectional dependence of errors is an important issue in the analysis of large panels. This section introduces the notions of weak and strong cross-sectional dependence and the notion of exponent of cross-sectional dependence to characterize the correlation structure of $\{z_{it}\}$ over the cross-sectional dimension, i, at a given point in time, t. Consider the double index process $\{z_{it}, i \in \mathbb{N}, t \in \mathbb{Z}\}$, where z_{it} is defined on a suitable probability space, the index t refers to an ordered set such as time, and i refers to units of an unordered population. We make the following assumption:

ASSUMPTION CSD.1: For each $t \in T \subseteq Z$, $\mathbf{z}_t = (z_{1t}, ..., z_{Nt})'$ has mean $E(\mathbf{z}_t) = 0$, and variance $Var(\mathbf{z}_t) = \boldsymbol{\Sigma}_t$, where $\boldsymbol{\Sigma}_t$ is an $N \times N$ symmetric, nonnegative definite matrix. The (i, j)-th element of $\boldsymbol{\Sigma}_t$, denoted by $\sigma_{ij,t}$, is bounded such that $0 < \sigma_{ii,t} \leq K$, for $i = 1, 2, ..., N$, where K is a finite constant independent of N.

Instead of assuming unconditional mean and variances, one could consider conditioning on a given information set, Ω_{t-1}, for $t = 1, 2, ..., T$, as done in Chudik et al. (2011). The assumption of zero means can also be relaxed to $E(\mathbf{z}_t) = \boldsymbol{\mu}$ (or $E(\mathbf{z}_t | \Omega_{t-1}) = \boldsymbol{\mu}_{t-1}$). The covariance matrix, $\boldsymbol{\Sigma}_t$, fully characterizes cross-sectional correlations of the double index process $\{z_{it}\}$, and this section discusses summary measures based on the elements of $\boldsymbol{\Sigma}_t$ that can be used to characterize the extent of the cross-sectional dependence in $\mathbf{z}_t$.

Summary measures of cross-sectional dependence based on $\boldsymbol{\Sigma}_t$ can be constructed in a number of different ways. One possible measure, that has received a great deal of attention in the literature, is the largest eigenvalue of $\mathbf{\Sigma}_t$, denoted by $\lambda_1(\mathbf{\Sigma}_t)$. See, for example, Bai and Silverstein (1998), Hachem et al. (2005) and Yin et al. (1988). However, the existing work in this area suggests that the estimates of $\lambda_1(\mathbf{\Sigma}_t)$ based on sample estimates of $\mathbf{\Sigma}_t$ could be very poor when N is large relative to T, and consequently using estimates of $\lambda_1(\mathbf{\Sigma}_t)$ for the analysis of cross-sectional dependence might be problematic in cases where T is not sufficiently large relative to N. Accordingly, other measures based on matrix norms of $\mathbf{\Sigma}_t$ have also been used in the literature. One prominent choice is the absolute column sum matrix norm, defined by $\|\mathbf{\Sigma}_t\|_1 = \max_{j\in\{1,2,\dots,N\}} \sum_{i=1}^N |\sigma_{ij,t}|$, which is equal to the absolute row sum matrix norm of $\mathbf{\Sigma}_t$, defined by $\|\mathbf{\Sigma}_t\|_\infty = \max_{i\in\{1,2,\dots,N\}} \sum_{j=1}^N |\sigma_{ij,t}|$, due to the symmetry of $\mathbf{\Sigma}_t$. It is easily seen that $|\lambda_1(\mathbf{\Sigma}_t)| \leq \sqrt{\|\mathbf{\Sigma}_t\|_1 \|\mathbf{\Sigma}_t\|_\infty} = \|\mathbf{\Sigma}_t\|_1$. See Chudik et al. (2011). Another possible measure of cross-sectional dependence can be based on the behavior of (weighted) cross-sectional averages which is often of interest in panel data econometrics, as well as in macroeconomics and finance where the object of the analysis is often the study of aggregates or portfolios of asset returns. In view of this, Bailey et al. (2012) and Chudik et al. (2011) suggest to summarize the extent of cross-sectional dependence based on the behavior of cross-sectional averages $\bar{z}_{wt} = \sum_{i=1}^N w_{it} z_{it} = \mathbf{w}_t' \mathbf{z}_t$, at a point in time t, for $t \in \mathcal{T}$, where $\mathbf{z}_t$ satisfies Assumption CSD.1 and the sequence of weight vectors $\mathbf{w}_t$ satisfies the following assumption.

ASSUMPTION CSD.2: Let $\mathbf{w}_t = (w_{1t}, \dots, w_{Nt})'$, for $t \in \mathcal{T} \subseteq \mathbb{Z}$ and $N \in \mathbb{N}$, be a vector of non-stochastic weights. For any $t \in \mathcal{T}$, the sequence of weight vectors $\{\mathbf{w}_t\}$ of growing dimension $(N \to \infty)$ satisfies the 'granularity' conditions:

$$\|\mathbf{w}_t\| = \sqrt{\mathbf{w}_t' \mathbf{w}_t} = O\left(N^{-\frac{1}{2}}\right), \tag{1}$$

$$\frac{w_{jt}}{\|\mathbf{w}_t\|} = O\left(N^{-\frac{1}{2}}\right) \quad \text{uniformly in } j \in \mathbb{N}. \tag{2}$$

Assumption CSD.2, known in finance as the granularity condition, ensures that the weights $\{w_{it}\}$ are not dominated by a few of the cross section units.[1] Although we have assumed the weights to be non-stochastic, this is done for expositional convenience and can be relaxed by allowing the weights, w_t, to be random but distributed independently of z_t. Chudik et al. (2011) define the concepts of weak and strong cross-sectional dependence based on the limiting behavior of $\bar{z}_{wt}$ at a

[1] Conditions (1)-(2) imply existence of a finite constant K (which does not depend on i or N) such that $|w_{it}| < KN^{-1}$ for any $i = 1, 2, \dots, N$ and any $N \in \mathbb{N}$.

given point in time $t \in \mathcal{T}$, as $N \to \infty$.

Definition 1 (Weak and strong cross-sectional dependence) *The process $\{z_{it}\}$ is said to be cross-sectionally weakly dependent (CWD) at a given point in time $t \in \mathcal{T}$, if for* any *sequence of weight vectors $\{\mathbf{w}_t\}$ satisfying the granularity conditions (1)-(2) we have*

$$\lim_{N \to \infty} Var(\mathbf{w}_t' \mathbf{z}_t) = 0. \tag{3}$$

$\{z_{it}\}$ is said to be cross-sectionally strongly dependent (CSD) at a given point in time $t \in \mathcal{T}$, if there exists a sequence of weight vectors $\{\mathbf{w}_t\}$ satisfying (1)-(2) and a constant K independent of N such that for any N sufficiently large (and as $N \to \infty$)

$$Var(\mathbf{w}_t' \mathbf{z}_t) \geq K > 0. \tag{4}$$

The above concepts can also be defined conditional on a given information set, Ω_{t-1}, see Chudik et al. (2011). The choice of the conditioning set largely depends on the nature of the underlying processes and the purpose of the analysis. For example, in the case of dynamic stationary models, the information set could contain all lagged realizations of the process $\{z_{it}\}$, that is $\Omega_{t-1} = \{\mathbf{z}_{t-1}, \mathbf{z}_{t-2}, ...\}$, whilst for dynamic non-stationary models, such as unit root processes, the information included in Ω_{t-1}, could start from a finite past. Conditioning information set could also contain contemporaneous realizations, which might be useful in applications where a particular unit has a dominant influence on the rest of the units in the system. For further details, see Chudik and Pesaran (2013c).

The following proposition establishes the relationship between weak cross-sectional dependence and the asymptotic behavior of the largest eigenvalue of $\mathbf{\Sigma}_t$.

Proposition 1 *The following statements hold:*

(i) The process $\{z_{it}\}$ is CWD at a point in time $t \in \mathcal{T}$, if $\lambda_1(\mathbf{\Sigma}_t)$ is bounded in N or increases at the rate slower than N.

(ii) The process $\{z_{it}\}$ is CSD at a point in time $t \in \mathcal{T}$, if and only if for any N sufficiently large (and as $N \to \infty$), $N^{-1}\lambda_1(\mathbf{\Sigma}_t) \geq K > 0$.

Proof. First, suppose $\lambda_1(\mathbf{\Sigma}_t)$ is bounded in N or increases at the rate slower than N. We have

$$Var(\mathbf{w}_t'\mathbf{z}_t) = \mathbf{w}_t'\mathbf{\Sigma}_t\mathbf{w}_t \leq \left(\mathbf{w}_t'\mathbf{w}_t\right)\lambda_1(\mathbf{\Sigma}_t), \tag{5}$$

and under the granularity conditions (1)-(2) it follows that

$$\lim_{N\to\infty} Var(\mathbf{w}_t'\mathbf{z}_t) = 0,$$

namely that $\{z_{it}\}$ is CWD, which proves (i). Proof of (ii) is provided in Chudik, Pesaran, and Tosetti (2011) ■

It is often of interest to know not only whether $\bar{z}_{wt}$ converges to its mean, but also the rate at which this convergence (if at all) takes place. To this end, Bailey et al. (2012) propose to characterize the degree of cross-sectional dependence by an exponent of cross-sectional dependence defined by the rate of change of $Var(\bar{z}_{wt})$ in terms of N. Note that in the case where z_{it} are independently distributed across i, we have $Var(\bar{z}_{wt}) = O\left(N^{-1}\right)$, whereas in the case of strong cross-sectional dependence $Var(\bar{z}_{wt}) \geq K > 0$. There is, however, a range of possibilities in between, where $Var(\bar{z}_{wt})$ decays but at a rate slower than N^{-1}. In particular, using a factor framework, Bailey et al. (2012) show that in general

$$Var(\bar{z}_{wt}) = \kappa_0 N^{2(\alpha-1)} + \kappa_1 N^{-1} + O(N^{\alpha-2}), \tag{6}$$

where $\kappa_i > 0$ for $i = 0$ and 1, are bounded in N, which will be time invariant in the case of stationary processes. Since the rate at which $Var(\bar{z}_{wt})$ tends to zero with N cannot be faster than N^{-1}, the range of α identified by $Var(\bar{z}_{wt})$ lies in the restricted interval $-1 < 2\alpha - 2 \leq 0$ or $1/2 < \alpha \leq 1$. Note that (3) holds for all values of $\alpha < 1$, whereas (4) holds only for $\alpha = 1$. Hence the process with $\alpha < 1$ is CWD, and a CSD process has the exponent $\alpha = 1$. Bailey et al. (2012) show that under certain conditions on the underlying factor model, α is identified in the range $1/2 < \alpha \leq 1$, and can be consistently estimated. Alternative bias-adjusted estimators of α are proposed and shown by Monte Carlo experiments to have satisfactory small sample properties.

A particular form of a CWD process arises when pair-wise correlations take non-zero values only across finite subsets of units that do not spread widely as the sample size increases. A similar situation arises in the case of spatial processes, where direct dependence exists only amongst adjacent observations, and the indirect dependence is assumed to decay with distance. For further details see Pesaran and Tosetti (2011).

Since $\lambda_1(\mathbf{\Sigma}_t) \leq \|\mathbf{\Sigma}_t\|_1$, it follows from (5) that both the spectral radius and the column norm of the covariance matrix of a CSD process will be increasing at the rate N. Similar situations also arise in the case of time series processes with long memory or strong temporal dependence where autocorrelation coefficients are not absolutely summable. Along the cross section dimension, common factor models represent examples of strong cross-sectional dependence.

3 Common Factor Models

Consider the m factor model for $\{z_{it}\}$

$$z_{it} = \gamma_{i1} f_{1t} + \gamma_{i2} f_{2t} + ... + \gamma_{im} f_{mt} + e_{it}, \quad i = 1, 2, ..., N, \tag{7}$$

which can be written more compactly as

$$\mathbf{z}_t = \mathbf{\Gamma}\mathbf{f}_t + \mathbf{e}_t, \tag{8}$$

where $\mathbf{f}_t = (f_{1t}, f_{2t}, ..., f_{mt})'$, $\mathbf{e}_t = (e_{1t}, e_{2t}, ..., e_{Nt})'$, and $\mathbf{\Gamma} = (\gamma_{ij})$, for $i = 1, 2, ..., N$, $j = 1, 2, ..., m$, is an $N \times m$ matrix of fixed coefficients, known as factor loadings. The common factors, $\mathbf{f}_t$, simultaneously affect all cross-sectional units, albeit with different degrees as measured by $\boldsymbol{\gamma}_i = (\gamma_{i1}, \gamma_{i2}, ..., \gamma_{im})'$. Examples of observed common factors that tend to affect all households' and firms' consumption and investment decisions include interest rates and oil prices. Aggregate demand and supply shocks represent examples of common unobserved factors. In multifactor models, interdependence arises from reaction of units to some external events. Further, according to this representation, correlation between any pair of units does not depend on how far these observations are apart, and violates the distance decay effect that underlies the spatial interaction model.

The following assumptions are typically made regarding the common factors, $f_{\ell t}$, and the idiosyncratic errors, e_{it}.

ASSUMPTION CF.1: The $m \times 1$ vector $\mathbf{f}_t$ is a zero mean covariance stationary process, with absolutely summable autocovariances, distributed independently of $e_{it'}$ for all i, t, t', such that

$E(f_{\ell t}^2) = 1$ and $E(f_{\ell t} f_{\ell' t}) = 0$, for $\ell \neq \ell' = 1, 2, ..., m$.

ASSUMPTION CF.2: $Var\left(e_{it}\right) = \sigma_i^2 < K < \infty$, e_{it} and e_{jt} are independently distributed for all $i \neq j$ and for all t. Specifically, $\max_i \left(\sigma_i^2\right) = \sigma_{\max}^2 < K < \infty$.

Assumption CF.1 is an identification condition, since it is not possible to separately identify $\mathbf{f}_t$ and $\mathbf{\Gamma}$. The above factor model with a fixed number of factors and cross-sectionally independent idiosyncratic errors is often referred to as an exact factor model. Under the above assumptions, the covariance of $\mathbf{z}_t$ is given by

$$E\left(\mathbf{z}_t \mathbf{z}_t'\right) = \mathbf{\Gamma}\mathbf{\Gamma}' + \mathbf{V},$$

where $\mathbf{V}$ is a diagonal matrix with elements σ_i^2 on the main diagonal.

The assumption that the idiosyncratic errors, e_{it}, are cross-sectionally independent is not necessary and can be relaxed. The factor model that allows the idiosyncratic shocks, e_{it}, to be cross-sectionally weakly correlated is known as the approximate factor model. See Chamberlain (1983). In general, the correlation patterns of the idiosyncratic errors can be characterized by

$$\mathbf{e}_t = \mathbf{R}\boldsymbol{\varepsilon}_t, \tag{9}$$

where $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \varepsilon_{2t}, ..., \varepsilon_{Nt})' \sim (\mathbf{0}, \mathbf{I}_N)$. In the case of this formulation $\mathbf{V} = \mathbf{R}\mathbf{R}'$, which is no longer diagonal when $\mathbf{R}$ is not diagonal, and further identification restrictions are needed so that the factor specification can be distinguished from the cross-sectional dependence assumed for the idiosyncratic errors. To this end it is typically assumed that $\mathbf{R}$ has bounded row and column sum matrix norms (so that the cross-sectional dependence of $\mathbf{e}_t$ is sufficiently weak) and the factor loadings are such that $\lim_{N\to\infty}(N^{-1}\mathbf{\Gamma}'\mathbf{\Gamma})$ is a full rank matrix.

A leading example of $\mathbf{R}$ arises in the context of the first-order spatial autoregressive, SAR(1), model, defined by

$$\mathbf{e}_t = \rho \mathbf{W}\mathbf{e}_t + \mathbf{\Lambda}\boldsymbol{\varepsilon}_t, \tag{10}$$

where $\mathbf{\Lambda}$ is a diagonal matrix with strictly positive and bounded elements, $0 < \sigma_i < \infty$, ρ is a spatial autoregressive coefficient, and the matrix $\mathbf{W}$ is the 'connection or spatial' weight matrix which is taken as given. Assuming that $(\mathbf{I}_N - \rho\mathbf{W})$ is invertible, we then have $\mathbf{R} = (\mathbf{I}_N - \rho\mathbf{W})^{-1}\mathbf{\Lambda}$. In the spatial literature, $\mathbf{W}$ is assumed to have non-negative elements and is typically row-standardized so that $\|\mathbf{W}\|_\infty = 1$. Under these assumptions, $|\rho| < 1$ ensures that $|\rho| \|\mathbf{W}\|_\infty < 1$, and we have

$$\begin{aligned}
\|\mathbf{R}\|_\infty &= \|\mathbf{\Lambda}\|_\infty \left\|\mathbf{I}_N + \rho\mathbf{W}+\rho^2\mathbf{W}^2 + ...\right\|_\infty \\
&\leq \|\mathbf{\Lambda}\|_\infty \left[1 + |\rho| \|\mathbf{W}\|_\infty + |\rho|^2 \|\mathbf{W}\|_\infty^2 + ...\right] = \frac{\|\mathbf{\Lambda}\|_\infty}{1 - |\rho| \|\mathbf{W}\|_\infty} < K < \infty,
\end{aligned}$$

where $\|\mathbf{\Lambda}\|_\infty = \max_i(\sigma_i) < \infty$. Similarly, $\|\mathbf{R}\|_1 < K < \infty$, if it is further assumed that $|\rho| \|\mathbf{W}\|_1 < 1$. In general, $\mathbf{R} = (\mathbf{I}_N - \rho\mathbf{W})^{-1} \mathbf{\Lambda}$ has bounded row and column sum matrix norms if $|\rho| < \min(1/\|\mathbf{W}\|_1, 1/\|\mathbf{W}\|_\infty)$. In the case where $\mathbf{W}$ is a row and column stochastic matrix (often assumed in the spatial literature) this sufficient condition reduces to $|\rho| < 1$, which also ensures the invertibility of $(\mathbf{I}_N - \rho\mathbf{W})$. Note that for a doubly stochastic matrix $\rho(\mathbf{W}) = \|\mathbf{W}\|_1 = \|\mathbf{W}\|_\infty = 1$, where $\rho(\mathbf{W})$ is the spectral radius of $\mathbf{W}$. It turns out that almost all spatial models analyzed in the spatial econometrics literature characterize weak forms of cross-sectional dependence. See Sarafidis and Wansbeek (2012) for further discussion.

Turning now to the factor representation, to ensure that the factor component of (8) represents strong cross-sectional dependence (so that it can be distinguished from the idiosyncratic errors) it is sufficient that the absolute column sum matrix norm of $\|\mathbf{\Gamma}\|_1 = \max_{j\in\{1,2,...,N\}} \sum_{i=1}^N |\gamma_{ij}|$ rises with N at the rate N, which implies that $\lim_{N\to\infty}(N^{-1}\mathbf{\Gamma}'\mathbf{\Gamma})$ is a full rank matrix, as required earlier.

The distinction between weak and strong cross-sectional dependence in terms of factor loadings are formalized in the following definition.

Definition 2 (Strong and weak factors) *The factor $f_{\ell t}$ is said to be strong if*

$$\lim_{N\to\infty} N^{-1}\sum_{i=1}^{N} |\gamma_{i\ell}| = K > 0. \tag{11}$$

The factor $f_{\ell t}$ is said to be weak if

$$\lim_{N\to\infty} \sum_{i=1}^{N} |\gamma_{i\ell}| = K < \infty. \tag{12}$$

It is also possible to consider intermediate cases of semi-weak or semi-strong factors. In general, let α_ℓ be a positive constant in the range $0 \leq \alpha_\ell \leq 1$ and consider the condition

$$\lim_{N\to\infty} N^{-\alpha_\ell} \sum_{i=1}^{N} |\gamma_{i\ell}| = K < \infty. \tag{13}$$

Strong and weak factors correspond to the two values of $\alpha_\ell = 1$ and $\alpha_\ell = 0$, respectively. For any other values of $\alpha_\ell \in (0,1)$ the factor $f_{\ell t}$ can be said to be semi-strong or semi-weak. It will prove useful to associate the semi-weak factors with values of $0 < \alpha_\ell < 1/2$, and the semi-strong factors with values of $1/2 \leq \alpha_\ell < 1$. In a multi-factor set up the overall exponent can be defined by $\alpha = \max(\alpha_1, \alpha_2, ..., \alpha_m)$.

Example 1 *Suppose that z_{it} are generated according to the simple factor model, $z_{it} = \gamma_i f_t + e_{it}$, where f_t is independently distributed of γ_i, and $e_{it} \sim IID(0, \sigma_i^2)$, for all i and t, σ_i^2 is non-stochastic for expositional simplicity and bounded, $E\left(f_t^2\right) = \sigma_f^2 < \infty$, $E(f_t) = 0$ and f_t is independently distributed of $e_{it'}$ for all i, t and t'. The factor loadings are given by*

$$\begin{aligned} \gamma_i &= \mu + v_i, \textit{ for } i = 1, 2, ..., [N^{\alpha_\gamma}] & (14)\\ \gamma_i &= 0 \textit{ for } i = [N^{\alpha_\gamma}] + 1, [N^{\alpha_\gamma}] + 2, ..., N, & (15) \end{aligned}$$

for some constant $\alpha_\gamma \in [0,1]$, where $[N^{\alpha_\gamma}]$ is the integer part of N^{α_γ}, $\mu \neq 0$, and v_i are IID with mean 0 and the finite variance, σ_v^2.[2] *Note that $\sum_{i=1}^N |\gamma_i| = O_p\left([N^{\alpha_\gamma}]\right)$ and the factor f_t with loadings γ_i is strong for $\alpha_\gamma = 1$, weak for $\alpha_\gamma = 0$ and semi-weak or semi-strong for $0 < \alpha_\gamma < 1$. Consider the variance of the (simple) cross-sectional averages $\bar{z}_t = N^{-1} \sum_{i=1}^N z_{it}$*

$$Var_N(\bar{z}_t) = Var\left(\bar{z}_t \left| \{\gamma_i\}_{i=1}^N \right.\right) = \bar{\gamma}_N^2 \sigma_f^2 + N^{-1} \bar{\sigma}_N^2, \tag{16}$$

where (dropping the integer part sign, $[.]$, for further clarity)

$$\begin{aligned} \bar{\gamma}_N &= N^{-1} \sum_{i=1}^{N} \gamma_i = N^{-1} \sum_{i=1}^{N^{\alpha_\gamma}} \gamma_i = \mu N^{\alpha_\gamma - 1} + N^{\alpha_\gamma - 1} \left(\frac{1}{N^{\alpha_\gamma}} \sum_{i=1}^{N^{\alpha_\gamma}} v_i \right)\\ \bar{\sigma}_N^2 &= N^{-1} \sum_{i=1}^{N} \sigma_i^2 > 0. \end{aligned}$$

[2] *The assumption of zero loadings for $i > [N^{\alpha_\gamma}]$ could be relaxed so long as $\sum_{i=[N^{\alpha_\gamma}]+1}^{N} |\gamma_i| = O_p(1)$. But for expositional simplicity we maintain $\gamma_i = 0$ for $i = [N^{\alpha_\gamma}] + 1, [N^{\alpha_\gamma}] + 2, ..., N$.*

But, noting that

$$E\left(\bar{\gamma}_N\right) = \mu N^{\alpha_\gamma - 1},\ Var(\bar{\gamma}_N) = N^{\alpha_\gamma - 2}\sigma_v^2,$$

we have

$$E\left(\bar{\gamma}_N^2\right) = \left[E\left(\bar{\gamma}_N\right)\right]^2 + Var(\bar{\gamma}_N) = \mu^2 N^{2(\alpha_\gamma - 1)} + N^{\alpha_\gamma - 2}\sigma_v^2.$$

Therefore, using this result in (16), we now have

$$\begin{aligned} Var\left(\bar{z}_t\right) &= E\left[Var_N\left(\bar{z}_t\right)\right] = \sigma_f^2\mu^2 N^{2(\alpha_\gamma - 1)} + \bar{\sigma}_N^2 N^{-1} + \sigma_v^2\sigma_f^2 N^{\alpha_\gamma - 2} \qquad (17) \\ &= \sigma_f^2\mu^2 N^{2(\alpha_\gamma - 1)} + \bar{\sigma}_N^2 N^{-1} + O\left(N^{\alpha_\gamma - 2}\right). \qquad (18) \end{aligned}$$

Thus the exponent of cross-sectional dependence of z_{it}, denoted as α_z, and the exponent α_γ coincide in this example, so long as $\alpha_\gamma > 1/2$. When $\alpha_\gamma = 1/2$, one can not use $Var\left(\bar{z}_t\right)$ to distinguish the factor effects from those of the idiosyncratic terms. Of course, this does not necessarily mean that other more powerful techniques can not be found to distinguish such weak factor effects from the effects of the idiosyncratic terms. Finally, note also that in this example $\sum_{i=1}^N \gamma_i^2 = O_p\left(N^{\alpha_\gamma}\right)$, and the largest eigenvalue of the $N \times N$ covariance matrix, $Var\left(\mathbf{z}_t\right)$, also rises at the rate of N^{α_γ}.

The relationship between the notions of CSD and CWD and the definitions of weak and strong factors are explored in the following theorem.

Theorem 2 *Consider the factor model (8) and suppose that Assumptions CF.1-CF.2 hold, and there exists a positive constant $\alpha = \max(\alpha_1, \alpha_2, ..., \alpha_m)$ in the range $0 \leq \alpha \leq 1$, such that condition (13) is met for any $\ell = 1, 2, .., m$. Then the following statements hold:*

(i) The process $\{z_{it}\}$ is cross-sectionally weakly dependent at a given point in time $t \in \mathcal{T}$ if $\alpha < 1$, which includes cases of weak, semi-weak or semi-strong factors, $f_{\ell t}$, for $\ell = 1, 2, ..., m$.

(ii) The process $\{z_{it}\}$ is cross-sectionally strongly dependent at a given point in time $t \in \mathcal{T}$ if and only if there exists at least one strong factor.

Proof is provided in Chudik, Pesaran, and Tosetti (2011).

Since a factor structure can lead to strong as well as weak forms of cross-sectional dependence, cross-sectional dependence can also be characterized more generally by the following N factor representation

$$z_{it} = \sum_{j=1}^{N} \gamma_{ij} f_{jt} + \varepsilon_{it}, \text{ for } i = 1, 2, ..., N,$$

where ε_{it} is independently distributed across i. Under this formulation, to ensure that the variance of z_{it} is bounded in N, we also require that

$$\sum_{\ell=1}^{N} |\gamma_{i\ell}| \leq K < \infty, \text{ for } i = 1, 2, ..., N. \tag{19}$$

z_{it} can now be decomposed as

$$z_{it} = z_{it}^{s} + z_{it}^{w}, \tag{20}$$

where

$$z_{it}^{s} = \sum_{\ell=1}^{m} \gamma_{i\ell} f_{\ell t}; \quad z_{it}^{w} = \sum_{\ell=m+1}^{N} \gamma_{i\ell} f_{\ell t} + \varepsilon_{it}, \tag{21}$$

and $\gamma_{i\ell}$ satisfy conditions (11) for $\ell = 1, ..., m$, where m must be finite in view of the absolute summability condition (19) that ensures finite variances. Remaining loadings $\gamma_{i\ell}$ for $\ell = m+1, m+2, ..., N$ must satisfy either (12) or (13) for some $\alpha < 1$.[3] In the light of Theorem 2, it can be shown that z_{it}^{s} is CSD and z_{it}^{w} is CWD. Also, notice that when z_{it} is CWD, we have a model with no strong factors and potentially an infinite number of weak or semi-strong factors. Seen from this perspective, spatial models considered in the literature can be viewed as an N weak factor model.

Consistent estimation of factor models with weak or semi-strong factors may be problematic, as evident from the following example.

Example 3 *Consider the single factor model with known factor loadings*

$$z_{it} = \gamma_i f_t + \varepsilon_{it}, \quad \varepsilon_{it} \sim IID\left(0, \sigma^2\right).$$

The least squares estimator of f_t, which is the best linear unbiased estimator, is given by

$$\hat{f}_t = \frac{\sum_{i=1}^{N} \gamma_i z_{it}}{\sum_{i=1}^{N} \gamma_i^2}, \quad Var\left(\hat{f}_t\right) = \frac{\sigma^2}{\sum_{i=1}^{N} \gamma_i^2}.$$

In the weak factor case where $\sum_{i=1}^{N} \gamma_i^2$ is bounded in N, then $Var\left(\hat{f}_t\right)$ does not vanish as $N \to \infty$,

[3] Note that the number of factors with $\alpha_\ell > 0$ is limited by the absolute summability condition (19).

and $\hat{f}_t$ need not be a consistent estimator of f_t. See also Onatski (2012).

Presence of weak or semi-strong factors in errors does not affect consistency of conventional panel data estimators, but affects inference, as is evident from the following example.

Example 4 *Consider the following panel data model*

$$y_{it} = \beta x_{it} + u_{it},\ u_{it} = \gamma_i f_t + \varepsilon_{it},$$

where

$$x_{it} = \delta_i f_t + v_{it}.$$

To simplify the exposition we assume that, ε_{it}, v_{js} and $f_{t'}$ are independently, and identically distributed across all i, j, t, s and t', as $\varepsilon_{it} \sim IID\left(0, \sigma_\varepsilon^2\right)$, $v_{it} \sim IID\left(0, \sigma_v^2\right)$, and $f_t \sim IID\left(0, 1\right)$. The pooled estimator of β satisfies

$$\sqrt{NT}\left(\hat{\beta}_P - \beta\right) = \frac{\frac{1}{\sqrt{NT}}\sum_{i=1}^N \sum_{t=1}^T x_{it} u_{it}}{\frac{1}{NT}\sum_{i=1}^N \sum_{t=1}^T x_{it}^2}, \tag{22}$$

where the denominator converges in probability to $\sigma_v^2 + \lim_{N\to\infty} N^{-1} \sum_{i=1}^N \delta_i^2 > 0$, while the numerator can be expressed, after substituting for x_{it} and u_{it}, as

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^N \sum_{t=1}^T x_{it} u_{it} = \frac{1}{\sqrt{NT}}\sum_{i=1}^N \sum_{t=1}^T \gamma_i \delta_i f_t^2 + \frac{1}{\sqrt{NT}}\sum_{i=1}^N \sum_{t=1}^T \left(\delta_i f_t \varepsilon_{it} + \gamma_i v_{it} f_t + v_{it}\varepsilon_{it}\right). \tag{23}$$

Under the above assumptions it is now easily seen that the second term in the above expression is $O_p(1)$, but the first term can be written as

$$\begin{aligned} \frac{1}{\sqrt{NT}}\sum_{i=1}^N \sum_{t=1}^T \gamma_i \delta_i f_t^2 &= \frac{1}{\sqrt{N}}\sum_{i=1}^N \gamma_i \delta_i \cdot \frac{1}{\sqrt{T}}\sum_{t=1}^T f_t^2 \\ &= \frac{1}{\sqrt{N}}\sum_{i=1}^N \gamma_i \delta_i \cdot O_p\left(T^{1/2}\right). \end{aligned}$$

Suppose now that f_t is a factor such that loadings γ_i and δ_i are given by (14)-(15) with the exponents α_γ and α_δ ($0 \leq \alpha_\gamma, \alpha_\delta \leq 1$), respectively, and let $\alpha = \min\left(\alpha_\gamma, \alpha_\delta\right)$. It then follows that $\sum_{i=1}^N \gamma_i \delta_i =$

$O_p\left(N^{\alpha}\right)$, *and*

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\gamma_i\delta_i f_t^2 = O_p(N^{\alpha-1/2}T^{1/2}).$$

Therefore, even if $\alpha < 1$ *the first term in (23) diverges, and overall we have* $\hat{\beta}_P - \beta = O_p(N^{\alpha-1}) + O_p(T^{-1/2}N^{-1/2})$. *It is now clear that even if* f_t *is not a strong factor, the rate of convergence of* $\hat{\beta}_P$ *and its asymptotic variance will still be affected by the factor structure of the error term. In the case where* $\alpha = 0$, *and the errors are spatially dependent, the variance matrix of the pooled estimator also depends on the nature of the spatial dependence which must be taken into account when carrying out inference on* β. *See Pesaran and Tosetti (2011) for further results and discussions.*

Weak, strong and semi-strong common factors may be used to represent very general forms of cross-sectional dependence. For example, a factor process with an infinite number of weak factors, and no idiosyncratic errors can be used to represent spatial processes. In particular, the spatial model (9) can be represented by $e_{it} = \sum_{j=1}^{N}\gamma_{ij}f_{jt}$, where $\gamma_{ij} = r_{ij}$ and $f_{jt} = \varepsilon_{jt}$. Strong factors can be used to represent the effect of the cross section units that are "dominant" or pervasive, in the sense that they impact all the other units in the sample and their effect does not vanish as N tends to infinity, (Chudik and Pesaran (2013c)). As argued in Holly, Pesaran, and Yagamata (2011), a large city may play a dominant role in determining house prices nationally. Semi-strong factors may exist if there is a cross section unit or an unobserved common factor that affects only a subset of the units and the number of affected units rise more slowly than the total number of units. Estimates of the exponent of cross-sectional dependence reported by Bailey, Kapetanios, and Pesaran (2012, Tables 1 and 2) suggest that for typical large macroeconomic data sets the estimates of α fall in the range of $0.77 - 0.92$, which fall short of 1 assumed in the factor literature. For cross country quarterly real GDP growth, inflation and real equity prices the estimates of α are much closer to unity and tend to be around 0.97.

4 Large Panels with Strictly Exogenous Regressors and a Factor Error Structure

Consider the following heterogeneous panel data model

$$y_{it} = \boldsymbol{\alpha}_i' \mathbf{d}_t + \boldsymbol{\beta}_i' \mathbf{x}_{it} + u_{it}, \tag{24}$$

where $\mathbf{d}_t$ is a $n \times 1$ vector of observed common effects (including deterministics such as intercepts or seasonal dummies), $\mathbf{x}_{it}$ is a $k \times 1$ vector of observed individual-specific regressors on the ith cross-section unit at time t, and disturbances, u_{it}, have the following common factor structure

$$u_{it} = \gamma_{i1} f_{1t} + \gamma_{i2} f_{2t} + ... + \gamma_{im} f_{mt} + e_{it} = \boldsymbol{\gamma}_i' \mathbf{f}_t + e_{it}, \tag{25}$$

in which $\mathbf{f}_t = (f_{1t}, f_{2t}, ..., f_{mt})'$ is an m-dimensional vector of unobservable common factors, and $\boldsymbol{\gamma}_i = (\gamma_{i1}, \gamma_{i2}, ..., \gamma_{im})'$ is the associated $m \times 1$ vector of factor loadings. The number of factors, m, is assumed to be fixed relative to N, and in particular $m << N$. The idiosyncratic errors, e_{it}, could be CWD, for example, being generated by a spatial process, or, more generally, by a weak factor structure. For estimation purposes, as in the case of panels with group effects, the factor loadings, $\boldsymbol{\gamma}_i$, could be either random or fixed unknown coefficients. We distinguish between the homogeneous coefficient case where $\boldsymbol{\beta}_i = \boldsymbol{\beta}$ for all i, and the heterogeneous case where $\boldsymbol{\beta}_i$ are random draws from a given distribution. In the latter case, we assume that the object of interest is the mean coefficients, $\boldsymbol{\beta} = E(\boldsymbol{\beta}_i)$, for all i. When the regressors, $\mathbf{x}_{it}$, are strictly exogenous and the deviations $\boldsymbol{\upsilon}_i = \boldsymbol{\beta}_i - \boldsymbol{\beta}$ are distributed independently of the errors and the regressors, the mean coefficients, $\boldsymbol{\beta}$, can be consistently estimated using pooled as well as mean group estimation procedures. But only mean group estimation will be consistent if the regressors are weakly exogenous and/or if the deviations are correlated with the regressors/errors.[4]

The assumption of slope homogeneity is also crucially important for the derivation of the asymptotic distribution of the pooled or the mean group estimators of $\boldsymbol{\beta}$. Under slope homogeneity, the asymptotic distribution of the estimator of $\boldsymbol{\beta}$ typically converges at the rate of $\sqrt{NT}$, whilst under slope heterogeneity the rate is $\sqrt{N}$. In view of the uncertainty regarding the assumption of

[4] Pooled estimation is carried out assuming that $\boldsymbol{\beta}_i = \boldsymbol{\beta}$ for all i, whilst mean group estimation allows for slope heterogeneity and estimates $\boldsymbol{\beta}$ by the average of the individual estimates of $\boldsymbol{\beta}_i$ (Pesaran and Smith (1995)).

slope heterogeneity, non-parametric estimators of the variance matrix of the pooled and mean group estimators are proposed.[5] In the following sub-sections we review a number of different estimators of $\boldsymbol{\beta}$ proposed in the literature.

4.1 PC estimators

The principal components (PC) approach proposed by Coakley, Fuertes, and Smith (2002) and Bai (2009), by requiring that $N^{-1}\boldsymbol{\Gamma}'\boldsymbol{\Gamma}$ tends to a positive definite matrix, implicitly assumes that all the unobserved common factors in (25) are strong. Coakley, Fuertes, and Smith (2002) consider the panel data model with strictly exogenous regressors and homogeneous slopes (i.e., $\boldsymbol{\beta}_i = \boldsymbol{\beta}$), and propose a two-stage estimation procedure. In the first stage, PCs are extracted from the OLS residuals as proxies for the unobserved variables, and in the second step the estimated factors are treated as observable and the following augmented regression is estimated

$$y_{it} = \boldsymbol{\alpha}_i'\mathbf{d}_t + \boldsymbol{\beta}'\mathbf{x}_{it} + \boldsymbol{\gamma}_i'\hat{\mathbf{f}}_t + \varepsilon_{it}, \text{ for } i = 1, 2, ..., N;\ t = 1, 2, ..., T, \tag{26}$$

where $\hat{\mathbf{f}}_t$ is an $m \times 1$ vector of principal components of the residuals computed in the first stage. The resultant estimator of $\boldsymbol{\beta}$ is consistent for N and T large, so long as $\mathbf{f}_t$ and the regressors, $\mathbf{x}_{it}$, are uncorrelated. However, if the factors and the regressors are correlated, as it is likely to be the case in practice, the two-stage estimator becomes inconsistent (Pesaran (2006)).

Building on Coakley, Fuertes, and Smith (2002), Bai (2009) has proposed an iterative method which consists of alternating the PC method applied to OLS residuals and the least squares estimation of (26), until convergence. In particular, to simplify the exposition suppose $\boldsymbol{\alpha}_i = \mathbf{0}$. Then the least squares estimator of $\boldsymbol{\beta}$ and $\mathbf{F}$ is the solution of the following set of non-linear equations:

$$\hat{\boldsymbol{\beta}}_{PC} = \left(\sum_{i=1}^{N} \mathbf{X}_i'\mathbf{M}_{\hat{F}}\mathbf{X}_i\right)^{-1} \sum_{i=1}^{N} \mathbf{X}_i'\mathbf{M}_{\hat{F}}\mathbf{y}_i,$$
$$\frac{1}{NT}\sum_{i=1}^{N}\left(\mathbf{y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}}_{PC}\right)\left(\mathbf{y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}}_{PC}\right)'\hat{\mathbf{F}} = \hat{\mathbf{F}}\hat{\mathbf{V}},$$

where $\mathbf{X}_i = (\mathbf{x}_{i1}, \mathbf{x}_{i2}, ..., \mathbf{x}_{iT})'$ is the matrix of observations on $\mathbf{x}_{it}$, $\mathbf{y}_i = (y_{i1}, y_{i2}, ..., y_{iT})'$ is the vector of observations on y_{it}, $\mathbf{M}_{\hat{F}} = \mathbf{I}_T - \hat{\mathbf{F}}\left(\hat{\mathbf{F}}'\hat{\mathbf{F}}\right)^{-1}\hat{\mathbf{F}}'$, $\hat{\mathbf{F}} = \left(\hat{\mathbf{f}}_1, \hat{\mathbf{f}}_2, ..., \hat{\mathbf{f}}_T\right)'$, and $\hat{\mathbf{V}}$ is a diagonal matrix

[5] Tests of slope homogeneity hypothesis in static and dynamic panels are discussed in Pesaran and Yamagata (2008).

with the m largest eigenvalues of the matrix $\frac{1}{NT}\sum_{i=1}^{N}\left(\mathbf{y}_i-\mathbf{X}_i\hat{\boldsymbol{\beta}}_{PC}\right)\left(\mathbf{y}_i-\mathbf{X}_i\hat{\boldsymbol{\beta}}_{PC}\right)'$ arranged in a decreasing order. The solution $\hat{\boldsymbol{\beta}}_{PC}$, $\hat{\mathbf{F}}$ and $\hat{\boldsymbol{\gamma}}_i=\left(\hat{\mathbf{F}}'\hat{\mathbf{F}}\right)^{-1}\hat{\mathbf{F}}'\left(\mathbf{y}_i-\mathbf{X}_i\hat{\boldsymbol{\beta}}_{PC}\right)$ minimizes the sum of squared residuals function,

$$SSR_{NT}\left(\boldsymbol{\beta},\{\boldsymbol{\gamma}_i\}_{i=1}^{N},\{\mathbf{f}_t\}_{t=1}^{T}\right)=\sum_{i=1}^{N}\left(\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta}-\mathbf{F}\boldsymbol{\gamma}_i\right)'\left(\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta}-\mathbf{F}\boldsymbol{\gamma}_i\right),$$

where $\mathbf{F}=(\mathbf{f}_1,\mathbf{f}_2,...,\mathbf{f}_T)'$. This function is a Gaussian quasi maximum likelihood function of the model and in this respect, Bai's iterative principal components estimator can also be seen as a quasi maximum likelihood estimator, since it minimizes the quasi likelihood function.

Bai (2009) shows that such an estimator is consistent even if common factors are correlated with the explanatory variables. Specifically, the least square estimator of $\boldsymbol{\beta}$ obtained from the above procedure, $\hat{\boldsymbol{\beta}}_{PC}$, is consistent if both N and T tend to infinity, without any restrictions on the ratio T/N. When in addition $T/N\rightarrow K>0$, $\hat{\boldsymbol{\beta}}_{PC}$ converges at the rate $\sqrt{NT}$, but the limiting distribution of $\sqrt{NT}\left(\hat{\boldsymbol{\beta}}_{PC}-\boldsymbol{\beta}\right)$ does not necessarily have a zero mean. Nevertheless, Bai shows that the asymptotic bias can be consistently estimated and proposes a bias corrected estimator.

But it is important to bear in mind that PC-based estimators generally require the determination of the unknown number of strong factors (PCs), m, to be included in the second stage of estimation, and this can introduce some degree of sampling uncertainty into the analysis. There is now a large literature that considers the estimation of m, assuming all the m factors to be strong. See, for example, Bai and Ng (2002 and 2007), Kapetanios (2004 and 2010), Amengual and Watson (2007), Hallin and Liska (2007), Onatski (2009 and 2010), Ahn and Horenstein (2013), Breitung and Pigorsch (2013), Choi and Jeong (2013) and Harding (2013). There are also a number of useful surveys by Bai and Ng (2008), Stock and Watson (2011) and Breitung and Choi (2013), amongst others, that can be consulted for detailed discussions of these methods and additional references. An extensive Monte Carlo investigation into the small sample performance of different selection/estimation methods is provided in Choi and Jeong (2013).

4.2 CCE estimators

Pesaran (2006) suggests the Common Correlated Effects (CCE) estimation procedure that consists of approximating the linear combinations of the unobserved factors by cross-sectional averages of the dependent and explanatory variables, and then running standard panel regressions augmented with these cross-sectional averages. Both pooled and mean group versions are proposed, depending on the assumption regarding the slope homogeneity.

Under slope heterogeneity the CCE approach assumes that $\boldsymbol{\beta}_i' s$ follow the random coefficient model

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{\upsilon}_i, \quad \boldsymbol{\upsilon}_i \sim IID(\mathbf{0}, \boldsymbol{\Omega}_{\boldsymbol{\upsilon}}) \quad \text{for } i = 1, 2, ..., N,$$

where the deviations, $\boldsymbol{\upsilon}_i$, are distributed independently of $e_{jt}, \mathbf{x}_{jt}$, and $\mathbf{d}_t$, for all i, j and t. Since in many empirical applications where cross-sectional dependence is caused by unobservable factors, these factors are correlated with the regressors, and the following model for the individual-specific regressors in (24) is adopted

$$\mathbf{x}_{it} = \mathbf{A}_i'\mathbf{d}_t + \boldsymbol{\Gamma}_i'\mathbf{f}_t + \mathbf{v}_{it}, \tag{27}$$

where $\mathbf{A}_i$ and $\boldsymbol{\Gamma}_i$ are $n \times k$ and $m \times k$ factor loading matrices with fixed components, $\mathbf{v}_{it}$ is the idiosyncratic component of $\mathbf{x}_{it}$ distributed independently of the common effects $\mathbf{f}_{t'}$ and errors $e_{jt'}$ for all i, j, t and t'. However, $\mathbf{v}_{it}$ is allowed to be serially correlated, and cross-sectionally weakly correlated.

Equations (24), (25) and (27) can be combined into the following system of equations

$$\mathbf{z}_{it} = \begin{pmatrix} y_{it} \\ \mathbf{x}_{it} \end{pmatrix} = \mathbf{B}_i'\mathbf{d}_t + \mathbf{C}_i'\mathbf{f}_t + \boldsymbol{\xi}_{it}, \tag{28}$$

where

$$\begin{aligned} \boldsymbol{\xi}_{it} &= \begin{pmatrix} e_{it} + \boldsymbol{\beta}_i'\mathbf{v}_{it} \\ \mathbf{v}_{it} \end{pmatrix}, \\ \mathbf{B}_i &= \begin{pmatrix} \boldsymbol{\alpha}_i & \mathbf{A}_i \end{pmatrix} \begin{pmatrix} 1 & \mathbf{0} \\ \boldsymbol{\beta}_i & \mathbf{I}_k \end{pmatrix}, \mathbf{C}_i = \begin{pmatrix} \boldsymbol{\gamma}_i & \boldsymbol{\Gamma}_i \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \boldsymbol{\beta}_i & \mathbf{I}_k \end{pmatrix}. \end{aligned}$$

Consider the weighted average of $\mathbf{z}_{it}$ using the weights w_i satisfying the granularity conditions (1)-(2):

$$\bar{\mathbf{z}}_{wt} = \bar{\mathbf{B}}_w'\mathbf{d}_t + \bar{\mathbf{C}}_w'\mathbf{f}_t + \bar{\boldsymbol{\xi}}_{wt},$$

where

$$\bar{\mathbf{z}}_{wt} = \sum_{i=1}^{N} w_i \mathbf{z}_{it},$$

$$\bar{\mathbf{B}}_w = \sum_{i=1}^{N} w_i \mathbf{B}_i \ , \ \ \bar{\mathbf{C}}_w = \sum_{i=1}^{N} w_i \mathbf{C}_i, \quad \text{and} \ \ \bar{\boldsymbol{\xi}}_{wt} = \sum_{i=1}^{N} w_i \boldsymbol{\xi}_{it}.$$

Assume that[6]

$$Rank(\bar{\mathbf{C}}_w) = m \leq k+1, \tag{29}$$

we have

$$\mathbf{f}_t = (\bar{\mathbf{C}}_w \bar{\mathbf{C}}_w')^{-1} \bar{\mathbf{C}}_w \left(\bar{\mathbf{z}}_{wt} - \bar{\mathbf{B}}_w' \mathbf{d}_t - \bar{\boldsymbol{\xi}}_{wt} \right). \tag{30}$$

Under the assumption that e_{it}'s and $\mathbf{v}_{it}$'s are CWD processes, it is possible to show that (see Pesaran and Tosetti (2011))

$$\bar{\boldsymbol{\xi}}_{wt} \overset{q.m.}{\to} \mathbf{0}, \tag{31}$$

which implies

$$\mathbf{f}_t - (\bar{\mathbf{C}}_w \bar{\mathbf{C}}_w')^{-1} \bar{\mathbf{C}}_w \left(\bar{\mathbf{z}}_{wt} - \bar{\mathbf{B}}' \mathbf{d}_t \right) \overset{q.m.}{\to} 0, \text{ as } N \to \infty, \tag{32}$$

where

$$\mathbf{C} = \lim_{N \to \infty} (\bar{\mathbf{C}}_w) = \tilde{\boldsymbol{\Gamma}} \begin{pmatrix} 1 & \mathbf{0} \\ \boldsymbol{\beta} & \mathbf{I}_k \end{pmatrix}, \tag{33}$$

$\tilde{\boldsymbol{\Gamma}} = (E(\boldsymbol{\gamma}_i), E(\boldsymbol{\Gamma}_i))$, and $\boldsymbol{\beta} = E(\boldsymbol{\beta}_i)$. Therefore, the unobservable common factors, $\mathbf{f}_t$, can be well approximated by a linear combination of observed effects, $\mathbf{d}_t$, the cross-sectional averages of the dependent variable, $\bar{y}_{wt}$, and those of the individual-specific regressors, $\bar{\mathbf{x}}_{wt}$.

When the parameters of interest are the cross-sectional means of the slope coefficients, $\boldsymbol{\beta}$, we can consider two alternative estimators, the CCE Mean Group (CCEMG) estimator, originally proposed by Pesaran and Smith (1995), and the CCE Pooled (CCEP) estimator. Let $\bar{\mathbf{M}}_w$ be defined by

$$\bar{\mathbf{M}}_w = \mathbf{I}_T - \bar{\mathbf{H}}_w (\bar{\mathbf{H}}_w' \bar{\mathbf{H}}_w)^{+} \bar{\mathbf{H}}_w', \tag{34}$$

where $\mathbf{A}^{+}$ denotes the Moore-Penrose inverse of matrix $\mathbf{A}$, $\bar{\mathbf{H}}_w = (\mathbf{D}, \bar{\mathbf{Z}}_w)$, and $\mathbf{D}$ and $\bar{\mathbf{Z}}_w$ are, respectively, the matrices of the observations on $\mathbf{d}_t$ and $\bar{\mathbf{z}}_{wt} = (\bar{y}_{wt}, \bar{\mathbf{x}}_{wt}')'$.

[6] This assumption can be relaxed. See Pesaran (2006).

The CCEMG is a simple average of the estimators of the individual slope coefficients[7]

$$\hat{\boldsymbol{\beta}}_{CCEMG} = N^{-1}\sum_{i=1}^{N}\hat{\boldsymbol{\beta}}_{CCE,i}, \tag{35}$$

where

$$\hat{\boldsymbol{\beta}}_{CCE,i} = (\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{X}_i)^{-1}\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{y}_i. \tag{36}$$

Pesaran (2006) shows that, under some general conditions, $\hat{\boldsymbol{\beta}}_{CCEMG}$ is asymptotically unbiased for $\boldsymbol{\beta}$, and as $(N,T)\to\infty$,

$$\sqrt{N}(\hat{\boldsymbol{\beta}}_{CCEMG}-\boldsymbol{\beta}) \xrightarrow{d} N(\mathbf{0},\boldsymbol{\Sigma}_{CCEMG}), \tag{37}$$

where $\boldsymbol{\Sigma}_{CCEMG}=\boldsymbol{\Omega}_v$. A consistent estimator of the variance of $\hat{\boldsymbol{\beta}}_{CCEMG}$, denoted by $Var\left(\hat{\boldsymbol{\beta}}_{CCEMG}\right)$, can be obtained by adopting the non-parametric estimator:

$$\widehat{Var}\left(\hat{\boldsymbol{\beta}}_{CCEMG}\right) = N^{-1}\hat{\boldsymbol{\Sigma}}_{CCEMG} = \frac{1}{N\left(N-1\right)}\sum_{i=1}^{N}(\hat{\boldsymbol{\beta}}_{CCE,i}-\hat{\boldsymbol{\beta}}_{CCEMG})(\hat{\boldsymbol{\beta}}_{CCE,i}-\hat{\boldsymbol{\beta}}_{CCEMG})'. \tag{38}$$

The CCEP estimator is given by

$$\hat{\boldsymbol{\beta}}_{CCEP} = \left(\sum_{i=1}^{N} w_i\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{X}_i\right)^{-1}\sum_{i=1}^{N} w_i\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{y}_i. \tag{39}$$

Under some general conditions, Pesaran (2006) proves that $\hat{\boldsymbol{\beta}}_{CCEP}$ is asymptotically unbiased for $\boldsymbol{\beta}$, and, as $(N,T)\to\infty$,

$$\left(\sum_{i=1}^{N} w_i^2\right)^{-1/2}\left(\hat{\boldsymbol{\beta}}_{CCEP}-\boldsymbol{\beta}\right) \xrightarrow{d} N(0,\boldsymbol{\Sigma}_{CCEP}),$$

where

$$\begin{aligned}
\boldsymbol{\Sigma}_{CCEP} &= \boldsymbol{\Psi}^{*-1}\mathbf{R}^{*}\boldsymbol{\Psi}^{*-1},\\
\boldsymbol{\Psi}^{*} &= \lim_{N\to\infty}\left(\sum_{i=1}^{N} w_i\boldsymbol{\Sigma}_i\right), \quad \mathbf{R}^{*} = \lim_{N\to\infty}\left[N^{-1}\sum_{i=1}^{N}\tilde{w}_i^2(\boldsymbol{\Sigma}_i\boldsymbol{\Omega}_v\boldsymbol{\Sigma}_i)\right],\\
\boldsymbol{\Sigma}_i &= p\lim_{T\to\infty}\left(T^{-1}\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{X}_i\right), \text{ and } \tilde{w}_i = \frac{w_i}{\sqrt{N^{-1}\sum_{i=1}^{N} w_i^2}}.
\end{aligned}$$

[7]Pesaran (2006) also considered a weighted average of individual $\hat{\mathbf{b}}_i$, with weights inversely proportional to the individual variances.

A consistent estimator of $Var\left(\hat{\boldsymbol{\beta}}_{CCEP}\right)$, denoted by $\widehat{Var}\left(\hat{\boldsymbol{\beta}}_{CCEP}\right)$, is given by

$$\widehat{Var}\left(\hat{\boldsymbol{\beta}}_{CCEP}\right)=\left(\sum_{i=1}^{N}w_i^2\right)\hat{\boldsymbol{\Sigma}}_{CCEP}=\left(\sum_{i=1}^{N}w_i^2\right)\hat{\boldsymbol{\Psi}}^{*-1}\hat{\mathbf{R}}^{*}\hat{\boldsymbol{\Psi}}^{*-1}, \tag{40}$$

where

$$\begin{aligned}\hat{\boldsymbol{\Psi}}^{*} &= \sum_{i=1}^{N}w_i\left(\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{X}_i}{T}\right),\\ \hat{\mathbf{R}}^{*} &= \frac{1}{N-1}\sum_{i=1}^{N}\tilde{w}_i^2\left(\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{X}_i}{T}\right)(\hat{\boldsymbol{\beta}}_{CCE,i}-\hat{\boldsymbol{\beta}}_{CCEMG})(\hat{\boldsymbol{\beta}}_{CCE,i}-\hat{\boldsymbol{\beta}}_{CCEMG})'\left(\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{X}_i}{T}\right).\end{aligned}$$

The rate of convergence of $\hat{\boldsymbol{\beta}}_{CCEMG}$ and $\hat{\boldsymbol{\beta}}_{CCEP}$ is $\sqrt{N}$ when $\boldsymbol{\Omega}_v\neq\mathbf{0}$. Note that even if $\boldsymbol{\beta}_i$ were observed for all i, the estimate of $\boldsymbol{\beta}=E\left(\boldsymbol{\beta}_i\right)$ cannot converge at a faster rate than $\sqrt{N}$. If the individual slope coefficients $\boldsymbol{\beta}_i$ are homogeneous (namely if $\boldsymbol{\Omega}_v=\mathbf{0}$), $\hat{\boldsymbol{\beta}}_{CCEMG}$ and $\hat{\boldsymbol{\beta}}_{CCEP}$ are still consistent and converge at the rate $\sqrt{NT}$ rather than $\sqrt{N}$.

Advantage of the nonparametric estimators $\hat{\boldsymbol{\Sigma}}_{CCEMG}$ and $\hat{\boldsymbol{\Sigma}}_{CCEP}$ is that they do not require knowledge of the weak cross-sectional dependence of e_{it} (provided it is sufficiently weak) nor the knowledge of serial correlation of e_{it}. An important question is whether the non-parametric variance estimators $\widehat{Var}\left(\hat{\boldsymbol{\beta}}_{CCEMG}\right)$ and $\widehat{Var}\left(\hat{\boldsymbol{\beta}}_{CCEP}\right)$ can be used in both cases of homogeneous and heterogeneous slopes. As established in Pesaran and Tosetti (2011), the asymptotic distribution of $\hat{\boldsymbol{\beta}}_{CCEMG}$ and $\hat{\boldsymbol{\beta}}_{CCEP}$ depends on nuisance parameters when slopes are homogeneous ($\boldsymbol{\Omega}_v=\mathbf{0}$), including the extent of cross-sectional correlations of e_{it} and their serial correlation structure. However, it can be shown that the robust non-parametric estimators $\widehat{Var}\left(\hat{\boldsymbol{\beta}}_{CCEMG}\right)$ and $\widehat{Var}\left(\hat{\boldsymbol{\beta}}_{CCEP}\right)$ are consistent when the regressor-specific components, $\mathbf{v}_{it}$, are independently distributed across i.

The CCE continues to be applicable even if the rank condition (29) is not satisfied. Failure of the rank condition can occur if there is an unobserved factor for which the average of the loadings in the y_{it} and $\mathbf{x}_{it}$ equations tends to a zero vector. This could happen if, for example, the factor in question is weak, in the sense defined above. Another possible reason for failure of the rank condition is if the number of unobservable factors, m, is larger than $k+1$, where k is the number of the unit-specific regressors included in the model. In such cases, common factors cannot be estimated from cross-sectional averages. However, it is possible to show that the cross-sectional means of the slope coefficients, $\boldsymbol{\beta}_i$, can still be consistently estimated, under the additional assumption that the unobserved factor loadings, $\boldsymbol{\gamma}_i$, in equation (25) are independently and identically distributed across i, and of e_{jt}, $\mathbf{v}_{jt}$, and $\mathbf{g}_t = (\mathbf{d}_t', \mathbf{f}_t')'$ for all i, j and t, and uncorrelated with the loadings attached to the regressors, $\boldsymbol{\Gamma}_i$. The consequences of the correlation between loadings $\boldsymbol{\gamma}_i$ and $\boldsymbol{\Gamma}_i$ for the performance of CCE estimators in the rank deficient case are documented in Sarafidis and Wansbeek (2012).

An advantage of the CCE approach is that it yields consistent estimates under a variety of situations. Kapetanios, Pesaran, and Yagamata (2011) consider the case where the unobservable common factors follow unit root processes and could be cointegrated. They show that the asymptotic distribution of panel estimators in the case of I(1) factors is similar to that in the stationary case. Pesaran and Tosetti (2011) prove consistency and asymptotic normality for CCE estimators when $\{e_{it}\}$ are generated by a spatial process. Chudik, Pesaran, and Tosetti (2011) prove consistency and asymptotic normality of the CCE estimators when errors are subject to a finite number of unobserved strong factors and an infinite number of weak and/or semi-strong unobserved common factors as in (20)-(21), provided that certain conditions on the loadings of the infinite factor structure are satisfied. A further advantage of the CCE approach is that it does not require an a priori knowledge of the number of unobserved common factors.

In a Monte Carlo (MC) study, Coakley, Fuertes, and Smith (2006) compare ten alternative estimators for the mean slope coefficient in a linear heterogeneous panel regression with strictly exogenous regressors and unobserved common (correlated) factors. Their results show that, overall, the mean group version of the CCE estimator stands out as the most efficient and robust. These conclusions are in line with those in Kapetanios and Pesaran (2007) and Chudik, Pesaran, and Tosetti (2011), who investigate the small sample properties of CCE estimators and the estimators based on principal components. The MC results show that PC augmented methods do not perform as well as the CCE approach, and can lead to substantial size distortions, due, in part, to the small sample errors in the number of factors selection procedure. In a recent theoretical study, Westerlund and Urbain (2011) investigate the merits of the CCE and PC estimators in the case of homogeneous slopes and known number of unobserved common factors and find that, although the PC estimates of factors are more efficient than the cross-sectional averages, the CCE estimators of slope coefficients generally perform the best.

5 Dynamic Panel Data Models with a Factor Error Structure

The problem of estimation of panels subject to cross-sectional error dependence becomes much more complicated once the assumption of strict exogeneity of the unit-specific regressors is relaxed. One important example, is the panel data model with lagged dependent variables and unobserved common factors (possibly correlated with the regressors):[8]

$$y_{it} = \lambda_i y_{i,t-1} + \boldsymbol{\beta}_i' \mathbf{x}_{it} + u_{it}, \tag{41}$$

$$u_{it} = \boldsymbol{\gamma}_i' \mathbf{f}_t + e_{it}, \tag{42}$$

for $i = 1, 2, ..., N$; $t = 1, 2, ..., T$. It is assumed that $|\lambda_i| < 1$, and the dynamic processes have started a long time in the past. As in the previous section, we distinguish between the case of homogeneous coefficients, where $\lambda_i = \lambda$ and $\boldsymbol{\beta}_i = \boldsymbol{\beta}$ for all i, and the heterogeneous case, where λ_i and $\boldsymbol{\beta}_i$ are randomly distributed across units and the object of interest are the mean coefficients $\lambda = E(\lambda_i)$ and $\boldsymbol{\beta} = E(\boldsymbol{\beta}_i)$. This distinction is more important for dynamic panels, since not only the rate of convergence is affected by the presence of coefficient heterogeneity, but, as shown by Pesaran and Smith (1995), pooled least squares estimators are no longer consistent in the case of dynamic panel data models with heterogeneous coefficients.

It is convenient to define the vector of regressors $\boldsymbol{\zeta}_{it} = (y_{i,t-1}, \mathbf{x}_{it}')'$ and the corresponding parameter vector $\boldsymbol{\pi}_i = \left(\lambda_i, \boldsymbol{\beta}_i'\right)'$ so that (41) can be written as

$$y_{it} = \boldsymbol{\pi}_i' \boldsymbol{\zeta}_{it} + u_{it}. \tag{43}$$

5.1 Quasi maximum likelihood estimator

Moon and Weidner (2010) assume $\boldsymbol{\pi}_i = \boldsymbol{\pi}$ for all i and develop a Gaussian quasi maximum likelihood estimator (QMLE) of the homogeneous coefficient vector $\boldsymbol{\pi}$.[9] The QMLE of $\boldsymbol{\pi}$ is

$$\hat{\boldsymbol{\pi}}_{QMLE} = \underset{\boldsymbol{\pi} \in \mathbb{B}}{\arg\min} L_{NT}(\boldsymbol{\pi}),$$

[8] Fixed effects and observed common factors (denoted by $\mathbf{d}_t$ previously) can also be included in the model. They are excluded to simplify the notations.

[9] See also Lee, Moon, and Weidner (2012) for an extension of this framework to panels with measurement errors.

where $\mathbb{B}$ is a compact parameter set assumed to contain the true parameter values, and the objective function is the profile likelihood function:

$$L_{NT}(\boldsymbol{\pi}) = \min_{\{\boldsymbol{\gamma}_i\}_{i=1}^N, \{\mathbf{f}_t\}_{t=1}^T} \frac{1}{NT} \sum_{i=1}^{N} (\mathbf{y}_i - \boldsymbol{\Xi}_i \boldsymbol{\pi} - \mathbf{F}\boldsymbol{\gamma}_i)' (\mathbf{y}_i - \boldsymbol{\Xi}_i \boldsymbol{\pi} - \mathbf{F}\boldsymbol{\gamma}_i),$$

where

$$\boldsymbol{\Xi}_i = \begin{pmatrix} y_{i1} & \mathbf{x}'_{i,2} \\ y_{i,2} & \mathbf{x}'_{i,3} \\ \vdots & \vdots \\ y_{i,T-1} & \mathbf{x}'_{iT} \end{pmatrix}.$$

Both $\hat{\boldsymbol{\pi}}_{QMLE}$ and $\hat{\boldsymbol{\beta}}_{PC}$ minimize the same objective function and therefore, when the same set of regressors is considered, these two estimators are numerically the same, but there are important differences in their bias-corrected versions and in other aspects of the analysis of Bai (2009) and the analysis of Moon and Weidner (2010). The latter paper allows for more general assumptions on regressors, including the possibility of weak exogeneity, and adopts a quadratic approximation of the profile likelihood function, which allows the authors to work out the asymptotic distribution and to conduct inference on the coefficients.

Moon and Weidner (MW) show that $\hat{\boldsymbol{\pi}}_{QMLE}$ is a consistent estimator of $\boldsymbol{\pi}$, as $(N,T) \to \infty$ without any restrictions on the ratio T/N. To derive the asymptotic distribution of $\hat{\boldsymbol{\pi}}_{QMLE}$, MW require $T/N \to \varkappa$, $0 < \varkappa < \infty$, as $(N,T) \to \infty$, and assume that the idiosyncratic errors, e_{it}, are cross-sectionally independent. Under certain high level assumptions, they show that $\sqrt{NT}\,(\hat{\boldsymbol{\pi}}_{QMLE} - \boldsymbol{\pi})$ converges to a normal distribution with a non-zero mean, which is due to two types of asymptotic bias. The first follows from the heteroskedasticity of the error terms, as in Bai (2009), and the second is due to the presence of weakly exogenous regressors. The authors provide consistent estimators of these two components, and propose a bias-corrected QMLE.

There are, however, two important considerations that should be born in mind when using the QMLE proposed by MW. First, it is developed for the case of full slope homogeneity, namely under $\boldsymbol{\pi}_i = \boldsymbol{\pi}$ for all i. This assumption, for example, rules out the inclusion of fixed effects into the model which can be quite restrictive in practice. Although, the unobserved factor component, $\boldsymbol{\gamma}_i'\mathbf{f}_t$, does in principle allow for fixed effects if the first element of $\mathbf{f}_t$ can be constrained to be unity at the estimation stage. A second consideration is the small sample properties of QMLE in the case of models with fixed effects, which are of primarily interest in empirical applications. Simulations reported in Chudik and Pesaran (2013b) suggests that the bias correction does not go far enough and the QMLE procedure could yield tests which are grossly over-sized. To check the robustness of the QMLE to the presence of fixed effects, we carried out a small Monte Carlo experiment in the case of a homogeneous AR(1) panel data model with fixed effects, $\lambda_i = 0.70$, and $N = T = 100$. Using $R = 2,000$ replications, the bias of the bias-corrected QMLE, $\hat{\lambda}_{QMLE}$, turned out to be -0.024, and tests based on $\hat{\lambda}_{QMLE}$ were grossly oversized with the size exceeding 60%.

5.2 PC estimators for dynamic panels

Song (2013) extends Bai (2009)'s approach to dynamic panels with heterogeneous coefficients. The focus of Song's analysis is on the estimation of unit-specific coefficients $\boldsymbol{\pi}_i = \left(\lambda_i, \boldsymbol{\beta}_i'\right)'$. In particular, Song proposes an iterated least squares estimator of $\boldsymbol{\pi}_i$, and shows as in Bai (2009) that the solution can be obtained by alternating the PC method applied to the least squares residuals and the least squares estimation of (41) until convergence. In particular, the least squares estimators of $\boldsymbol{\pi}_i$ and $\mathbf{F}$ are the solution to the following set of non-linear equations

$$\hat{\boldsymbol{\pi}}_{i,PC} = \left(\boldsymbol{\Xi}_i'\mathbf{M}_{\hat{F}}\boldsymbol{\Xi}_i\right)^{-1}\boldsymbol{\Xi}_i'\mathbf{M}_{\hat{F}}\mathbf{y}_i, \text{ for } i = 1, 2, ..., N, \tag{44}$$

$$\frac{1}{NT}\sum_{i=1}^{N}\left(\mathbf{y}_i - \boldsymbol{\Xi}_i\hat{\boldsymbol{\pi}}_{i,PC}\right)\left(\mathbf{y}_i - \boldsymbol{\Xi}_i\hat{\boldsymbol{\pi}}_{i,PC}\right)'\hat{\mathbf{F}} = \hat{\mathbf{F}}\hat{\mathbf{V}}. \tag{45}$$

Song (2013) establishes consistency of $\hat{\boldsymbol{\pi}}_{i,PC}$ when $(N, T) \to \infty$ without any restrictions on T/N. If in addition $T/N^2 \to 0$, Song (2013) shows that $\hat{\boldsymbol{\pi}}_{i,PC}$ is $\sqrt{T}$ consistent, and derives the asymptotic distribution under some additional requirements including the cross-sectional independence of e_{it}. Song (2013) does not provide theoretical results on the estimation of the mean coefficients $\boldsymbol{\pi} = E\left(\boldsymbol{\pi}_i\right)$, but he considers the following mean group estimator based on the individual estimates $\hat{\boldsymbol{\pi}}_{i,PC}$,

$$\hat{\boldsymbol{\pi}}^s_{PCMG} = \frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{\pi}}_{i,PC},$$

in a Monte Carlo study and finds that $\hat{\boldsymbol{\pi}}^s_{PCMG}$ has satisfactory small sample properties in terms of bias and root mean squared error. But he does not provide any results on the asymptotic distribution of $\hat{\boldsymbol{\pi}}^s_{PCMG}$. However, results of a Monte Carlo study presented in Chudik and Pesaran (2013b) suggest that $\sqrt{N}\left(\hat{\boldsymbol{\pi}}^s_{PCMG} - \boldsymbol{\pi}\right)$ is asymptotically normally distributed with mean zero and a covariance matrix that can be estimated by (as in the case of the CCEMG estimator),

$$\widehat{Var}\left(\hat{\boldsymbol{\pi}}^s_{PCMG}\right) = \frac{1}{N(N-1)} \sum_{i=1}^{N} \left(\hat{\boldsymbol{\pi}}^s_i - \hat{\boldsymbol{\pi}}^s_{MG}\right)\left(\hat{\boldsymbol{\pi}}^s_i - \hat{\boldsymbol{\pi}}^s_{MG}\right)'.$$

The test results based on this conjecture tend to perform well so long as T is sufficiently large. However, as with the other PC based estimators, knowledge of the number of factors and the assumption that the factors under consideration are strong continue to play an important role in the small sample properties of the tests based on $\hat{\boldsymbol{\pi}}^s_{MGPC}$.

5.3 Dynamic CCE estimators

The CCE approach as it was originally proposed in Pesaran (2006) does not cover the case where the panel includes a lagged dependent variable or weakly exogenous regressors.[10] Extension of the CCE approach to dynamic panels with heterogeneous coefficients and weakly exogenous regressors is proposed by Chudik and Pesaran (2013b). In what follows we refer to this extension as dynamic CCE.

The inclusion of a lagged dependent variable amongst the regressors has three main consequences for the estimation of the mean coefficients. The first is the well known time series bias,[11] which affects the individual specific estimates and is of order $O\left(T^{-1}\right)$. The second consequence is that the full rank condition becomes necessary for consistent estimation of the mean coefficients unless the $\mathbf{f}_t$ is serially uncorrelated. The third complication arises from the interaction of dynamics and coefficient heterogeneity, which leads to infinite lag order relationships between unobserved common factors and cross-sectional averages of the observables when N is large. This issue also arises in cross-sectional aggregation of heterogeneous dynamic models. See Granger (1980) and Chudik and Pesaran (2013a).

To illustrate these complications, using (41) and recalling assumption $|\lambda_i| < 1$, for all i, then we have

$$y_{it} = \sum_{\ell=0}^{\infty} \lambda_i^{\ell} \boldsymbol{\beta}_i' \mathbf{x}_{i,t-\ell} + \sum_{\ell=0}^{\infty} \lambda_i^{\ell} \boldsymbol{\gamma}_i' \mathbf{f}_{t-\ell} + \sum_{\ell=0}^{\infty} \lambda_i^{\ell} e_{i,t-\ell}. \quad (46)$$

Taking weighted cross-sectional averages, and assuming independence of λ_i, $\boldsymbol{\beta}_i$, and $\boldsymbol{\gamma}_i$, strict

[10] See Everaert and Groote (2012) who derive the asymptotic bias of the CCE pooled estimator in the case of dynamic *homogeneous* panels.

[11] This bias was first quantified in the case of a simple AR(1) model by Hurwicz (1950).

exogeneity of $\mathbf{x}_{it}$, and weak cross-sectional dependence of $\{e_{it}\}$, we obtain (following the arguments in Chudik and Pesaran (2013a)),

$$\overline{y}_{wt} = a(L)\,\boldsymbol{\gamma}'\mathbf{f}_t + a(L)\,\boldsymbol{\beta}'\overline{\mathbf{x}}_{wt} + \xi_{wt}, \tag{47}$$

where $a(L) = \sum_{\ell=0}^{\infty} a_\ell L^\ell$, with $a_\ell = E\left(\lambda_i^\ell\right)$, $\boldsymbol{\beta} = E(\boldsymbol{\beta}_i)$, and $\boldsymbol{\gamma} = E(\boldsymbol{\gamma}_i)$. Under the assumption that the idiosyncratic errors are cross-sectionally weakly dependent, we have $\xi_{wt} \xrightarrow{p} 0$, as $N \to \infty$, with the rate of convergence depending on the degree of cross-sectional dependence of $\{e_{it}\}$ and the granularity of $\mathbf{w}$. In the case where $\mathbf{w}$ satisfies the usual granularity conditions (1)-(2), and the exponent of cross-sectional dependence of e_{it} is $\alpha_e \leq 1/2$, we have $\xi_{wt} = O_p\left(N^{-1/2}\right)$. In the special case where $\boldsymbol{\beta} = 0$ and $m = 1$, (47) reduces to

$$\overline{y}_{wt} = \gamma a(L)\, f_t + O_p\left(N^{-1/2}\right).$$

The extent to which f_t can be accurately approximated by $\overline{y}_{wt}$ and its lagged values depends on the rate at which, $a_\ell = E\left(\lambda_i^\ell\right)$, the coefficients in the polynomial lag operator, $a(L)$, decay with ℓ, and the size of the cross section dimension, N. The coefficients in $a(L)$ are given by the moments of λ_i and therefore these coefficients need not be absolute summable if the support of λ_i is not sufficiently restricted in the neighborhood of the unit circle (see Granger (1980) and Chudik and Pesaran (2013a)). Assuming that for all i the support of λ_i lies strictly within the unit circle, it is then easily seen that α_ℓ will then decay exponentially and for N sufficiently large, f_t can be well approximated by $\overline{y}_{wt}$ and a number of its lagged values.[12] The number of lagged values of $\overline{y}_{wt}$ needed to approximate f_t rises with T but at a slower rate.[13]

In the general case where $\boldsymbol{\beta}$ is nonzero, $\mathbf{x}_{it}$ are weakly exogenous, and $m \geq 1$, Chudik and Pesaran (2013b) show that there exists the following large N distributed lag relationship between the unobserved common factors and cross-sectional averages of the dependent variable and the

[12] For example if λ_i is distributed uniformly over the range $(0, b)$ where $0 < b < 1$, we have $\alpha_\ell = E(\lambda_i^\ell) = b^\ell/(1+\ell)$, which decays exponentially with ℓ.

[13] The number of lags cannot increase too fast, otherwise there will not be a sufficient number of observations to accurately estimate the parameters, whilst at the same time a sufficient number of lags are needed to ensure that the factors are well approximated. Setting the number of lags equal to $T^{1/3}$ seems to be a good choice, balancing the effects of the above two opposing considerations. See Berk (1974), Said and Dickey (1984), and Chudik and Pesaran (2013c) for a related discussion on the choice of lag truncation for estimation of infinite order autoregressive models.

regressors, $\bar{\mathbf{z}}_{wt} = (\bar{y}_{wt}, \bar{\mathbf{x}}'_{wt})'$,

$$\mathbf{\Lambda}(L)\,\tilde{\mathbf{\Gamma}}'\mathbf{f}_t = \bar{\mathbf{z}}_{wt} + O_p\left(N^{-1/2}\right),$$

where as before $\tilde{\mathbf{\Gamma}} = E(\boldsymbol{\gamma}_i, \mathbf{\Gamma}_i)$ and the decay rate of the matrix coefficients in $\mathbf{\Lambda}(L)$ depends on the heterogeneity of λ_i and $\boldsymbol{\beta}_i$ and other related distributional assumptions. The existence of a large N relationship between the unobserved common factors and cross-sectional averages of variables is not surprising considering that only the components with the largest exponents of cross-sectional dependence can survive cross-sectional aggregation with granular weights. Assuming $\tilde{\mathbf{\Gamma}}$ has full row rank, i.e. $rank\left(\tilde{\mathbf{\Gamma}}\right) = m$, and the distributions of coefficients are such that $\mathbf{\Lambda}^{-1}(L)$ exists and has exponentially decaying coefficients yields the following unit-specific dynamic CCE regressions,

$$y_{it} = \lambda_i y_{i,t-1} + \boldsymbol{\beta}'_i \mathbf{x}_{it} + \sum_{\ell=0}^{p_T} \boldsymbol{\delta}'_{i\ell}\bar{\mathbf{z}}_{w,t-\ell} + e_{yit}, \tag{48}$$

where $\bar{\mathbf{z}}_{wt}$ and its lagged values are used to approximate $\mathbf{f}_t$. The error term e_{yit} consists of three parts: an idiosyncratic term, e_{it}, an error component due to the truncation of possibly infinite distributed lag function, and an $O_p\left(N^{-1/2}\right)$ error component due to the approximation of unobserved common factors based on large N relationships.

Chudik and Pesaran (2013b) consider the least squares estimates of $\boldsymbol{\pi}_i = \left(\lambda_i, \boldsymbol{\beta}'_i\right)'$ based on the above dynamic CCE regressions, denoted as $\hat{\boldsymbol{\pi}}_i = \left(\hat{\lambda}_i, \hat{\boldsymbol{\beta}}'_i\right)'$, and the mean group estimate of $\boldsymbol{\pi} = E(\boldsymbol{\pi}_i)$ based on $\hat{\boldsymbol{\pi}}_i$. To define these estimators, we introduce the following data matrices

$$\tilde{\mathbf{\Xi}}_i = \begin{pmatrix} y_{ip_T} & \mathbf{x}'_{i,p_T+1} \\ y_{i,p_T+1} & \mathbf{x}'_{i,p_T+2} \\ \vdots & \vdots \\ y_{i,T-1} & \mathbf{x}'_{iT} \end{pmatrix}, \quad \bar{\mathbf{Q}}_w = \begin{pmatrix} \bar{\mathbf{z}}'_{w,p_T+1} & \bar{\mathbf{z}}'_{w,p_T} & \cdots & \bar{\mathbf{z}}'_{w,1} \\ \bar{\mathbf{z}}'_{w,p_T+2} & \bar{\mathbf{z}}'_{w,p_T+1} & \cdots & \bar{\mathbf{z}}'_{w,2} \\ \vdots & \vdots & & \vdots \\ \bar{\mathbf{z}}'_{w,T} & \bar{\mathbf{z}}'_{w,T-1} & \cdots & \bar{\mathbf{z}}'_{w,T-p_T} \end{pmatrix}, \tag{49}$$

and the projection matrix $\bar{\mathbf{M}}_q = \mathbf{I}_{T-p_T} - \bar{\mathbf{Q}}_w\left(\bar{\mathbf{Q}}'_w\bar{\mathbf{Q}}_w\right)^+\bar{\mathbf{Q}}'_w$, where $\mathbf{I}_{T-p_T}$ is a $(T-p_T)\times(T-p_T)$ dimensional identity matrix.[14] p_T should be set such that p_T^2/T tends to zero as p_T and T both tend to infinity. Monte Carlo experiments reported in Chudik and Pesaran (2013b) suggest that setting $p_T = T^{1/3}$ could be a good choice in practice.

[14]Matrices $\mathbf{\Xi}_i$, $\bar{\mathbf{Q}}_w$, and $\bar{\mathbf{M}}_q$ depend also on p_T, N and T, but we omit these subscripts to simplify notations.

The individual estimates, $\widehat{\boldsymbol{\pi}}_i$, can now be written as

$$\widehat{\boldsymbol{\pi}}_i = \left(\tilde{\boldsymbol{\Xi}}_i'\bar{\mathbf{M}}_q\tilde{\boldsymbol{\Xi}}_i\right)^{-1}\tilde{\boldsymbol{\Xi}}_i'\bar{\mathbf{M}}_q\tilde{\mathbf{y}}_i, \tag{50}$$

where $\tilde{\mathbf{y}}_i = (y_{i,p_T+1}, y_{i,p_T+2}, ..., y_{i,T})'$. The mean group estimator of $\boldsymbol{\pi} = E(\boldsymbol{\pi}_i) = \left(\lambda, \boldsymbol{\beta}'\right)'$ is given by

$$\widehat{\boldsymbol{\pi}}_{MG} = \frac{1}{N}\sum_{i=1}^{N}\widehat{\boldsymbol{\pi}}_i. \tag{51}$$

Chudik and Pesaran (2013b) show that $\widehat{\boldsymbol{\pi}}_i$ and $\widehat{\boldsymbol{\pi}}_{MG}$ are consistent estimators of $\boldsymbol{\pi}_i$ and $\boldsymbol{\pi}$, respectively, assuming that the rank condition is satisfied and $(N, T, p_T) \to \infty$ such that $p_T^3/T \to \varkappa$, $0 < \varkappa < \infty$, but without any restrictions on the ratio N/T. The rank condition is necessary for the consistency of $\widehat{\boldsymbol{\pi}}_i$ because the unobserved factors are allowed to be correlated with the regressors. If the unobserved common factors were serially uncorrelated (but still correlated with $\mathbf{x}_{it}$), then $\widehat{\boldsymbol{\pi}}_{MG}$ is consistent also in the rank deficient case, despite the inconsistency of $\widehat{\boldsymbol{\pi}}_i$, so long as factor loadings are independently, identically distributed across i. The convergence rate of $\widehat{\boldsymbol{\pi}}_{MG}$ is $\sqrt{N}$ due to the heterogeneity of the slope coefficients. Chudik and Pesaran (2013b) show that $\widehat{\boldsymbol{\pi}}_{MG}$ converges to a normal distribution as $(N, T, p_T) \to \infty$ such that $p_T^3/T \to \varkappa_1$ and $T/N \to \varkappa_2$, $0 < \varkappa_1, \varkappa_2 < \infty$. The ratio N/T needs to be restricted for conducting inference, due to the presence of small time series bias. In the full rank case, the asymptotic variance of $\widehat{\boldsymbol{\pi}}_{MG}$ is given by the variance of $\boldsymbol{\pi}_i$ alone. When the rank condition does not hold, but factors are serially uncorrelated, then the asymptotic variance depends also on other parameters, including the variance of factor loadings. In both cases the asymptotic variance can be consistently estimated non-parametrically, as in (38).

Monte Carlo experiments in Chudik and Pesaran (2013b) show that the dynamic CCE approach performs reasonably well (in terms of bias, RMSE, size and power). This is particularly the case when the parameter of interest is the average slope of the regressors ($\boldsymbol{\beta}$), where the small sample results are quite satisfactory even if N and T are relatively small (around 40). But the situation is different if the parameter of interest is the mean coefficient of the lagged dependent variable (λ). In the case of λ, the CCEMG estimator suffers form the well known time series bias and tests based on it tend to be over-sized, unless T is sufficiently large. To mitigate the consequences of this bias, Chudik and Pesaran (2013b) consider application of half-panel jackknife procedure (Dhaene and Jochmans (2012)), and the recursive mean adjustment procedure (So and Shin (1999)), both of which are easy to implement. The proposed jackknife bias-corrected CCEMG estimator is found to be more effective in mitigating the time series bias, but it can not fully deal with the size distortion when T is relatively small. Improving the small T sample properties of the CCEMG estimator of λ in the heterogeneous panel data models still remains a challenge to be taken on in the future.

The application of CCE approach to static panels with weakly exogenous regressors (namely without lagged dependent variables) has not yet been investigated in the literature. In order to investigate whether the standard CCE mean group and pooled estimators could be applied in this setting, we conducted Monte Carlo experiments. We used the following data generating process

$$y_{it} = c_{yi} + \beta_{0i} x_{it} + \beta_{1i} x_{i,t-1} + u_{it},\ u_{it} = \boldsymbol{\gamma}_i' \mathbf{f}_t + \varepsilon_{it}, \tag{52}$$

and

$$x_{it} = c_{xi} + \alpha_{xi} y_{i,t-1} + \boldsymbol{\gamma}_{xi}' \mathbf{f}_t + v_{it}, \tag{53}$$

for $i = 1, 2, ..., N$, and $t = -99, ..., 0, 1, 2, ..., T$ with the starting values $y_{i,-100} = x_{i,-100} = 0$. This set up allows for feedbacks from $y_{i,t-1}$ to the regressors, thus rendering x_{it} weakly exogenous. The size of the feedback is measured by α_{xi}. The unobserved common factors in $\mathbf{f}_t$ and the unit-specific components v_{it} are generated as independent stationary $AR(1)$ processes:

$$\begin{aligned} f_{t\ell} &= \rho_{f\ell} f_{t-1,\ell} + \varsigma_{ft\ell},\ \varsigma_{ft\ell} \sim IIDN\left(0, 1 - \rho_{f\ell}^2\right), \\ v_{it} &= \rho_{xi} v_{i,t-1} + \varsigma_{it},\ \varsigma_{it} \sim IIDN\left(0, \sigma_{vi}^2\right), \end{aligned} \tag{54}$$

for $i = 1, 2, ..., N$, $\ell = 1, 2, .., m$, and for $t = -99, ..., 0, 1, 2, ..., T$ with the starting values $f_{\ell,-100} = 0$ and $v_{i,-100} = 0$. The first 100 time observations ($t = -99, -98, ..., 0$) are discarded. We generate ρ_{xi}, for $i = 1, 2, ...N$ as $IIDU[0, 0.95]$, and set $\rho_{f\ell} = 0.6$, for $\ell = 1, 2, ..., m$. We also set $\sigma_{vi} = \sqrt{1 - [E(\rho_{xi})]^2}$ for all i.

The fixed effects are generated as $c_{yi} \sim IIDN(1, 1)$, $c_{xi} = c_{yi} + \varsigma_{c_x i}$, where $\varsigma_{c_x i} \sim IIDN(0, 1)$, thus allowing for dependence between x_{it} and c_{yi}. We set $\beta_{1i} = -0.5$ for all i, and generate β_{0i} as $IIDU(0.5, 1)$. We consider two possibilities for the feedback coefficients α_{xi}: weakly exogenous regressors where we generate α_{xi} as draws from $IIDU(0, 1)$ (in which case $E(\alpha_{xi}) = 0.5$), and strictly exogenous regressors where we set $\alpha_{xi} = 0$ for all i. We consider $m = 3$ unobserved common factors, with all factor loadings generated independently in the same way as in Chudik and Pesaran (2013b). Similarly, the idiosyncratic errors, ε_{it}, are generated as in Chudik and Pesaran (2013b) to be heteroskedastic and weakly cross-sectionally dependent. We consider the following combinations of sample sizes: $N \in \{40, 50, 100, 150, 200\}$, $T \in \{20, 50, 100, 150, 200\}$, and set the number of replications to $R = 2000$.

The small sample results for the CCE mean group and pooled estimators (with lagged augmentations) in the case of these experiments with weakly exogenous regressors are presented on the upper panel of Table 1. The rank condition in these experiment does not hold, but this does not seem to cause any major problems for the CCE mean group estimator, which performs very well (in terms of bias and RMSE) for $T > 50$ and for all values of N. Also tests based on this estimator are correctly sized and have good power properties. When $T \leq 50$, we observe a negative bias and the tests are oversized (the rejection rates are in the range of 9 to 75 percent, depending on the sample size). The CCE pooled estimator, however, is no longer consistent in the case of weakly exogenous regressors with heterogeneous coefficients, due to the bias caused by the correlation between the slope coefficients and the regressors. For comparison, we also provide, at the bottom panel of Table 1, the results of the same experiments but with strictly exogenous regressors ($\alpha_{xi} = 0$), where the bias is negligible and all tests are correctly sized.

6 Tests of Error Cross-Sectional Dependence

In this section we provide an overview of alternative approaches to testing the cross-sectional independence or weak dependence of the errors in the following panel data model

$$y_{it} = a_i + \boldsymbol{\beta}_i'\mathbf{x}_{it} + u_{it}, \tag{55}$$

where a_i and $\boldsymbol{\beta}_i$ for $i = 1, 2, ..., N$ are assumed to be fixed unknown coefficients, and $\mathbf{x}_{it}$ is a k-dimensional vector of regressors. We consider both cases where the regressors are strictly and weakly exogenous, as well as when they include lagged values of y_{it}.

The literature on testing for error cross-sectional dependence in large panels follows two separate strands, depending on whether the cross section units are ordered or not. In the case of ordered data sets (which could arise when observations are spatial or belong to given economic or social networks) tests of cross-sectional independence that have high power with respect to such ordered alternatives have been proposed in the spatial econometrics literature. A prominent example of such tests is Moran's I test. See Moran (1948) with further developments by Anselin (1988), Anselin and Bera (1998), Haining (2003), and Baltagi, Song, and Koh (2003).

In the case of cross section observations that do not admit an ordering, tests of cross-sectional dependence are typically based on estimates of pair-wise error correlations (ρ_{ij}) and are applicable when T is sufficiently large so that relatively reliable estimates of ρ_{ij} can be obtained. An early test of this type is the Lagrange multiplier (LM) test of Breusch and Pagan (1980, pp. 247-248) which tests the null hypothesis that *all* pair-wise correlations are zero, namely that $\rho_{ij} = 0$ for all $i \neq j$. This test is based on the average of the *squared* estimates of pair-wise correlations, and under standard regularity conditions it is shown to be asymptotically (as $T \to \infty$) distributed as χ^2 with $N(N-1)/2$ degrees of freedom. The LM test tends to be highly over-sized in the case of panels with relatively large N.

In what follows we review the various attempts made in the literature to develop tests of cross-sectional dependence when N is large and the cross-section units are unordered. But before proceeding further, we first need to consider the appropriateness of the null hypothesis of cross-sectional "independence" or "uncorrelatedness", that underlie the LM test of Breusch and Pagan (1980), namely that all ρ_{ij} are zero for all $i \neq j$, when N is large. The null that underlies the LM test is sensible when N is small and fixed as $T \to \infty$. But when N is relatively large and rising with T, it is unlikely to matter if out of the total $N(N-1)/2$ pair-wise correlations only a few are non-zero. Accordingly, Pesaran (2013) argues that the null of cross-sectionally uncorrelated errors, defined by

$$H_0 : E\left(u_{it}u_{jt}\right) = 0, \text{for all } t \text{ and } i \neq j, \tag{56}$$

is restrictive for large panels and the null of a sufficiently weak cross-sectional dependence could be more appropriate since mere incidence of isolated error dependencies are of little consequence for estimation or inference about the parameters of interest, such as the individual slope coefficients, $\boldsymbol{\beta}_i$, or their average value, $E(\boldsymbol{\beta}_i) = \boldsymbol{\beta}$.

Consider the panel data model (55), and let $\hat{u}_{it}$ be the OLS estimator of u_{it} defined by

$$\hat{u}_{it} = y_{it} - \hat{a}_i - \hat{\boldsymbol{\beta}}_i'\mathbf{x}_{it}, \tag{57}$$

with $\hat{a}_i$, and $\hat{\boldsymbol{\beta}}_i$ being the OLS estimates of a_i and $\boldsymbol{\beta}_i$, based on the T sample observations, $y_t, \mathbf{x}_{it}$, for $t = 1, 2, ..., T$. Consider the sample estimate of the pair-wise correlation of the residuals, $\hat{u}_{it}$ and $\hat{u}_{jt}$, for $i \neq j$

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

In the case where the u_{it} is symmetrically distributed and the regressors are strictly exogenous, then under the null hypothesis of no cross-sectional dependence, $\hat{\rho}_{ij}$ and $\hat{\rho}_{is}$ are cross-sectionally uncorrelated for all i, j and s such that $i \neq j \neq s$. This follows since

$$E\left(\hat{\rho}_{ij}\hat{\rho}_{is}\right) = \sum_{t=1}^{T}\sum_{t'=1}^{T} E\left(\hat{\eta}_{it}\hat{\eta}_{it'}\hat{\eta}_{jt}\hat{\eta}_{st'}\right) = \sum_{t=1}^{T}\sum_{t'=1}^{T} E\left(\hat{\eta}_{it}\hat{\eta}_{it'}\right) E\left(\hat{\eta}_{jt}\right) E\left(\hat{\eta}_{st'}\right) = 0. \tag{58}$$

where $\hat{\eta}_{it} = \hat{u}_{it} \Big/ \left(\sum_{t=1}^{T} \hat{u}_{it}^2\right)^{1/2}$. Note when $\mathbf{x}_{it}$ is strictly exogenous for each i, $\hat{u}_{it}$, being a linear function of u_{it}, for $t = 1, 2, .., T$, will also be symmetrically distributed with zero means, which ensures that η_{it} is also symmetrically distributed around its mean which is zero. Further, under (56) and when N is finite, it is known that (see Pesaran (2004))

$$\sqrt{T}\hat{\rho}_{ij} \overset{a}{\sim} N(0,1), \tag{59}$$

for a given i and j, as $T \to \infty$. The above result has been widely used for constructing tests based on the sample correlation coefficient or its transformations. Noting that, from (59), $T\hat{\rho}_{ij}^2$ is asymptotically distributed as a χ_1^2, it is possible to consider the following statistic

$$CD_{LM} = \sqrt{\frac{1}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \left(T\hat{\rho}_{ij}^2 - 1\right)}. \tag{60}$$

Based on the Euclidean norm of the matrix of sample correlation coefficients, (60) is a version of the Lagrange Multiplier test statistic due to Breusch and Pagan (1980). Frees (1995) first explored the finite sample properties of the LM statistic, calculating its moments for fixed values of T and N, under the normality assumption. He advanced a non-parametric version of the LM statistic based on the Spearman rank correlation coefficient. Dufour and Khalaf (2002) have suggested to apply Monte Carlo exact tests to correct the size distortions of CD_{LM} in finite samples. However, these tests, being based on the bootstrap method applied to the CD_{LM}, are computationally intensive, especially when N is large.

An alternative adjustment to the LM test is proposed by Pesaran, Ullah, and Yamagata (2008), where the LM test is centered to have a zero mean for a fixed T. These authors also propose a correction to the variance of the LM test. The basic idea is generally applicable, but analytical bias corrections can be obtained only under the assumption that the regressors, $\mathbf{x}_{it}$, are strictly exogenous and the errors, u_{it} are normally distributed. Under these assumptions, Pesaran, Ullah, and Yamagata (2008) show that the exact mean and variance of $(N-k)\,\hat{\rho}_{ij}^2$ are given by:

$$\begin{aligned}
\mu_{Tij} &= E\left[(N-k)\,\hat{\rho}_{ij}^2\right] = \frac{1}{T-k} Tr\left[E\left(\mathbf{M}_i\mathbf{M}_j\right)\right], \\
v_{Tij}^2 &= Var\left[(N-k)\,\hat{\rho}_{ij}^2\right] = \left\{Tr\left[E\left(\mathbf{M}_i\mathbf{M}_j\right)\right]\right\}^2 a_{1T} + 2\left\{Tr\left[E\left(\mathbf{M}_i\mathbf{M}_j\right)^2\right]\right\} a_{2T},
\end{aligned}$$

where $a_{1T} = a_{2T} - \left(\frac{1}{T-k}\right)^2$, and $a_{2T} = 3\left[\frac{(T-k-8)(T-k+2)+24}{(T-k+2)(T-k-2)(T-k-4)}\right]^2$, $\mathbf{M}_i = \mathbf{I}_T - \tilde{\mathbf{X}}_i\left(\tilde{\mathbf{X}}_i'\tilde{\mathbf{X}}_i\right)^{-1}\tilde{\mathbf{X}}_i'$ and $\tilde{\mathbf{X}}_i$ is $T\times(k+1)$ matrix of observations on $(1,\mathbf{x}_{it}')'$. The adjusted LM statistic is now given by

$$LM_{Adj} = \sqrt{\frac{2}{N(N-1)}}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\frac{(T-k)\,\hat{\rho}_{ij}^2-\mu_{Tij}}{v_{Tij}}, \tag{61}$$

which is asymptotically $N(0,1)$ under H_0, $T\to\infty$ followed by $N\to\infty$. The asymptotic distribution of LM_{Adj} is derived under sequential asymptotics, but it might be possible to establish it under the joint asymptotics following the method of proof in Schott (2005) or Pesaran (2013).

The application of the LM_{Adj} test to dynamic panels or panels with weakly exogenous regressors is further complicated by the fact that the bias corrections depend on the true values of the unknown parameters and will be difficult to implement. The implicit null of LM tests when T and $N\to\infty$, jointly rather than sequentially could also differ from the null of uncorrelatedness of all pair-wise correlations. To overcome some of these difficulties, Pesaran (2004) has proposed a test that has exactly mean zero for fixed values of T and N. This test is based on the average of pair-wise correlation coefficients

$$CD_P = \sqrt{\frac{2T}{N(N-1)}}\left(\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\hat{\rho}_{ij}\right). \tag{62}$$

As it is established in (58), under the null hypothesis $\hat{\rho}_{ij}$ and $\hat{\rho}_{is}$ are uncorrelated for all $i\neq j\neq s$, but they need not be independently distributed when T is finite. Therefore, the standard central limit theorems cannot be applied to the elements of the double sum in (62) when $(N,T)\to\infty$

jointly, and as shown in Pesaran (2013, Theorem 2) the derivation of the limiting distribution of CD_P statistic involves a number of complications. It is also important to bear in mind that the implicit null of the test in the case of large N depends on the rate at which T expands with N. Indeed, as argued in Pesaran (2004), under the null hypothesis of $\rho_{ij} = 0$ for all $i \neq j$, we continue to have $E\left(\hat{\rho}_{ij}\right) = 0$, even when T is fixed, so long as u_{it} are symmetrically distributed around zero, and the CD_P test continues to hold.

Pesaran (2013) extends the analysis of CD_P test and shows that the implicit null of the test is weak cross-sectional dependence. In particular, the implicit null hypothesis of the test depends on the relative expansion rates of N and T.[15] Using the exponent of cross-sectional dependence, α, developed in Bailey, Kapetanios, and Pesaran (2012) and discussed above, Pesaran (2013) shows that when $T = O\left(N^{\epsilon}\right)$ for some $0 < \epsilon \leq 1$, the implicit null of the CD_P test is given by $0 \leq \alpha < (2 - \epsilon)/4$. This yields the range $0 \leq \alpha < 1/4$ when $(N, T) \to \infty$ at the same rate such that $T/N \to \varkappa$ for some finite positive constant $\varkappa$, and the range $0 \leq \alpha < 1/2$ when T is small relative to N. For larger values of α, as shown by Bailey, Kapetanios, and Pesaran (2012), α can be estimated consistently using the variance of the cross-sectional averages.

Monte Carlo experiments reported in Pesaran (2013) show that the CD_P test has good small sample properties for values of α in the range $0 \leq \alpha \leq 1/4$, even in cases where T is small relative to N, as well as when the test is applied to residuals from pure autoregressive panels so long as there are no major asymmetries in the error distribution.

Other statistics have also been proposed in the literature to test for zero contemporaneous correlation in the errors of panel data model (55).[16] Using results from the literature on *spacing* discussed in (Pyke (1965)), Ng (2006) considers a statistic based on the q^{th} differences of the cumulative normal distribution associated to the $N(N-1)/2$ pair-wise correlation coefficients ordered from the smallest to the largest, in absolute value. Building on the work of John (1971), and under the assumption of normal disturbances, strictly exogenous regressors, and homogeneous slopes, Baltagi, Feng, and Kao (2011) propose a test of the null hypothesis of sphericity, defined by

$$H_0^{BFK} : \mathbf{u}_t \sim IIDN\left(\mathbf{0}, \sigma_u^2 \mathbf{I}_N\right),$$

[15] Pesaran (2013) also derives the exact variance of the CD_P test under the null of cross sectional independence and proposes a slightly modified version of the CD_P test distributed exactly with mean zero and a unit variance.

[16] A recent review is provided by Moscone and Tosetti (2009).

based on the statistic

$$J_{BFK} = \frac{T\left(tr(\hat{\mathbf{S}})/N\right)^{-2} tr(\hat{\mathbf{S}}^2)/N - T - N}{2} - \frac{1}{2} - \frac{N}{2\,(T-1)}, \tag{63}$$

where $\hat{\mathbf{S}}$ is the $N \times N$ sample covariance matrix, computed using the fixed effects residuals under the assumption of slope homogeneity, $\boldsymbol{\beta}_i = \boldsymbol{\beta}$. Under H_0^{BFK}, errors u_{it} are cross-sectionally independent and homoskedastic and the J_{BFK} statistic converges to a standardized normal distribution as $(N,T) \to \infty$ such that $N/T \to \varkappa$ for some finite positive constant $\varkappa$. The rejection of H_0^{BFK} could be caused by cross-sectional dependence, heteroskedasticity, slope heterogeneity, and/or non-normal errors. Simulation results reported in Baltagi, Feng, and Kao (2011) show that this test performs well in the case of homoskedastic, normal errors, strictly exogenous regressors, and homogeneous slopes, although it is oversized for panels with large N and small T, and is sensitive to non-normality of disturbances. Joint assumption of homoskedastic errors and homogeneous slopes is quite restrictive in applied work and therefore the use of the J_{BFK} statistics as a test of cross-sectional dependence should be approached with care.

A slightly modified version of the CD_{LM} statistic, given by

$$LM_S = \sqrt{\frac{T+1}{N(N-1)\,(T+2)}} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \left[(T-1)\,\hat{\rho}_{ij}^2 - 1 \right] \tag{64}$$

has also been considered by Schott (2005), who shows that when the LM_S statistic is computed based on normally distributed observations, as opposed to panel residuals, it converges to $N(0,1)$ under $\rho_{ij} = 0$ for all $i \neq j$ as $(N,T) \to \infty$ such that $N/T \to \varkappa$ for some $0 < \varkappa < \infty$. Monte Carlo simulations reported in Jensen and Schmidt (2011) suggests that the LM_S test has good size properties for various sample sizes when applied to panel residuals in the case when slopes are homogeneous and estimated using the fixed effects approach. However, the LM_S test can lead to severe over-rejection when the slopes are in fact heterogeneous and the fixed effects estimators are used. The over-rejection of the LM_S test could persist even if mean group estimates are used in the computation of the residuals to take care of slope heterogeneity. This is because for relatively small values of T, unlike the LM_{Adj} statistic defined by (61), the LM_S statistic defined by (64) is not guaranteed to have a zero mean exactly.

The problem of testing for cross-sectional dependence in limited dependent variable panel data models with strictly exogenous covariates has also been investigated by Hsiao, Pesaran, and Pick (2012). In this paper the authors derive a LM test and show that in terms of the generalized residuals of Gourieroux et al. (1987), the test reduces to the LM test of Breusch and Pagan (1980). However, not surprisingly as with the linear panel data models, the LM test based on generalized residuals tends to over-reject in panels with large N. They then develop a CD type test based on a number of different residuals, and using Monte Carlo experiments they find that the CD test preforms well for most combinations of N and T.

Sarafidis et al. (2009) propose a test for the null hypothesis of homogeneous cross-sectional dependence

$$H_0 : Var\left(\boldsymbol{\gamma}_i\right) = \mathbf{0}, \tag{65}$$

in a lagged dependent variable model with regressors and residual factor structure (41)-(42) with cross-sectionally uncorrelated idiosyncratic innovations e_{it} against the alternative of heterogeneous cross-sectional dependence

$$H_1 : Var\left(\boldsymbol{\gamma}_i\right) \neq \mathbf{0}. \tag{66}$$

Following Sargan (1988) and exploring two different sets of moment conditions, one valid only under the null and the other valid under both hypotheses, Sarafidis et al. (2009) derive Sargan's difference test based on the first-differenced as well as system based GMM estimators in a large N and fixed T setting. The null hypothesis (65) does not imply that the errors are cross-sectionally uncorrelated, and it allows to examine whether any cross section dependence of errors remains after including time dummies, or after the data is transformed in terms of deviations from time-specific averages. In such cases the CD_P test lacks power and the test by Sarafidis et al. (2009) could have some merits.

The existing literature on testing for error cross-sectional dependence, with the exception of Sarafidis et al. (2009), have mostly focused on the case of strictly exogenous regressors. This assumption is required for both LM_{Adj} and J_{BFK} tests, while Pesaran (2004) shows that the CD_P test is also applicable to autoregressive panel data models so long as the errors are symmetrically distributed. The properties of the CD_P test for dynamic panels that include weakly or strictly exogenous regressors have not yet been investigated.

We conduct Monte Carlo experiments to investigate the performance of these tests in the case of dynamic panels and to shed light also on the performance of LM_S test in the case of heterogeneous slopes. We generate the dependent variable and the regressors in the same way as described in Section 5.3 with the following two exceptions. First, we introduce lags of the dependent variable in (60):

$$y_{it} = c_{yi} + \lambda_i y_{i,t-1} + \beta_{0i} x_{it} + \beta_{1i} x_{i,t-1} + u_{it}, \tag{67}$$

and generate λ_i as $IIDU\left(0, 0.8\right)$. As discussed in Chudik and Pesaran (2013b) the lagged dependent variable coefficients, λ_i, and the feedback coefficients, α_{xi}, in (53) need to be chosen such as to ensure the variances of y_{it} remain bounded. We generate α_{xi} as $IIDU\left(0, 0.35\right)$, which ensures that this condition is met and $E\left(\alpha_{xi}\right) = 0.35/2$. For comparison purposes, we also consider the case of strictly exogenous regressors where we set $\lambda_i = \alpha_{xi} = 0$ for all i. The second exception is the generation of the reduced form errors. In order to consider different options for cross-sectional dependence, we use the following residual factor model to generate the errors u_{it}.

$$u_{it} = \gamma_i g_t + \varepsilon_{it}, \tag{68}$$

where $\varepsilon_{it} \sim IIDN\left(0, \frac{1}{2}\sigma_i^2\right)$ with $\sigma_i^2 \sim \chi^2\left(2\right)$, $g_t \sim IIDN\left(0, 1\right)$ and the factor loadings are generated as

$$\begin{aligned} \gamma_i &= v_{\gamma i}, \text{ for } i = 1, 2, ..., M_\alpha, \\ \gamma_i &= 0, \text{ for } i = M_\alpha + 1, M_\alpha + 2, ..., N, \end{aligned}$$

where $M_\alpha = [N^\alpha]$, $v_{\gamma i} \sim IIDU\left[\mu_v - 0.5, \mu_v + 0.5\right]$. We set $\mu_v = 1$, and consider four values of the exponent of the cross-sectional dependence for the errors, namely $\alpha = 0, 0.25, 0.5$ and 0.75. We also consider the following combinations of $N \in \{40, 50, 100, 150, 200\}$, and $T \in \{20, 50, 100, 150, 200\}$, and use 2000 replications for all experiments.

Table 2 presents the findings for the CD_p, LM_{Adj} and LM_S tests. The rejection rates for J_{BFK} in all cases, including the cross-sectionally independent case of $\alpha = 0$, were all close to 100%, in part due to the error variance heteroskedasticity, and are not included in Table 2. The top panel of Table 2 reports the test results for the case of strictly exogenous regressors, and the bottom part gives the results for the panel data models with weakly exogenous regressors. We see that the CD_P test continues to perform well even when the panel data model contains a lagged dependent variable and other weakly exogenous regressors, for the combination of N and T samples considered. The results also confirm the theoretical finding discussed above that shows the implicit null of the CD_P test is $0 \leq \alpha \leq 0.25$. In contrast, the LM_{Adj} test tends to over-reject when the panel includes dynamics and T is small compared to N. The reported rejection rate when $N = 200$ and $T = 20$ is 14.25 percent.[17] Furthermore, the findings also suggest that the LM_{Adj} test has power when the cross-sectional dependence is very weak, namely in the case when the exponent of cross-sectional dependence is $\alpha = 0.25$. LM_S also over-rejects when T is small relative to N, but the over-rejection is much more severe as compared to LM_{adj} test since in the weakly exogenous regressor case it is not centered at zero for a fixed T.

The over-rejection of the J_{BFK} test in these experiments is caused by a combination of several factors, including heteroskedastic errors and heterogeneous coefficients. In order to distinguish between these effects, we also conducted experiments with homoskedastic errors where we set $Var\left(\varepsilon_{it}\right) = \sigma_i^2 = 1$, for all i, and strictly exogenous regressors (by setting $\alpha_{xi} = 0$ for all i), and consider two cases for the coefficients: heterogeneous and homogeneous (we set $\beta_{i0} = E\left(\beta_{i0}\right) = 0.75$, for all i). The results under homoskedastic errors and homogeneous slopes are summarized in the upper part of Table 3. As to be expected, the J_{BFK} test has good size and power when $T > 20$ and $\alpha = 0$. But the test tends to over-reject when $T = 20$ and N relatively large even under these restrictions. The bottom part of Table 3 presents findings for the experiments with slope heterogeneity, whilst maintaining the assumptions of homoskedastic errors and strictly exogenous regressors. We see that even a small degree of slope heterogeneity can cause the J_{BFK} test to over-reject badly .

Finally, it is important to bear in mind that even the CD_P test is likely to over-reject in the case of models with weakly exogenous regressors if N is much larger than T. Only in the case of models with strictly exogenous regressors, and pure autoregressive models with symmetrically distributed disturbances, we would expect the CD_P test to perform well even if N is much larger than T. To illustrate this property we provide empirical size and power results when $N = 1,000$ and $T = 10$ in Table 4. As can be seen the CD_P test has the correct size when we consider panel data models with strictly exogenous regressors or in the case of pure AR(1) models, which is in contrast to the case of panels with weakly exogenous regressors where the size of the CD_P test is close to 70 percent. It is clear that the small sample properties of the CD_p test for very large N

[17] The rejection rates based on the LM_{Adj} test were above 90 percent for the sample size $N = 500, 1000$ and $T = 10$.

and small T panels very much depends on whether the panel includes weakly exogenous regressors.

7 Application of CCE estimators and CD tests to unbalanced panels

CCE estimators can be readily extended to unbalanced panels, a situation which frequently arises in practice. Denote the set of cross section units with the available data on y_{it} and $\mathbf{x}_{it}$ in period t as $\mathcal{N}_t$ and the number of elements in the set by $\#\mathcal{N}_t$. Initially, we suppose that data coverage for the dependent variables and regressors is the same and later we relax this assumption. The main complication of applying CCE estimator to the case of unbalanced panels is the inclusion of cross-sectional averages in the individual regressions. There are two possibilities regarding the units to include in the computation of cross-sectional averages, either based on the same number of units or based on a varying number of units. In both cases, cross-sectional averages should be constructed using at least a minimum number of units, say $N_{\min}$, which based on the current Monte Carlo evidence suggests the value of $N_{\min} = 20$. If the same units are used, we have

$$\bar{y}_t = \frac{1}{\#\mathcal{N}} \sum_{i \in \mathcal{N}} y_{it}, \text{ and similarly } \bar{\mathbf{x}}_t = \frac{1}{\#\mathcal{N}} \sum_{i \in \mathcal{N}} \mathbf{x}_{it},$$

for $t = \underline{t}, \underline{t}+1, ..., \bar{t}$ where $\mathcal{N} = \bigcap_{t=\underline{t}}^{\bar{t}} \mathcal{N}_t$ and the starting and ending points of the sample $\underline{t}$ and $\bar{t}$ are chosen to maximize the use of data subject to the constraint $\#\mathcal{N} \geq N_{\min}$. The second possibility utilizes data in a more efficient way,

$$\bar{y}_t = \frac{1}{\#\mathcal{N}_t} \sum_{i \in \mathcal{N}_t} y_{it}, \text{ and } \bar{\mathbf{x}}_t = \frac{1}{\#\mathcal{N}_t} \sum_{i \in \mathcal{N}_t} \mathbf{x}_{it},$$

for $t = \underline{t}, \underline{t}+1, ..., \bar{t}$, where $\underline{t}$ and $\bar{t}$ are chosen such that $\#\mathcal{N}_t \geq N_{\min}$ for all $t = \underline{t}, \underline{t}+1, ..., \bar{t}$. Both procedures are likely to perform similarly when $\#\mathcal{N}$ is reasonably large, and the occurrence of missing observations is random. In cases where new cross section units are added to the panel over time and such additions can have systematic influences on the estimation outcomes, it might be advisable to de-mean or de-trend the observations for individual cross section units before computing the cross section averages to be used in the CCE regressions.

Now suppose that the cross section coverage differs for each variable. For example, the dependent variable can be available only for OECD countries, whereas some of the regressors could be available for a larger set of countries. Then, it is preferable to utilize also data on non-OECD countries to maximize the number of units for the computation of CS averages for each of the individual variables.

The CD and LM tests can also be readily extended to unbalanced panels. Denote by $\mathcal{T}_i$, the set of dates over which time series observations on y_{it} and $\mathbf{x}_{it}$ are available for the i^{th} individual, and the number of the elements in the set by $\#\mathcal{T}_i$. For each i compute the OLS residuals based on full set of time series observations for that individual. As before, denote these residuals by $\hat{u}_{it}$, for $t \in \mathcal{T}_i$, and compute the pair-wise correlations of $\hat{u}_{it}$ and $\hat{u}_{jt}$ using the common set of data points in $\mathcal{T}_i \cap \mathcal{T}_j$. Since, the estimated residuals need not sum to zero over the common sample period ρ_{ij} could be estimated by

$$\hat{\rho}_{ij} = \frac{\sum_{t\in\mathcal{T}_i\cap\mathcal{T}_j}\left(\hat{u}_{it}-\bar{\hat{u}}_i\right)\left(\hat{u}_{jt}-\bar{\hat{u}}_j\right)}{\left[\sum_{t\in\mathcal{T}_i\cap\mathcal{T}_j}\left(\hat{u}_{it}-\bar{\hat{u}}_i\right)^2\right]^{1/2}\left[\sum_{t\in\mathcal{T}_i\cap\mathcal{T}_j}\left(\hat{u}_{jt}-\bar{\hat{u}}_j\right)^2\right]^{1/2}},$$

where

$$\bar{\hat{u}}_i = \frac{\sum_{t\in\mathcal{T}_i\cap\mathcal{T}_j}\hat{u}_{it}}{\#\left(\mathcal{T}_i\cap\mathcal{T}_j\right)}.$$

The CD (similarly the LM type) statistics for the unbalanced panel can then be computed as usual by

$$CD_P = \sqrt{\frac{2}{N(N-1)}}\left(\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\sqrt{T_{ij}}\hat{\rho}_{ij}\right), \tag{69}$$

where $T_{ij} = \#\left(\mathcal{T}_i\cap\mathcal{T}_j\right)$ Under the null hypothesis $CD_P \sim N(0,1)$ for $T_i > k+1$, $T_{ij} > 3$, and sufficiently large N.

8 Concluding Remarks

This paper provides a review of the literature on large panel data models with cross-sectional error dependence. The survey focusses on large N and T panel data models where a natural ordering across cross section dimension is not available. This excludes the literature on spatial panel econometrics, which is recently reviewed by Lee and Yu (2010 and 2013). We provide a brief account of the concepts of weak and strong cross-sectional dependence, and discuss the exponent of cross-sectional dependence that characterizes the different degrees of cross-sectional dependence.

We then attempt a synthesis of the literature on estimation and inference in large N and T panel data models with a common factor error structure. We distinguish between strictly and weakly exogenous regressors and panels with homogeneous and heterogeneous slope coefficients. We also provide an overview of tests of error cross-sectional dependence in static and dynamic panel data models.

Table 1: Small sample properties of CCEMG and CCEP estimators of mean slope coefficients in panel data models with weakly and strictly exogenous regressors

	Bias (x100)					RMSE (x100)					Size (x100)					Power (x100)				
(N,T)	**20**	**50**	**100**	**150**	**200**	**20**	**50**	**100**	**150**	**200**	**20**	**50**	**100**	**150**	**200**	**20**	**50**	**100**	**150**	**200**
	Experiments with weakly exogenous regressors																			
	CCEMG																			
40	-5.70	-1.46	-0.29	0.00	0.11	7.82	3.65	2.80	2.67	2.61	23.70	9.35	6.20	6.05	6.25	86.80	94.05	96.00	96.30	96.95
50	-5.84	-1.56	-0.39	0.04	0.11	7.56	3.43	2.56	2.41	2.33	29.50	9.30	7.00	6.70	6.20	93.40	96.75	98.75	98.70	99.20
100	-5.88	-1.50	-0.41	-0.05	0.07	6.82	2.63	1.83	1.70	1.64	46.70	13.10	6.00	5.75	5.25	99.75	99.95	100.00	100.00	100.00
150	-6.11	-1.59	-0.45	-0.11	0.08	6.73	2.36	1.53	1.35	1.30	66.05	16.15	6.60	4.75	4.80	100.00	100.00	100.00	100.00	100.00
200	-6.04	-1.55	-0.43	-0.12	0.01	6.54	2.17	1.37	1.18	1.18	74.65	19.70	7.35	4.50	6.10	100.00	100.00	100.00	100.00	100.00
	CCEP																			
40	-3.50	-0.09	0.76	0.98	1.23	6.58	3.71	3.33	3.24	3.35	14.80	6.75	7.50	7.55	9.85	72.30	78.45	80.55	82.70	82.55
50	-3.55	-0.27	0.70	1.08	1.19	6.07	3.31	2.96	3.00	2.96	14.00	5.70	6.20	8.65	8.80	79.70	86.90	88.55	88.70	90.90
100	-3.56	-0.10	0.76	1.08	1.17	5.11	2.42	2.22	2.27	2.26	21.75	5.50	6.75	9.10	10.45	96.05	97.80	98.80	98.95	99.30
150	-3.78	-0.10	0.74	1.10	1.16	4.86	1.98	1.87	1.99	1.98	30.45	5.85	7.60	11.45	12.60	99.15	99.75	99.95	99.95	100.00
200	-3.66	-0.19	0.80	1.08	1.13	4.56	1.77	1.67	1.78	1.77	35.65	6.25	8.35	12.50	12.45	100.00	100.00	100.00	100.00	100.00
	Experiments with strictly exogenous regressors																			
	CCEMG																			
40	0.19	-0.05	0.02	0.07	0.04	6.43	3.91	3.06	2.91	2.75	6.20	6.40	4.60	6.40	5.55	36.20	74.40	89.95	93.90	95.60
50	-0.02	0.08	0.11	-0.05	-0.02	5.72	3.48	2.83	2.68	2.46	5.25	6.10	5.90	6.75	5.75	43.90	82.20	93.70	96.80	98.05
100	-0.06	0.01	0.02	-0.05	-0.01	4.13	2.52	2.02	1.79	1.78	5.55	6.45	4.90	4.95	6.20	69.95	97.60	99.75	99.95	100.00
150	0.06	0.03	0.00	0.02	0.01	3.29	2.03	1.62	1.50	1.42	5.40	6.00	5.50	5.05	5.30	85.65	99.95	100.00	100.00	100.00
200	-0.06	0.03	-0.02	-0.03	-0.01	2.87	1.75	1.39	1.33	1.23	4.50	5.30	4.85	6.50	5.15	94.10	100.00	100.00	100.00	100.00
	CCEP																			
40	0.21	0.17	0.02	-0.01	-0.02	5.78	3.85	3.16	3.08	2.85	6.40	6.45	5.95	7.10	6.35	74.55	72.90	88.10	92.15	93.50
50	0.03	-0.01	-0.13	0.02	-0.02	5.20	3.48	2.84	2.59	2.54	5.60	6.25	6.25	6.00	5.95	83.35	83.30	94.80	96.30	97.30
100	-0.01	-0.06	0.05	-0.04	0.07	3.67	2.56	2.03	1.89	1.76	5.60	6.15	5.00	5.35	5.65	98.50	97.75	99.85	100.00	100.00
150	0.05	0.02	0.02	0.01	0.01	2.95	2.02	1.65	1.52	1.49	4.50	5.20	5.50	4.95	5.60	99.80	99.95	100.00	100.00	100.00
200	-0.09	-0.04	-0.06	0.03	0.02	2.57	1.74	1.43	1.38	1.28	6.05	5.75	5.15	5.75	4.95	100.00	100.00	100.00	100.00	100.00

Notes: Observations are generated as $y_{it} = c_{yi} + \beta_{0i}x_{it} + \beta_{1i}x_{i,t-1} + u_{it}$, $u_{it} = \gamma_i'\mathbf{f}_t + \varepsilon_{it}$, and $x_{it} = c_{xi} + \alpha_{xi}y_{i,t-1} + \gamma_{xi}'\mathbf{f}_t + v_{it}$, (see (52)-(53)), where $\beta_{0i} \sim IIDU(0.5, 1)$, $\beta_{1i} = -0.5$ for all i, and $m = 3$ (number of unobserved common factors). Fixed effects are generated as $c_{yi} \sim IIDN\,(1, 1)$, and $c_{xi} = c_{yi} + IIDN\,(0, 1)$. In the case of weakly exogenous regressors, $\alpha_{xi} \sim IIDU(0, 1)$ (with $E\,(\alpha_{xi}) = 0.5$), and under the case of strictly exogenous regressors $\alpha_{xi} = 0$ for all i. The errors are generated to be heteroskedastic and weakly cross-sectionally dependent. See Section 5.3 for a more detailed description of the MC design.

Table 2: Size and power of CD_P LM_{adj} and LM_S tests in the case of panels with weakly and strictly exogenous regressors

(nominal size is set at 5%)

	$\alpha=0$					$\alpha=0.25$					$\alpha=0.5$					$\alpha=0.75$				
(N,T)	**20**	**50**	**100**	**150**	**200**	**20**	**50**	**100**	**150**	**200**	**20**	**50**	**100**	**150**	**200**	**20**	**50**	**100**	**150**	**200**
	Experiments with strictly exogenous regressors																			
	CD_P **test**																			
40	5.65	6.25	5.15	5.15	4.95	6.20	6.50	6.20	6.25	6.75	24.60	46.70	74.60	86.35	91.95	99.50	100.00	100.00	100.00	100.00
50	5.40	4.80	5.30	5.15	5.40	5.05	5.15	6.85	7.25	6.50	28.55	52.60	78.85	90.45	96.10	99.70	100.00	100.00	100.00	100.00
100	5.45	5.45	5.40	4.40	5.45	5.10	6.15	6.75	6.60	8.20	32.50	60.15	82.55	92.95	97.45	99.95	100.00	100.00	100.00	100.00
150	4.80	4.75	4.65	4.95	5.05	5.05	5.70	5.85	5.15	6.10	31.90	56.45	83.45	92.80	97.50	100.00	100.00	100.00	100.00	100.00
200	5.85	4.70	5.25	6.60	4.50	6.00	5.80	5.30	5.55	6.40	30.00	57.60	83.65	94.15	97.95	100.00	100.00	100.00	100.00	100.00
	LM_{Adj} **test**																			
40	4.75	5.25	5.50	4.30	5.20	6.80	7.65	15.95	28.30	36.55	43.05	93.35	99.80	100.00	100.00	99.70	100.00	100.00	100.00	100.00
50	6.05	5.25	4.00	4.95	4.95	6.05	6.45	12.40	19.70	31.50	47.85	95.85	100.00	100.00	100.00	99.70	100.00	100.00	100.00	100.00
100	7.00	5.10	4.75	4.70	4.80	7.35	8.80	18.40	34.30	46.25	53.75	98.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
150	6.55	4.85	5.10	5.15	5.40	7.45	6.70	11.95	18.55	28.65	49.85	98.55	99.95	100.00	100.00	100.00	100.00	100.00	100.00	100.00
200	7.75	4.95	5.15	3.90	5.10	8.75	6.45	8.50	13.25	19.00	52.05	98.70	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	LM_S **test**																			
40	11.35	5.30	4.70	5.70	5.30	15.75	10.85	20.10	28.15	41.50	63.35	95.65	99.70	99.95	100.00	99.80	100.00	100.00	100.00	100.00
50	17.65	6.70	5.90	5.40	4.90	18.90	11.40	15.65	21.05	31.45	73.85	97.05	99.95	99.90	100.00	99.95	100.00	100.00	100.00	100.00
100	44.70	9.40	5.80	6.15	6.15	49.70	19.80	24.80	40.00	51.05	88.65	99.20	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
150	67.25	14.30	7.25	7.30	5.45	70.40	25.85	19.35	25.20	35.40	94.15	99.70	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
200	85.10	24.45	9.45	6.55	6.60	85.75	31.10	19.85	22.05	26.05	98.20	99.85	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	Experiments with lagged dependent variable and other weakly exogenous regressors																			
	CD_P **test**																			
40	5.90	4.75	5.55	5.05	4.90	6.00	6.10	5.70	6.40	5.90	26.00	49.75	72.55	84.80	93.10	99.80	100.00	100.00	100.00	100.00
50	6.35	5.00	5.15	5.60	4.40	5.90	6.75	6.00	5.95	6.45	28.90	55.00	77.90	90.50	96.00	99.75	100.00	100.00	100.00	100.00
100	6.55	5.50	5.05	5.10	3.95	6.85	6.85	6.80	6.75	8.15	31.75	59.15	81.30	93.65	97.10	100.00	100.00	100.00	100.00	100.00
150	7.55	5.90	4.50	5.60	4.35	8.30	5.75	6.80	5.25	6.45	34.30	56.15	80.95	94.15	97.00	100.00	100.00	100.00	100.00	100.00
200	8.10	4.75	5.05	5.65	4.60	10.30	6.00	7.25	6.40	6.45	35.75	61.10	83.20	94.05	98.45	100.00	100.00	100.00	100.00	100.00
	LM_{Adj} **test**																			
40	5.05	4.70	5.25	4.10	5.85	6.40	6.85	15.25	26.70	38.65	32.60	92.00	99.80	99.95	100.00	99.40	100.00	100.00	100.00	100.00
50	5.80	4.90	4.70	4.90	4.90	5.30	6.15	12.10	20.80	30.85	35.65	95.55	99.85	100.00	100.00	99.65	100.00	100.00	100.00	100.00
100	6.45	5.60	5.05	4.70	4.80	7.85	7.50	18.45	31.05	47.00	36.40	98.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
150	11.55	5.95	5.30	5.20	3.85	10.35	6.50	10.35	19.65	28.60	31.60	97.65	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
200	14.25	5.50	5.35	4.90	5.25	12.85	6.00	8.20	11.55	19.25	31.55	98.80	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	LM_S **test**																			
40	15.40	6.10	5.40	4.50	5.10	17.00	10.90	19.45	28.35	41.70	61.30	94.85	99.85	99.85	100.00	99.65	100.00	100.00	100.00	100.00
50	18.60	6.55	5.60	4.25	5.05	22.25	10.25	14.25	22.80	31.70	72.80	97.35	99.95	100.00	100.00	99.80	100.00	100.00	100.00	100.00
100	50.25	10.55	6.60	5.10	5.55	55.60	21.65	26.35	37.95	54.40	88.20	99.25	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
150	77.65	18.25	7.70	6.25	6.95	77.70	28.20	18.75	27.70	34.95	95.40	99.55	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
200	89.60	29.55	11.90	6.90	6.30	87.95	36.65	19.95	23.40	25.85	98.70	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

Notes: Observations are generated using the equations $y_{it} = c_{yi} + \lambda_i y_{i,t-1} + \beta_{0i} x_{it} + \beta_{1i} x_{i,t-1} + u_{it}$, $x_{it} = c_{xi} + \alpha_{xi} y_{i,t-1} + \gamma'_{xi}\mathbf{f}_t + v_{it}$, (see (67) and (53), respectively), and $u_{it} = \gamma_i \tilde{f}_t + \varepsilon_{it}$, (see (68)). Four values of $\alpha = 0, 0.25, 0.5$ and 0.75 are considered. Null of weak cross-sectional dependence is characterized by $\alpha = 0$ and $\alpha = 0.25$. In the case of panels with strictly exogenous regressors $\lambda_i = \alpha_{xi} = 0$, for all i. For a more detailed account of the MC design see Section 6. LM_S test statistic is computed using the fixed effects estimates.

Table 3: Size and power of the J_{BFK} test in the case of panel data models with strictly exogenous regressors and homoskedastic idiosyncratic shocks e_{it} (nominal size is set to 5%)

	$\alpha=0$					$\alpha=0.25$					$\alpha=0.5$					$\alpha=0.75$				
(N,T)	**20**	**50**	**100**	**150**	**200**	**20**	**50**	**100**	**150**	**200**	**20**	**50**	**100**	**150**	**200**	**20**	**50**	**100**	**150**	**200**
	Experiments with homogeneous slopes																			
40	7.85	5.60	5.60	5.20	5.90	21.85	53.80	79.40	86.65	92.50	82.70	99.30	100.00	100.00	100.00	99.70	100.00	100.00	100.00	100.00
50	8.90	5.90	6.00	6.10	4.20	17.85	44.90	73.75	83.75	89.10	84.35	99.90	100.00	100.00	100.00	99.85	100.00	100.00	100.00	100.00
100	9.70	6.10	5.65	5.30	5.50	19.35	52.30	81.30	91.90	95.55	88.25	100.00	100.00	100.00	100.00	99.95	100.00	100.00	100.00	100.00
150	15.00	5.90	5.30	5.10	5.60	14.65	39.60	69.80	83.95	91.00	87.95	99.95	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
200	21.30	6.60	5.30	4.60	5.60	15.90	27.45	58.70	75.45	84.55	87.10	99.95	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	Experiments with heterogeneous slopes																			
40	7.30	9.10	13.70	22.10	31.80	22.15	55.15	83.90	93.30	96.45	81.40	99.60	100.00	100.00	100.00	99.75	100.00	100.00	100.00	100.00
50	7.60	8.80	18.20	30.90	40.45	18.65	53.25	80.95	92.45	96.95	85.45	99.85	100.00	100.00	100.00	99.85	100.00	100.00	100.00	100.00
100	9.40	16.85	42.05	65.20	83.10	21.40	65.65	94.80	99.20	99.90	88.75	100.00	100.00	100.00	100.00	99.90	100.00	100.00	100.00	100.00
150	12.65	24.70	60.80	86.25	96.25	17.35	62.30	94.10	99.60	99.95	88.45	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
200	15.20	36.80	78.90	95.65	99.00	16.05	64.85	96.15	99.75	99.90	87.75	99.95	100.00	100.00	100.00	99.95	100.00	100.00	100.00	100.00

Notes: The data generating process is the same as the one used to generate the results in Table 2 with strictly exogenous regressors, but with two exceptions: error variances are assumed homoskedastic ($Var(\varepsilon_{it}) = \sigma_i^2 = 1$, for all i) and two possibilities are considered for the slope coefficients: heterogeneous and homogeneous (in the latter case $\beta_{i0} = E(\beta_{i0}) = 0.75$, for all i). Null of weak cross-sectional dependence is characterized by $\alpha = 0$ and $\alpha = 0.25$. See also the notes to Table 2. The J_{BFK} test statistic is computed using the fixed effects estimates.

Table 4: Size and power of the CD_P test for large N and short T panels with strictly and weakly exogenous regressors (nominal size is set to 5%)

	$\alpha = 0$	$\alpha = 0.25$	$\alpha = 0.5$	$\alpha = 0.75$
(N,T)	**10**	**10**	**10**	**10**
	Panel with strictly exogenous regressors			
1000	5.10	6.30	20.50	99.90
	Pure AR(1) panel			
1000	5.50	6.05	22.10	100.00
	Dynamic panel with weakly exogenous regressors			
1000	69.45	70.70	73.95	100.00

Notes: See the notes to Tables 1 and 2, and Section 6 for further details. In particular, note that null of weak cross-sectional dependence is characterized by $\alpha = 0$ and $\alpha = 0.25$, with alternatives of semi-strong and strong cross-sectional dependence given by values of $\alpha \geq 1/2$.

References

Ahn, S. C. and A. R. Horenstein (2013). Eigenvalue ratio test for the number of factors. *Econometrica 81*(3), 1203–1207.

Amengual, D. and M. W. Watson (2007). Consistent estimation of the number of dynamic factors in a large N and T panel. *Journal of Business and Economic Statistics 25*(1), 91–96.

Andrews, D. (2005). Cross section regression with common shocks. *Econometrica 73*, 1551–1585.

Anselin, L. (1988). *Spatial Econometrics: Methods and Models.* Dordrecht, The Netherlands: Kluwer Academic Publishers.

Anselin, L. (2001). Spatial econometrics. In B. H. Baltagi (Ed.), *A Companion to Theoretical Econometrics*. Blackwell, Oxford.

Anselin, L. and A. K. Bera (1998). Spatial dependence in linear regression models with an introduction to spatial econometrics. In A. Ullah and D. E. A. Giles (Eds.), *Handbook of Applied Economic Statistics*. Marcel Dekker, New York.

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica 77*, 1229–1279.

Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica 70*, 191–221.

Bai, J. and S. Ng (2007). Determining the number of primitive shocks in factor models. *Journal of Business and Economic Statistics 25*(1), 52–60.

Bai, J. and S. Ng (2008). Large dimensional factor analysis. *Foundations and Trends in Econometrics 3*(2), 89–168.

Bai, Z. D. and J. W. Silverstein (1998). No eigenvalues outside the support of the limiting spectral distribution of large dimensional sample covariance matrices. *Annals of Probability 26*(1), 316–345.

Bailey, N., G. Kapetanios, and M. H. Pesaran (2012). Exponents of cross-sectional dependence: Estimation and inference. CESifo Working Paper No. 3722, revised July 2013.

Baltagi, B. H., Q. Feng, and C. Kao (2011). Testing for sphericity in a fixed effects panel data model. *The Econometrics Journal 14*, 25–47.

Baltagi, B. H., S. Song, and W. Koh (2003). Testing panel data regression models with spatial error correlation. *Journal of Econometrics 117*, 123–150.

Berk, K. N. (1974). Consistent autoregressive spectral estimates. *The Annals of Statistics 2*, 489–502.

Breitung, J. and I. Choi (2013). Factor models. In N. Hashimzade and M. A. Thornton (Eds.), *Handbook Of Research Methods And Applications In Empirical Macroeconomics*, Chapter 11. Edward Elgar.

Breitung, J. and U. Pigorsch (2013). A canonical correlation approach for selecting the number of dynamic factors. *Oxford Bulletin of Economics and Statistics 75*(1), 23–36.

Breusch, T. S. and A. R. Pagan (1980). The Lagrange Multiplier test and its application to model specifications in econometrics. *Review of Economic Studies 47*, 239–253.

Chamberlain, G. (1983). Funds, factors and diversification in arbitrage pricing models. *Econometrica 51*, 1305–1324.

Chamberlain, G. (1984). Panel data. In Z. Griliches and M. Intrilligator (Eds.), *Handbook of Econometrics*, Volume 2, Chapter 22, pp. 1247–1318. North-Holland, Amsterdam.

Choi, I. and H. Jeong (2013). Model selection for factor analysis: Some new criteria and performance comparisons. Research Institute for Market Economy (RIME) Working Paper No.1209, Sogang University.

Chudik, A. and M. H. Pesaran (2013a). Aggregation in large dynamic panels. *forthcoming in Journal of Econometrics*.

Chudik, A. and M. H. Pesaran (2013b). Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors. CESifo Working Paper No. 4232.

Chudik, A. and M. H. Pesaran (2013c). Econometric analysis of high dimensional VARs featuring a dominant unit. *Econometric Reviews 32*, 592–649.

Chudik, A., M. H. Pesaran, and E. Tosetti (2011). Weak and strong cross section dependence and estimation of large panels. *The Econometrics Jounal 14*, C45–C90.

Cliff, A. and J. K. Ord (1973). *Spatial Autocorrelation*. Pion, London.

Cliff, A. and J. K. Ord (1981). *Spatial Processes: Models and Applications*. Pion, London.

Coakley, J., A. M. Fuertes, and R. Smith (2002). A principal components approach to cross-section dependence in panels. Birkbeck College Discussion Paper 01/2002.

Coakley, J., A. M. Fuertes, and R. Smith (2006). Unobserved heterogeneity in panel time series. *Computational Statistics and Data Analysis 50*, 2361–2380.

Dhaene, G. and K. Jochmans (2012). Split-panel jackknife estimation of fixed-effect models. Mimeo, 21 July 2012.

Dufour, J. M. and L. Khalaf (2002). Exact tests for contemporaneous correlation of disturbances in seemingly unrelated regressions. *Journal of Econometrics 106*, 143–170.

Everaert, G. and T. D. Groote (2012). Common correlated effects estimation of dynamic panels with cross-sectional dependence. Mimeo, 9 November 2012.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2000). The generalized dynamic factor model: Identification and estimation. *Review of Economics and Statistic 82*, 540–554.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2004). The generalized dynamic factor model: Consistency and rates. *Journal of Econometrics 119*, 231–235.

Forni, M. and M. Lippi (2001). The generalized factor model: Representation theory. *Econometric Theory 17*, 1113–1141.

Frees, E. W. (1995). Assessing cross sectional correlation in panel data. *Journal of Econometrics 69*, 393–414.

Geweke, J. (1977). The dynamic factor analysis of economic time series. In D. Aigner and A. Goldberger (Eds.), *Latent variables in socio-economic models*. Amsterdam: North-Holland.

Gourieroux, C., A. Monfort, E. Renault, and A. Trognon (1987). Generalised residuals. *Journal of Econometrics 34*, 5–32.

Granger, C. W. J. (1980). Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics 14*, 227–238.

Hachem, W., P. Loubaton, and J. Najim (2005). The empirical eigenvalue distribution of a gram matrix: From independence to stationarity. *Markov Processes and Related Fields 11*(4), 629–648.

Haining, R. P. (2003). *Spatial Data Analysis: Theory and Practice*. Cambridge University Press, Cambridge.

Hallin, M. and R. Liska (2007). The generalized dynamic factor model: determining the number of factors. *Journal of the American Statistical Association 102*, 603–617.

Harding, M. (2013). Estimating the number of factors in large dimensional factor models. mimeo, April 2013.

Holly, S., M. H. Pesaran, and T. Yagamata (2011). Spatial and temporal diffusion of house prices in the UK. *Journal of Urban Economics 69*, 2–23.

Hsiao, C., M. H. Pesaran, and A. Pick (2012). Diagnostic tests of cross-section independence for limited dependent variable panel data models. *Oxford Bulletin of Economics and Statistics 74*, 253–277.

Hurwicz, L. (1950). Least squares bias in time series. In T. C. Koopman (Ed.), *Statistical Inference in Dynamic Economic Models*. New York: Wiley.

Jensen, P. S. and T. D. Schmidt (2011). Testing cross-sectional dependence in regional panel data. *Spatial Economic Analysis 6*(4), 423–450.

John, S. (1971). Some optimal multivariate tests. *Biometrika 58*, 123–127.

Kapetanios, G. (2004). A new method for determining the number of factors in factor models with large datasets. Queen Mary University of London, Working Paper No. 525.

Kapetanios, G. (2010). A testing procedure for determining the number of factors in approximate factor models with large datasets. *Journal of Business and Economic Statistics 28*(3), 397–409.

Kapetanios, G. and M. H. Pesaran (2007). Alternative approaches to estimation and inference in large multifactor panels: Small sample results with an application to modelling of asset returns. In G. Phillips and E. Tzavalis (Eds.), *The Refinement of Econometric Estimation and Test Procedures: Finite Sample and Asymptotic Analysis*. Cambridge: Cambridge University Press.

Kapetanios, G., M. H. Pesaran, and T. Yagamata (2011). Panels with nonstationary multifactor error structures. *Journal of Econometrics 160*, 326–348.

Lee, L.-F. and J. Yu (2010). Some recent developments in spatial panel data model. *Regional Science and Urban Economics 40*, 255–271.

Lee, L.-F. and J. Yu (2013). Spatial panel data models. Mimeo, April, 2013.

Lee, N., H. R. Moon, and M. Weidner (2012). Analysis of interactive fixed effects dynamic linear panel regression with measurement error. *Economics Letters 117*(1), 239–242.

Moon, H. R. and M. Weidner (2010). Dynamic linear panel regression models with interactive fixed effects. Mimeo, July 2010.

Moran, P. A. P. (1948). The interpretation of statistical maps. *Biometrika 35*, 255–60.

Moscone, F. and E. Tosetti (2009). A review and comparison of tests of cross section independence in panels. *Journal of Economic Surveys 23*, 528–561.

Ng, S. (2006). Testing cross section correlation in panel data using spacings. *Journal of Business and economic statistics 24*, 12–23.

O'Connell, P. G. J. (1998). The overvaluation of purchasing power parity. *Journal of International Economics 44*, 1–19.

Onatski, A. (2009). Testing hypotheses about the number of factors in large factor models. *Econometrica 77*, 1447–1479.

Onatski, A. (2010). Determining the number of factors from empirical distribution of eigenvalues. *Review of Economics and Statistics 92*, 1004–1016.

Onatski, A. (2012). Asymptotics of the principal components estimator of large factor models with weakly influential factors. *Journal of Econometrics 168*, 244–258.

Pesaran, M. H. (2004). General diagnostic tests for cross section dependence in panels. CESifo Working Paper No. 1229.

Pesaran, M. H. (2006). Estimation and inference in large heterogeneous panels with multifactor error structure. *Econometrica 74*, 967–1012.

Pesaran, M. H. (2013). Testing weak cross-sectional dependence in large panels. *forthcoming in Econometric Reviews*.

Pesaran, M. H. and R. Smith (1995). Estimation of long-run relationships from dynamic heterogeneous panels. *Journal of Econometrics 68*, 79–113.

Pesaran, M. H. and E. Tosetti (2011). Large panels with common factors and spatial correlation. *Journal of Econometrics 161*(2), 182–202.

Pesaran, M. H., A. Ullah, and T. Yamagata (2008). A bias-adjusted LM test of error cross section independence. *The Econometrics Journal 11*, 105–127.

Pesaran, M. H. and T. Yamagata (2008). Testing slope homogeneity in large panels. *Journal of Econometrics 142*, 50–93.

Phillips, P. C. B. and D. Sul (2003). Dynamic panel estimation and homogeneity testing under cross section dependence. *The Econometrics Journal 6*, 217–259.

Phillips, P. C. B. and D. Sul (2007). Bias in dynamic panel estimation with fixed effects, incidental trends and cross section dependence. *Journal of Econometrics 137*, 162–188.

Pyke (1965). Spacings. *Journal of the royal statistical society, Series B 27*, 395–449.

Said, E. and D. A. Dickey (1984). Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika 71*, 599–607.

Sarafidis, V. and D. Robertson (2009). On the impact of error cross-sectional dependence in short dynamic panel estimation. *The Econometrics Journal 12*, 62–81.

Sarafidis, V. and T. Wansbeek (2012). Cross-sectional dependence in panel data analysis. *Econometric Reviews 31*, 483–531.

Sarafidis, V., T. Yagamata, and D. Robertson (2009). A test of cross section dependence for a linear dynamic panel model with regressors. *Journal of Econometrics 148*, 149–161.

Sargan, J. (1988). Testing for misspecification after estimation using instrumental variables. In E. Maasoumi (Ed.), *Contributions to Econometrics: John Denis Sargan*, Volume 1. Cambridge University Press.

Sargent, T. J. and C. A. Sims (1977). Business cycle modeling without pretending to have too much a-priori economic theory. In C. Sims (Ed.), *New methods in business cycle research.* Minneapolis: Federal Reserve Bank of Minneapolis.

Schott, J. R. (2005). Testing for complete independence in high dimensions. *Biometrika 92*, 951–956.

So, B. S. and D. W. Shin (1999). Recursive mean adjustment in time series inferences. *Statistics & Probability Letters 43*, 65–73.

Song, M. (2013). Asymptotic theory for dynamic heterogeneous panels with cross-sectional dependence and its applications. Mimeo, 30 January 2013.

Stock, J. H. and M. W. Watson (2011). Dynamic factor models. In M. P. Clements and D. F. Hendry (Eds.), *The Oxford Handbook of Economic Forecasting*. Oxford University Press.

Westerlund, J. and J. Urbain (2011). Cross-sectional averages or principal components? Research Memoranda 053, Maastricht : METEOR, Maastricht Research School of Economics of Technology and Organization.

Whittle, P. (1954). On stationary processes on the plane. *Biometrika 41*, 434–449.

Yin, Y. Q., Z. D. Bai, and P. R. Krishnainiah (1988). On the limit of the largest eigenvalue of the large dimensional sample covariance matrix. *Probability Theory and Related Fields 78*(4), 509–521.

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.