

Large Panels with Common Factors and Spatial Correlations*

M. Hashem Pesaran
Cambridge University and USC

Elisa Tosetti
Cambridge University

August 2007

Abstract

This paper considers the statistical analysis of large panel data sets where even after conditioning on common observed effects the cross section units might remain dependently distributed. This could arise when the cross section units are subject to unobserved common effects and/or if there are spill over effects due to spatial or other forms of local dependencies. The paper provides an overview of the literature on cross section dependence, introduces the concepts of time-specific weak and strong cross section dependence and shows that the commonly used spatial models are examples of weak cross section dependence. It is then established that the Common Correlated Effects (CCE) estimator of panel data model with a multifactor error structure, recently advanced by Pesaran (2006), continues to provide consistent estimates of the slope coefficient, even in the presence of spatial error processes. Small sample properties of the CCE estimator under various patterns of cross section dependence, including spatial forms, are investigated by Monte Carlo experiments. Results show that the CCE approach works well in the presence of weak and/or strong cross sectionally correlated errors. We also explore the role of certain characteristics of spatial processes in determining the performance of CCE estimators, such as the form and intensity of spatial dependence, and the sparseness of the spatial weight matrix.

Keywords: Panels, Common Correlated Effects, Strong and Weak Cross Section Dependence.

JEL Classification: C10, C31, C33

*We are grateful to Alexander Chudik and George Kapetanios for helpful comments and suggestions.

1 Introduction

Over the past few years there has been a growing literature, both empirical and theoretical, on econometric analysis of panel data models with cross sectionally dependent error processes. Cross correlations can be due to omitted common effects, spatial effects, or could arise as a result of interactions within socio-economic networks. Conditioning on variables specific to the cross section units alone does not deliver cross section error independence; an assumption required by the standard literature on panel data models. In the presence of such dependence, conventional panel estimators such as fixed or random effects can result in misleading inference and even inconsistent estimators. (Phillips and Sul, 2003). Further, conventional panel estimators will be inconsistent if regressors are correlated with the factors behind the error cross section dependence. (Pesaran, 2006). Correlation across units in panels has also serious drawbacks on commonly used panel unit root tests, since several of the existing tests assume independence (Levin, Lin and Chu, 2002; Im, Pesaran and Shin, 2003). As a result, when applied to cross sectional dependent panels, such unit root tests tend to have substantial size distortions (O'Connell, 1998). This potential problem has recently given major impetus to the research on panel unit root tests that allow for cross correlations (see, for example, Bai and Ng, 2004, Moon and Perron, 2004, and Pesaran, 2007). These and other related developments are reviewed in Breitung and Pesaran (2007).

In the case of panel data models where the cross section dimension, N , is small and the time series dimension, T , is large, the standard approach to cross section dependence is to consider the equations from different cross section units as a system of seemingly unrelated regression equations (SURE), and then estimate it by Generalized Least Squares techniques (Zellner, 1962). If the time series dimension is not sufficiently large, and in particular if $N > T$, the SURE approach is not feasible. The approach also fails to yield consistent estimators if the cross dependence is due to an observed common factor which is correlated with the included observed regressors.

Currently, there are two main strands in the literature for dealing with error cross section dependence in panels where N is large relative to T , namely the residual multifactor and the spatial econometric approaches. The multifactor approach assumes that the cross dependence can be characterized by a finite number of unobserved common factors, possibly due to economy-wide shocks that affect all units albeit with different intensities. Under this framework, the error term is a linear combination of few common time-specific effects with heterogeneous factor loadings plus an idiosyncratic (individual-specific) error term. Estimation of a panel with such multifactor residual structure can be addressed by full maximum likelihood procedure (Robertson and Symons, 2000), or by principal component analysis (Coakley, Fuertes and Smith, 2002). A major shortcoming of these techniques is that they are not applicable when the regressors are correlated with the common shocks, as in this case they lead to inconsistent estimators. Recently, Pesaran (2006) has suggested an estimation method, known as Common Correlated Effects (CCE), that consists of approximating the linear combinations of the unobserved factors by cross section averages of the dependent and explanatory variables and then running standard panel regressions augmented with the cross section averages. An advantage of this approach is that it yields consistent estimates also when the regressors are correlated with the factors. Monte Carlo studies have also shown that,

compared to other existing methods, CCE estimation is the most efficient and robust to alternative assumptions various hypothesis of non-stationarity of variables and factors (Coakley, Fuertes and Smith, 2006; Kapetanios and Pesaran, 2007).

The spatial approach assumes that the structure of cross section correlation is related to location and distance among units, defined according to a pre-specified metric. Proximity needs not be measured in terms of physical space, but can be defined using other types of metrics, such as economic (Conley, 1999, Pesaran, Schuermann and Weiner, 2004), policy, or social distance (Conley and Topa, 2002). Hence, cross section correlation is represented by means of a spatial process, which explicitly relates each unit to its neighbours (Whittle, 1954). Spatial econometric literature has suggested a number of processes for modeling spatial dependence. The most widely applied are the Spatial Moving Average (SMA) and the Spatial Error Component (SEC) specifications, and the Spatial Autoregressive (SAR) model. They differ in the range of dependence implied by their covariance matrices. Estimation of panels with spatially correlated errors is generally based on maximum likelihood techniques (Anselin, 1988), or on the generalized method of moments (Kelejian and Prucha, 1999; Conley, 1999). However, most of the work on spatial econometrics has so far been carried out in the context of a single cross section, or under strong cross section homogeneity and time invariance assumptions. While few authors consider heterogeneity and time dependency in spatial panels (Anselin, 1988; Kelejian et al., 2006), the study of local forms of correlation across units in more general framework of heterogeneous panel regressions still remains a challenging topic.

This paper considers estimation of a panel data model with common factors where the idiosyncratic errors could display spatial dependence. Specifically, we consider a panel data model where the error term is the sum of a multifactor structure and a spatial process. In this setting, we show that Pesaran's CCE approach continues to be valid, and under certain standard assumptions on the spatial error process provides consistent estimates of the slope coefficients. To this end, we introduce the concepts of time-specific weak and strong cross section dependence. We define a process to be cross sectionally weakly dependent at a given point in time if its weighted average at that time converges to its expectation in quadratic mean, as the cross section dimension is increased without bounds for all predetermined weights that satisfy certain 'granularity' conditions. If this condition does not hold, then the process is said to be cross sectionally strongly dependent. The distinctive feature of strong correlation is that it is pervasive, in the sense that it remains common to all units in a progressively larger cross sectional sample. Convergence properties of weighted averages is of great importance for the asymptotic theory of various estimators and tests commonly used in econometrics. From a more general perspective, the time-specific notions of weak and strong cross section dependence is also related to a consistent body of research in finance, the theory of asset pricing and portfolio choice (Chamberlain, 1983; Forni and Lippi, 2001).

We also investigate the nature of dependency conveyed by most widely applied processes for modeling contemporaneous correlation. We prove that common factor models provide a class of panel data models with both strong and weak dependence, while commonly used spatial processes are characterizations of weak cross section dependence.

We conclude the paper with a Monte Carlo study investigating the small sample performance of

the CCE estimator under alternative cross section error correlations. We examine the performance of the CCE estimator both when the errors display spatial dependence and when they are subject to unobserved common factors. The mixture case where both sources of cross section dependence exist is also considered. In particular, we show how certain characteristics of spatial processes, such as the spatial intensity and the sparseness of the spatial matrix, affect the performance of the CCE estimator.

The plan of the remainder of the paper is as follows. Section 2 sets out the panel regression model with common factors and spatial dependence and discusses the idea underlying the CCE approach. Section 3 introduces the concepts of strong and weak cross section dependence. Section 4 explore the relationship between the dependence structure of processes and the existence of dominant cross section units. Section 5 presents common factor models as examples of models with strong dependence. Section 6 investigates the nature of cross section dependence implied by most commonly used spatial processes. Section 7 introduces the CCE estimator in the context of panels with common factors and spatial dependence. Section 8 describes the Monte Carlo design and discusses the results. Finally, Section 9 provides some concluding remarks.

2 Panels with common factors and spatial dependence

Let y_{it} be the observation on the i^{th} cross section unit at time t for $i = 1, 2, \dots, N$; $t = 1, 2, \dots, T$, and suppose that it is generated as

$$y_{it} = \boldsymbol{\alpha}'_i \mathbf{d}_t + \boldsymbol{\beta}'_i \mathbf{x}_{it} + \boldsymbol{\gamma}'_i \mathbf{f}_t + e_{it}, \quad (1)$$

where \mathbf{d}_t is a $n \times 1$ vector of observed common effects, \mathbf{x}_{it} is a $k \times 1$ vector of observed individual specific regressors on the i^{th} cross section unit at time t , $\mathbf{f}_t = (f_{1t}, f_{2t}, \dots, f_{mt})'$ is an m -dimensional vector of unobservable variables (known as common shocks/factors), $\boldsymbol{\gamma}_i = (\gamma_{1i}, \gamma_{2i}, \dots, \gamma_{mi})'$ is the associated $m \times 1$ vector of factor loadings, and e_{it} is the unit-specific (idiosyncratic) error term. The common factors, \mathbf{f}_t , are viewed as shocks that simultaneously affect all cross section units, albeit with different degrees.

To model the correlation between the individual specific regressors, \mathbf{x}_{it} , and the common factors we suppose that

$$\mathbf{x}_{it} = \mathbf{A}'_i \mathbf{d}_t + \boldsymbol{\Gamma}'_i \mathbf{f}_t + \mathbf{v}_{it}, \quad (2)$$

where \mathbf{A}_i and $\boldsymbol{\Gamma}_i$ are $n \times k$ and $m \times k$ factor loading matrices with fixed components, and \mathbf{v}_{it} is the individual component of \mathbf{x}_{it} , assumed to be distributed independently of the common factors $\mathbf{g}_t = (\mathbf{d}'_t, \mathbf{f}'_t)'$.

Equations (1) and (2) can be written more compactly as

$$\mathbf{z}_{it} = \begin{pmatrix} y_{it} \\ \mathbf{x}_{it} \end{pmatrix} = \mathbf{B}'_i \mathbf{d}_t + \mathbf{C}'_i \mathbf{f}_t + \mathbf{D}'_i \mathbf{u}_{it}, \quad (3)$$

where

$$\begin{aligned}\mathbf{u}_{it} &= \begin{pmatrix} \mathbf{e}_{it} \\ \mathbf{v}_{it} \end{pmatrix}, \mathbf{D}_i = \begin{pmatrix} 1 & \mathbf{0} \\ \boldsymbol{\beta}_i & \mathbf{I}_k \end{pmatrix}, \\ \mathbf{B}_i &= (\boldsymbol{\alpha}_i \quad \mathbf{A}_i) \mathbf{D}_i, \mathbf{C}_i = (\boldsymbol{\gamma}_i \quad \boldsymbol{\Gamma}_i) \mathbf{D}_i.\end{aligned}$$

The idea underlying the CCE approach is that the unobserved factors can be well approximated by the cross section averages of \mathbf{z}_{it} . Hence, estimation can be carried out by least squares applied to auxiliary regressions where the observed regressors are augmented with these cross section averages plus the observed common factors, \mathbf{d}_t .

As a way of illustrating this result, consider a set of fixed weights $\{w_i\}$ that add up to unity and satisfy the condition $\sum_{i=1}^N w_i^2 = O(N^{-1})$. Let

$$\boldsymbol{\xi}_{it} = \mathbf{D}'_i \mathbf{u}_{it} = \begin{pmatrix} e_{it} + \boldsymbol{\beta}'_i \mathbf{v}_{it} \\ \mathbf{v}_{it} \end{pmatrix}. \quad (4)$$

Then the cross section average of (3), using the weights w_i yields

$$\bar{\mathbf{z}}_{wt} = \bar{\mathbf{B}}'_w \mathbf{d}_t + \bar{\mathbf{C}}'_w \mathbf{f}_t + \bar{\boldsymbol{\xi}}_{wt}, \quad (5)$$

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

If we assume

$$\text{Rank}(\bar{\mathbf{C}}_w) = m \leq k + 1, \text{ for all } N, \quad (6)$$

it follows that

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

Therefore, \mathbf{f}_t can be approximated by linear combinations of $(\bar{\mathbf{z}}_{wt}, \mathbf{d}_t)$, if

$$\bar{\boldsymbol{\xi}}_{wt} \xrightarrow{q.m.} \mathbf{0}. \quad (7)$$

In such a case we obtain

$$\mathbf{f}_t - (\mathbf{C} \mathbf{C}')^{-1} \mathbf{C} [\bar{\mathbf{z}}_{wt} - \bar{\mathbf{B}}'_w \mathbf{d}_t] \xrightarrow{q.m.} 0, \text{ as } N \rightarrow \infty, \quad (8)$$

where

$$\bar{\mathbf{C}}_w \xrightarrow{p} \mathbf{C} = \tilde{\boldsymbol{\Gamma}} \begin{pmatrix} 1 & \mathbf{0} \\ \boldsymbol{\beta} & \mathbf{I}_k \end{pmatrix}, \text{ as } N \rightarrow \infty, \quad (9)$$

$\tilde{\Gamma} = (E(\gamma_i), E(\Gamma_i))$, and $\beta = \mathbf{E}(\beta_i)$.¹ One of the main contributions of this paper is to investigate general conditions under which $\tilde{\xi}_{wt} \xrightarrow{q.m.} \mathbf{0}$, in particular in cases where the individual specific errors, e_{it} , display spatial dependence.

To this end we now introduce some definitions and notations to be used throughout the paper: $\lambda_1(\mathbf{A}) \geq \dots \geq \lambda_n(\mathbf{A})$ are the eigenvalues of a matrix $\mathbf{A} \in \mathbb{M}_n$, where \mathbb{M}_n is the space of real $n \times n$ matrices. \mathbf{A}^- denotes a generalized inverse of \mathbf{A} . \mathbb{R} is the set of real numbers, \mathbb{N} the set of natural numbers, and \mathbb{Z} is the set of integers. The following definitions establish some useful terminologies (see, for example, Horn and Johnson (1985, p.35, pp.290-291)).

Definition 1 Let $\mathbf{A} \in \mathbb{M}_n$. The spectral radius $\rho(\mathbf{A})$ of \mathbf{A} is

$$\rho(\mathbf{A}) \equiv \max_{1 \leq i \leq n} \{|\lambda_i(\mathbf{A})|\}.$$

Definition 2 Let $\mathbf{A} \in \mathbb{M}_n$. The column norm of \mathbf{A} is

$$\|\mathbf{A}\|_c = \max_{1 \leq j \leq n} \sum_{i=1}^n |a_{ij}|.$$

The row norm of \mathbf{A} is

$$\|\mathbf{A}\|_r = \max_{1 \leq i \leq n} \sum_{j=1}^n |a_{ij}|.$$

The Euclidean norm of \mathbf{A} is

$$\|\mathbf{A}\|_2 = [Tr(\mathbf{A}\mathbf{A}')]^{1/2}.$$

Definition 3 Let $\{a_n\}$ be a non-stochastic sequence (Lee, 2002).

- (i) $\{a_n\}$ is bounded (denoted by $O(1)$) if there exist a finite constant K that does not depend on n such that $|a_n| \leq K$ for all n .
- (ii) $\{a_n\}$ is at most of order b_n (denoted by $O(b_n)$) if there exists a constant K independent of n such that $|a_n| \leq Kb_n$ for all n .
- (iii) $\{a_n\}$ is bounded away from zero at a rate b_n if there exist a positive sequence $\{b_n\}$ and a constant $K > 0$ independent of n such that $K \leq \frac{|a_n|}{b_n}$ for n sufficiently large.

We make the following assumptions on the common factors, their loadings, and the individual or unit specific errors:

ASSUMPTION 1: The $(n + m) \times 1$ vector $\mathbf{g}_t = (\mathbf{d}'_t, \mathbf{f}'_t)'$ is a covariance stationary process, with absolute summable autocovariances, distributed independently of $e_{it'}$ and $\mathbf{v}_{it'}$ for all i, t, t' .²

¹Pesaran also shows that the CCE continues to be applicable even if the rank condition, (6), is not satisfied. In this case the mathematical details are much more complicated.

²This assumption can be relaxed to allow for unit roots in the common factors, along the lines shown in Kapetanios, Pesaran and Yagamata (2006).

ASSUMPTION 2: The individual-specific errors e_{it} and $\mathbf{v}_{jt'}$ are distributed independently for all i, j, t and t' , and for each i , \mathbf{v}_{it} follows a linear stationary process with absolute summable autocovariances given by

$$\mathbf{v}_{it} = \sum_{\ell=0}^{\infty} \mathbf{\Pi}_{i\ell} \boldsymbol{\zeta}_{i,t-\ell},$$

where for each i , $\boldsymbol{\zeta}_{it}$ is a $k \times 1$ vector of serially uncorrelated random variables with mean zero, the variance matrix \mathbf{I}_k , and finite fourth-order cumulants. The coefficient matrices $\mathbf{\Pi}_{i\ell}$ satisfy the condition

$$\text{Var}(\mathbf{v}_{it}) = \sum_{\ell=0}^{\infty} \mathbf{\Pi}_{i\ell} \mathbf{\Pi}'_{i\ell} = \boldsymbol{\Sigma}_{vi} \leq \bar{\boldsymbol{\Sigma}}_v < \infty,$$

for all i and some constant matrix $\bar{\boldsymbol{\Sigma}}_v$, where $\boldsymbol{\Sigma}_{vi}$ is a positive definite matrix.

ASSUMPTION 3: The unobserved factor loadings, $\boldsymbol{\gamma}_i$ and $\mathbf{\Gamma}_i$ are bounded, i.e. $\|\boldsymbol{\gamma}_i\|_2 < K$ and $\|\mathbf{\Gamma}_i\|_2 < K$, for all i .

ASSUMPTION 4: The slope coefficients $\boldsymbol{\beta}_i$ follow the random coefficient model

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \mathbf{v}_i, \quad \mathbf{v}_i \sim IID(0, \boldsymbol{\Omega}_v) \quad \text{for } i = 1, \dots, N,$$

where $\|\boldsymbol{\beta}\| < K$, $\|\boldsymbol{\Omega}_v\|_2 < K$, $\boldsymbol{\Omega}_v$ is a symmetric non negative definite matrix, and the random deviations, \mathbf{v}_i , are distributed independently of $\boldsymbol{\gamma}_j$, $\mathbf{\Gamma}_j$, e_{jt} , \mathbf{v}_{jt} , and \mathbf{g}_t for all i, j and t .

Finally, we assume that the individual specific errors e_{it} follow a *Spatial Autoregressive* (SAR) processes (Cliff and Ord, 1973, 1981):

$$e_{it} = \delta \bar{e}_{it} + \varepsilon_{it}, \tag{10}$$

where δ is a scalar parameter, and \bar{e}_{it} is the spatial lag of e_{it} , defined by

$$\bar{e}_{it} = \sum_{j=1}^N s_{ij,t} e_{jt}, \tag{11}$$

and ε_{it} are cross sectionally independently distributed random variables with mean zero and finite variances, σ_i^2 . But for each i the errors ε_{it} can be serially correlated.

Similar spatial models can also be specified for the individual-specific errors of the \mathbf{x}_{it} equation. But to keep the exposition simple we shall not be concerned with this problem explicitly, and assume that \mathbf{v}_{it} is cross sectionally weakly dependent (see below). As we shall see this assumption does allow for the possibility of spatial dependence across the individual specific errors of the \mathbf{x}_{it} equation.

In (11), $s_{ij,t}$ is the generic element of a $N \times N$ matrix \mathbf{S}_t , known as spatial weight matrix, which provides information on the ordering of cross sectional units and the various network or neighborhood linkages at time t (Anselin, 1988, 2002). In matrix form, the spatial model (10) can be written as

$$\mathbf{e}_t = \delta \mathbf{S}_t \mathbf{e}_t + \boldsymbol{\varepsilon}_t, \tag{12}$$

where $\mathbf{e}_t = (e_{1t}, \dots, e_{Nt})'$ and $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \dots, \varepsilon_{Nt})'$. Hence, the error term associated to a cross section unit is modelled as a weighted average of errors corresponding to its neighbouring cross section units, plus a white noise.

In the following, we introduce the notions of *time-specific* weak and strong cross section dependence for a stochastic process, on the basis of the convergence properties of its cross section average, as the number of units in the average is allowed to increase without bound. We investigate the conditions under which the CCE estimator continues to be valid in the presence of spatial dependence such as the SAR process defined by (12).

3 Weak and strong cross section dependence

In this section, we study the structure of correlation of a double index process $\{z_{it}, i \in \mathcal{S}, t \in \mathcal{T}\}$ where $\mathcal{S} \subseteq \mathbb{N}$, $\mathcal{T} \subseteq \mathbb{Z}$, and z_{it} are random variables defined on a probability space (Ω, \mathcal{F}, P) . In this paper the index t refers to an ordered set, the time, while the index i indicates the units of an unordered population. Our primary focus is on characterizing the correlation pattern of the double index process $\{z_{it}\}$ over the cross sectional dimension. To this end, we make the following assumptions:

ASSUMPTION 5: Let $\mathbf{w}_t = (w_{1t}, \dots, w_{Nt})'$ be a vector of pre-determined weights and assume that the following ‘granularity’ conditions hold for all $t \in \mathcal{T}$:

$$\|\mathbf{w}_t\|_2 = O\left(N^{-\frac{1}{2}}\right) \quad (13)$$

$$\frac{w_{jt}}{\|\mathbf{w}_t\|_2} = O\left(N^{-\frac{1}{2}}\right) \quad \text{for any } j \leq N \quad (14)$$

Also, although not necessary for our analysis, it is convenient to assume that the following normalization condition holds

$$\sum_{j=1}^N w_{jt} = 1. \quad (15)$$

ASSUMPTION 6: Let \mathcal{I}_t be the information set at time t containing at least $\mathbf{z}_t, \mathbf{z}_{t-1}, \dots$ and $\mathbf{w}_t, \mathbf{w}_{t-1}, \dots$, where $\mathbf{z}_t = (z_{1t}, \dots, z_{Nt})'$ and $\mathbf{w}_t = (w_{1t}, \dots, w_{Nt})'$. For each $t \in \mathcal{T}$, \mathbf{z}_t has conditional mean and variance

$$E(\mathbf{z}_t | \mathcal{I}_{t-1}) = \mathbf{0}, \quad (16)$$

$$Var(\mathbf{z}_t | \mathcal{I}_{t-1}) = \boldsymbol{\Sigma}_t, \quad (17)$$

where $\boldsymbol{\Sigma}_t$ is a $N \times N$ symmetric, nonnegative definite matrix, with generic $(i, j)^{th}$ element $\sigma_{ij,t}$, and such that $0 < \sigma_{ii,t} \leq K$, for $i = 1, \dots, N$, where K is a finite constant independent of N .

Assumption 5, known in finance as the granularity condition, ensures that the weights $\{w_{it}\}$ is not dominated by a few of the cross section units. In Assumption 6 we impose some regularity conditions on the time series properties of $\{z_{it}\}$. Assumption 6 is also standard in finance and specifies that \mathbf{z}_t has conditional means and variances. The first part, (16), can be relaxed to

$E(\mathbf{z}_t | \mathcal{I}_{t-1}) = \boldsymbol{\mu}_{t-1}$, with $\boldsymbol{\mu}_{t-1}$ being a pre-determined function of the elements of \mathcal{I}_{t-1} . But to keep the exposition simple and without loss of generality we have set $\boldsymbol{\mu}_{t-1} = \mathbf{0}$.

Consider now the weighted averages, $\bar{z}_{wt} = \sum_{i=1}^N w_{i,t-1} z_{it} = \mathbf{w}'_{t-1} \mathbf{z}_t$, for $t \in \mathcal{T}$, where \mathbf{z}_t and \mathbf{w}_{t-1} satisfy Assumptions 5 and 6. We are interested in the limiting behaviour of $\mathbf{w}'_{t-1} \mathbf{z}_t$ for each $t \in \mathcal{T}$ as $N \rightarrow \infty$.

Definition 4 (*weak 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 all weight vectors, \mathbf{w}_{t-1} , satisfying the granularity conditions (13)-(14) we have*

$$\lim_{N \rightarrow \infty} \text{Var}(\mathbf{w}'_{t-1} \mathbf{z}_t | \mathcal{I}_{t-1}) = 0. \quad (18)$$

If (18) holds for all $t \in \mathcal{T}$, then we say that the process is cross sectionally weakly dependent.

Proposition 5 *The process $\{z_{it}\}$ is CWD at a point in time $t \in \mathcal{T}$ if $\lambda_1(\boldsymbol{\Sigma}_t)$ is bounded.*

Proof. We note that by the Rayleigh-Ritz theorem³

$$\text{Var}(\mathbf{w}'_{t-1} \mathbf{z}_t | \mathcal{I}_{t-1}) = \mathbf{w}'_{t-1} \boldsymbol{\Sigma}_t \mathbf{w}_{t-1} \leq (\mathbf{w}'_{t-1} \mathbf{w}_{t-1}) \lambda_1(\boldsymbol{\Sigma}_t). \quad (19)$$

Since $\lambda_1(\boldsymbol{\Sigma}_t)$ is bounded for all N , then under the granularity conditions (13)-(14) we have

$$\lim_{N \rightarrow \infty} \text{Var}(\mathbf{w}'_{t-1} \mathbf{z}_t | \mathcal{I}_{t-1}) = 0.$$

■

A particular form of a CWD process arises when pairwise correlations take non-zero values only across finite subsets of units that do not spread widely as sample size increases. A similar case occurs in spatial processes, where for example local dependency exists only among adjacent observations. In Section 6 we will characterize the cross correlations implied by most commonly used spatial processes.

A number of remarks concerning the above concept of CWD is in order.

Remark 6 *The notion of weak dependence does not necessarily involve an ordering of the observations or the specification of a distance metric.*

Remark 7 *It is clear from (19) that if $\lambda_1(\boldsymbol{\Sigma}_t) = O(N^{1-\epsilon})$ for any $\epsilon \in (0, 1)$, then*

$$\lim_{N \rightarrow \infty} (\mathbf{w}'_{t-1} \mathbf{w}_{t-1}) \lambda_1(\boldsymbol{\Sigma}_t) = 0,$$

and the underlying process will be CWD. Hence, the bounded eigenvalue condition discussed in the literature is sufficient but not necessary for CWD. According to our definition a process could be CWD even if its maximum eigenvalue is rising with N , so long as its rate of increase is bounded appropriately.

³See Horn and Johnson (1985, p.176).

Remark 8 *The concept of “idiosyncratic” developed in Forni and Lippi (2001) can be derived from our notion of CWD as a special case. Forni and Lippi (2001) define as “idiosyncratic” any double sequence, $\{z_{it}\}$, if its weighted average computed both over cross section units and across time converges to zero in quadratic mean, for all sets of weights satisfying the granularity condition (13). They show that the boundedness of the largest eigenvalue of the spectral density matrix of $\{z_{it}\}$ (at all frequencies) is necessary and sufficient for the process to be idiosyncratic. To ensure the existence of the spectral density, Forni and Lippi assume that the underlying time series processes are stationary with absolutely summable autocovariances. In contrast, in our development of the CWD, we consider the asymptotic behaviour of the weighted averages at each point in time, which does not require any regularity assumptions on the time series properties of the underlying processes beyond those set out in Assumption 6.*

Remark 9 *Also in establishing the equivalence of the bounded eigenvalue conditions and the idiosyncratic property, Forni and Lippi consider weights that are not pre-determined, and in principle could depend on the elements of Σ (taken by Forni and Lippi to be time invariant). By focusing on pre-determined weights, as we shall see below, we are able to provide a more natural link to the literatures on panel estimation and inference and the portfolio decision theory.*

If condition (18) is not satisfied, then the incidence of non-zero pair-wise correlations across the elements of Σ_t would be pervasive. More specifically, we say the process $\{z_{it}\}$ is cross sectionally strongly dependent (CSD) at a point in time $t \in T$, if there exists a vector of weights, \mathbf{w}_{t-1} , that satisfy (13)-(14), and a constant K independent of N such that for N sufficiently large

$$\text{Var}(\mathbf{w}'_{t-1}\mathbf{z}_t | \mathcal{I}_{t-1}) \geq K > 0. \quad (20)$$

Proposition 10 *Let $\{z_{it}\}$ be cross sectionally strongly dependent at a point in time $t \in T$. Then $\lambda_1(\Sigma_t)$ and $\|\Sigma_t\|_c$ are bounded away from zero at rate $(\mathbf{w}'_{t-1}\mathbf{w}_{t-1})^{-1}$.*

Proof. For a strongly dependent process, under the granularity conditions (13)-(14) we can find a set of weights \mathbf{w}_{t-1} such that

$$0 < K \leq \text{Var}(\mathbf{w}'_{t-1}\mathbf{z}_t | \mathcal{I}_{t-1}) \leq (\mathbf{w}'_{t-1}\mathbf{w}_{t-1}) \lambda_1(\Sigma_t), \quad (21)$$

which proves that $\lambda_1(\Sigma_t)$ diverges to infinity at least at rate $(\mathbf{w}'_{t-1}\mathbf{w}_{t-1})^{-1}$. Further, we know that⁴

$$\lambda_1(\Sigma_t) \leq \|\Sigma_t\|_c.$$

It follows from (21) that

$$0 < K \leq (\mathbf{w}'_{t-1}\mathbf{w}_{t-1}) \lambda_1(\Sigma_t) \leq (\mathbf{w}'_{t-1}\mathbf{w}_{t-1}) \|\Sigma_t\|_c,$$

which proves that $\|\Sigma_t\|_c$ must tend to infinity at least at rate $(\mathbf{w}'_{t-1}\mathbf{w}_{t-1})^{-1}$. ■

⁴See Horn and Johnstone (1985, pp. 297-298).

From the above proposition, both the spectral radius and the column norm of the covariance matrix of a CSD process are unbounded in N . This result for a CSD process is similar to the condition of not absolutely summable autocorrelations that characterizes time series processes with strong autocorrelation (Robinson, 2003). We observe that for any process $\{z_{it}\}$ either weak or strongly correlated, $\lambda_1(\mathbf{\Sigma}_t)$ and $\|\mathbf{\Sigma}_t\|_c$ cannot diverge to infinity at a rate larger than N . Indeed, since

$$\lambda_1(\mathbf{\Sigma}_t) \leq \sum_{i=1}^N \sigma_{ii,t}, \text{ and } \|\mathbf{\Sigma}_t\|_c \leq \left(\max_{1 \leq j \leq N} \sigma_{jj,t}^{1/2} \right) \sum_{i=1}^N \sigma_{ii,t}^{1/2},$$

and given that by Assumption 6 $\sigma_{ii,t}$, the diagonal elements of $\mathbf{\Sigma}_t$, are finite, it follows that $\lambda_1(\mathbf{\Sigma}_t)$ and $\|\mathbf{\Sigma}_t\|_c$ approach infinity at most at rate N .

Remark 11 *The above concepts of weak and strong cross section dependence are also related to the notion of diversifiability provided by the asset pricing theory (Chamberlain, 1983). In this context, $\mathbf{\Sigma}_t$ represents the covariance matrix of a vector of random returns on N different assets, and $w_{i,t-1}$, for $i = 1, 2, \dots, N$, denote the proportion of investor's wealth allocated to the i^{th} asset. From Definition 4 it follows that the part of asset returns that is weakly dependent will be fully diversified by portfolios constructed using \mathbf{w}_{t-1} as the portfolio weights, and as $N \rightarrow \infty$. The part of asset returns that is strongly dependent can only be diversified with portfolio weights that are dependent on the factor loadings. Indeed, assume for simplicity that returns $\{r_{it}\}$ have the one-factor structure*

$$r_{it} = \gamma_i f_t + e_{it},$$

where $f_t \sim \text{IID}(0, 1)$, and $\{e_{it}\}$ is a CWD process distributed independently of f_t and γ_i . Consider the portfolio weights

$$w_i^* = \frac{1}{N} \left[1 - \frac{\mu_\gamma}{s_\gamma^2} (\gamma_i - \bar{\gamma}) \right],$$

where

$$\mu_\gamma = \frac{1}{N} \sum_{i=1}^N \gamma_i, \quad s_\gamma^2 = N^{-1} \sum_{i=1}^N (\gamma_i - \mu_\gamma)^2 > 0.$$

It is then easily verified that these weights satisfy the granularity conditions, (13)-(14), and since $\{e_{it}\}$ is CWD then

$$\lim_{N \rightarrow \infty} \text{Var} \left(\sum_{i=1}^N w_i^* r_{it} | f_t \right) = f_t^2 \lim_{N \rightarrow \infty} \left(\sum_{i=1}^N w_i^* \gamma_i \right)^2.$$

However

$$\sum_{i=1}^N w_i^* \gamma_i = 0, \text{ for all } N,$$

and the variance of the portfolio return, $\rho_{w^*t} = \sum_{i=1}^N w_i^* r_{it}$, tends to zero as $N \rightarrow \infty$. Note that for all $\mu_\gamma \neq 0$, the underlying return $\{r_{it}\}$ is a CSD process.

We conclude this section with two results concerning the relationship between strongly correlated and weakly correlated variables. Following Definition 4, we say that two processes $\{z_{it,a}\}$ and $\{z_{it,b}\}$ are weakly correlated if $\lim_{N \rightarrow \infty} E(\bar{z}_{wt,a} \bar{z}_{wt,b} | \mathcal{I}_{t-1}) = 0$, for all sets of weights at time t that satisfy the granularity conditions. The next proposition considers correlation of two processes

with different cross dependence structures. We then investigate the correlation structure of linear combinations of strongly correlated and weakly correlated variables.

Proposition 12 *Suppose that $\{z_{it,a}\}$ and $\{z_{it,b}\}$ are CSD and CWD processes, respectively. Then for all sets of weights $\{w_{i,t-1}^a\}$ and $\{w_{i,t-1}^b\}$ satisfying conditions (13)-(14) we have*

$$\lim_{N \rightarrow \infty} E(\bar{z}_{wt,a} \bar{z}_{wt,b} | \mathcal{I}_{t-1}) = 0.$$

Proof. Let $\{w_{i,t-1}^a\}$ and $\{w_{i,t-1}^b\}$ be two sets of weights satisfying conditions (13)-(14). For $t \in \mathcal{T}$, we have

$$[E(\bar{z}_{wt,a} \bar{z}_{wt,b} | \mathcal{I}_{t-1})]^2 \leq E(\bar{z}_{wt,a}^2 | \mathcal{I}_{t-1}) E(\bar{z}_{wt,b}^2 | \mathcal{I}_{t-1}).$$

Since $z_{it,a}$ is a CSD process then

$$\lim_{N \rightarrow \infty} E(\bar{z}_{wt,a}^2 | \mathcal{I}_{t-1}) > K > 0,$$

where K is a finite constant. Also from (18), and considering that $z_{it,b}$ is a CWD process we have

$$\lim_{N \rightarrow \infty} E(\bar{z}_{wt,b}^2 | \mathcal{I}_{t-1}) = 0.$$

Therefore, for all sets of weights satisfying (13)-(14), we obtain

$$\lim_{N \rightarrow \infty} E(\bar{z}_{wt,a} \bar{z}_{wt,b} | \mathcal{I}_{t-1}) = 0.$$

■

Proposition 13 *Consider two independent processes $\{z_{it,a}\}$ and $\{z_{it,b}\}$, and their linear combinations defined by*

$$z_{it,c} = \beta_a z_{it,a} + \beta_b z_{it,b}, \quad (22)$$

where β_a and β_b are non-zero fixed coefficients. Then the following statements hold:

- (i) *Suppose $\{z_{it,a}\}$ and $\{z_{it,b}\}$ are CSD, then $\{z_{it,c}\}$ is CSD,*
- (ii) *Suppose $\{z_{it,a}\}$ and $\{z_{it,b}\}$ are CWD, then $\{z_{it,c}\}$ is CWD,*
- (iii) *Suppose $\{z_{it,a}\}$ is CSD and $\{z_{it,b}\}$ is CWD, then $\{z_{it,c}\}$ is CSD.*

Proof. Let $\Sigma_{t,a}$ and $\Sigma_{t,b}$ be the covariance matrices of $z_{t,a} = (z_{1t,a}, \dots, z_{Nt,a})'$ and $z_{t,b} = (z_{1t,b}, \dots, z_{Nt,b})'$, and $\Sigma_{t,c}$ the covariance of their linear combination that is, given the assumption of independence between $z_{t,a}$ and $z_{t,b}$

$$\Sigma_{t,c} = \beta_a^2 \Sigma_{t,a} + \beta_b^2 \Sigma_{t,b}.$$

The variance of the weighted average $\mathbf{w}'_{t-1} \mathbf{z}_{t,c}$ satisfies

$$\text{Var}(\mathbf{w}'_{t-1} \mathbf{z}_{t,c} | \mathcal{I}_{t-1}) \geq \beta_j^2 \text{Var}(\mathbf{w}'_{t-1} \mathbf{z}_{t,j} | \mathcal{I}_{t-1}), \quad j = a, b,$$

which implies that, if there exists a weights vector \mathbf{w}_{t-1} satisfying the granularity conditions such that either $Var(\mathbf{w}'_{t-1}\mathbf{z}_{t,a}|\mathcal{I}_{t-1})$ or $Var(\mathbf{w}'_{t-1}\mathbf{z}_{t,b}|\mathcal{I}_{t-1})$ or both are nonzero, then also $Var(\mathbf{w}'_{t-1}\mathbf{z}_{t,c}|\mathcal{I}_{t-1})$ is bounded away from zero and $\{z_{it,c}\}$ is cross sectionally strongly dependent (this proves (i) and (iii)). Also, by Weyl's theorem⁵, the largest eigenvalue of $\Sigma_{t,c}$ satisfies

$$\lambda_1(\Sigma_{t,c}) \leq \beta_a^2 \lambda_1(\Sigma_{t,a}) + \beta_b^2 \lambda_1(\Sigma_{t,b}),$$

which establishes that if both $\lambda_1(\Sigma_{t,a})$ and $\lambda_1(\Sigma_{t,b})$ are bounded, then also $\lambda_1(\Sigma_{t,c})$ is bounded in N , and hence $\{z_{it,c}\}$ is cross sectionally weakly correlated (this proves (ii)). ■

The above result can be generalized to linear functions of more than two processes. In general, linear combinations of independent processes that are strongly (weakly) correlated is strongly (weakly) dependent, while linear combinations of a finite number of weakly and strongly correlated processes is strongly correlated, since on aggregation only terms involving the strong component will be of any relevance. This result will be employed in Section 5.

4 Strong dependence and dominant effects

In this section we introduce the notion of dominant effects in the covariance matrix of strongly dependent processes. We then explore the relationship between dominant units and the asymptotic behaviour of eigenvalues of the underlying covariance matrix. The discussion will introduce us to multifactor models which will be broadly investigated in Section 5.

Definition 14 Let $\mathbf{A} = (a_{ij}) \in \mathbb{M}_n$ with finite elements, a_{ij} . We say that \mathbf{A} has m dominant columns if there exists a m -dimensional subset $\mathcal{J}(m)$ of the index set $\mathcal{J} = \{1, 2, \dots, n\}$ such that

$$(i) \sum_{i=1}^N |a_{ij}| = O(N), \text{ for } j \in \mathcal{J}(m)$$

$$(ii) \sum_{i=1}^N |a_{ij}| = O(1) \text{ for } j \in \mathcal{J} \cap \mathcal{J}(m)^c.$$

It also follows that matrix $\mathbf{A} \in \mathbb{M}_n$ will have m dominant rows if \mathbf{A}' has m dominant columns according to the above definition.

Clearly, if the conditional variance Σ_t of a process $\{z_{it}\}$ has m dominant columns, it will also have m dominant rows. In this case, the units corresponding to the unbounded columns and rows have a strong relationship with all other units. This situation arises in a number of empirical applications, in particular when it is possible to identify few dominant individuals (regions), the leaders, that influence many others, the followers. For example, in the context of global macroeconomic modelling, few large economies, such as US, strongly affect small economies, yet impact of small countries on the rest of the world as a whole is negligible (Pesaran and Chudik, 2007). In the study of price competition among firms, it is possible to consider not only localized forms of rivalry, where firms compete only with their close neighbours, but also a global or monopolistic competition, in which all products compete with one or few products of a particular brand (Pinkse et al., 2002). A

⁵See Horn and Johnson (1985, p. 181).

further example can be taken from the housing market in a country such as England, where house price variations in few major urban centres like London have an impact on the dynamics of prices in all other smaller and medium-sized cities in the country.

In Proposition 10 we have proved that under CSD the largest eigenvalue of the covariance matrix diverges to infinity at the rate of N . Further, we have seen that this implies that Σ_t has (at least) one dominant cross section unit. Under some general conditions the reverse is also true, that is if Σ_t has unbounded column norm, then the largest eigenvalue of Σ_t is also unbounded in N . Since Σ_t is a symmetric positive semi-definite matrix it can be decomposed as $\Sigma_t = \mathbf{R}_t \mathbf{R}_t'$, where \mathbf{R}_t is an $N \times r$ ($r \leq N$) matrix with bounded row norm. The following inequalities hold⁶

$$\lambda_1(\Sigma_t) \leq \|\Sigma_t\|_c \leq \|\mathbf{R}_t\|_c \|\mathbf{R}_t\|_r, \quad (23)$$

$$\frac{\|\mathbf{R}_t\|_c}{\sqrt{N}} \leq [\lambda_1(\Sigma_t)]^{1/2}. \quad (24)$$

From (24), if $\|\mathbf{R}_t\|_c$ is bounded away from zero at a rate greater than \sqrt{N} , then $\lambda_1(\Sigma_t)$ also diverges to infinity. Note that if the largest eigenvalue is bounded, the column norm of \mathbf{R}_t (and hence of Σ_t) can still diverge to infinity at a rate at most equal to \sqrt{N} . As we will see in Section 6 below, this result has a nice intuitive interpretation when dealing with spatial processes, where the \mathbf{R}_t matrix is a function of the spatial weight matrix.

We now consider the case where a strongly dependent process has more than one characteristic root that increases with N . As we will see in the following section, this is the case of common factor models (Chamberlain and Rothschild, 1983, Forni and Lippi, 2001). The following theorem holds.

Theorem 15 *Consider a process $\{z_{it}\}$ which is cross sectionally strongly dependent at time $t \in \mathcal{T}$. Let m be a fixed number. Then the following statements hold:*

- (i) *Suppose $\lambda_m(\Sigma_t) = O(N)$, then the process underlying Σ_t has at least m dominant cross section units*
- (ii) *Suppose Σ_t has m dominant cross section units, then $\lambda_j(\Sigma_t) = O(N)$, for $1 \leq j \leq m$, and $\lambda_{m+1}(\Sigma_t)$ is bounded.*

Proof. Suppose for simplicity that the columns and rows of Σ_t are arranged so that first column vector has the largest absolute sum norm, the second column vector has the second largest absolute sum norm, and so on. Call Σ_t^1 the matrix obtained from Σ_t by deleting its first row and first column. Let Σ_t^2 be the the matrix obtained from Σ_t^1 by deleting its first row and first column, Σ_t^3 that obtained dropping the first row and column of Σ_t^2 and so forth, to obtain the set of interlacing matrices $\Sigma_t^1, \dots, \Sigma_t^m$. From the interlacing eigenvalues theorem (see Horn and Johnson, 1985; p. 186) we know that

$$\lambda_m(\Sigma_t^1) \leq \lambda_m(\Sigma_t) \leq \lambda_{m-1}(\Sigma_t^1) \leq \dots \leq \lambda_2(\Sigma_t) \leq \lambda_1(\Sigma_t^1) \leq \lambda_1(\Sigma_t)$$

⁶For (23) see Bernstein (2005, p.351), and for (24) see Bernstein (2005, p.368, eq. xiv).

From the above relation, the following inequalities can be established

$$\lambda_m(\boldsymbol{\Sigma}_t) \leq \lambda_{m-1}(\boldsymbol{\Sigma}_t^1) \leq \lambda_{m-2}(\boldsymbol{\Sigma}_t^2) \leq \dots \leq \lambda_2(\boldsymbol{\Sigma}_t^{m-2}) \leq \lambda_1(\boldsymbol{\Sigma}_t^{m-1}) \quad (25)$$

$$\lambda_{m+1}(\boldsymbol{\Sigma}_t) \leq \lambda_m(\boldsymbol{\Sigma}_t^1) \leq \lambda_{m-1}(\boldsymbol{\Sigma}_t^2) \leq \dots \leq \lambda_2(\boldsymbol{\Sigma}_t^{m-1}) \leq \lambda_1(\boldsymbol{\Sigma}_t^m) \quad (26)$$

$$\lambda_1(\boldsymbol{\Sigma}_t^{m-1}) \leq \dots \leq \lambda_1(\boldsymbol{\Sigma}_t^2) \leq \lambda_1(\boldsymbol{\Sigma}_t^1) \leq \lambda_1(\boldsymbol{\Sigma}_t) \quad (27)$$

Also, recall that

$$\lambda_1(\boldsymbol{\Sigma}_t^j) \leq \left\| \boldsymbol{\Sigma}_t^j \right\|_c \quad (28)$$

Suppose first that $\lambda_m(\boldsymbol{\Sigma}_t) = O(N)$. Then from (25) $\lambda_1(\boldsymbol{\Sigma}_t^{m-1})$ is also unbounded, implying (from (28)) that $\left\| \boldsymbol{\Sigma}_t^{m-1} \right\|_c = O(N)$. Hence, $\boldsymbol{\Sigma}_t$ has at least m dominant units, which proves (i). *Vice versa*, suppose that $\boldsymbol{\Sigma}_t$ has m dominant units. Then we know from (24) that $\boldsymbol{\Sigma}_t$ has at least one eigenvalue unbounded in N . Further, note that, by the definition of dominant units, $\left\| \boldsymbol{\Sigma}_t^m \right\|_c$ is bounded, and hence, from (28), $\lambda_1(\boldsymbol{\Sigma}_t^m)$ is also bounded. From (26) it follows that $\lambda_{m+1}(\boldsymbol{\Sigma}_t)$ is bounded, which proves (ii). ■

Note that in several cases the number of column vectors in $\boldsymbol{\Sigma}_t$ having unbounded sums largely exceeds the number of unbounded characteristic roots. In the extreme case, $\boldsymbol{\Sigma}_t$ could have N dominant units, with only one eigenvalues exploding to infinity, as in the following example of equicorrelation

$$\boldsymbol{\Sigma}_t = \sigma^2 \begin{pmatrix} 1 & \theta & \dots & \theta & \theta \\ \theta & 1 & \dots & \theta & \theta \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \theta & \theta & \dots & 1 & \theta \\ \theta & \theta & \dots & \theta & 1 \end{pmatrix}, \quad (29)$$

where $|\theta| < 1$. In this case all column sums are unbounded. However, the characteristic roots of the above matrix are

$$\begin{aligned} \lambda_1(\boldsymbol{\Sigma}_t) &= \sigma^2 [1 + (N - 1)\theta] \\ \lambda_j(\boldsymbol{\Sigma}_t) &= \sigma^2 (1 - \theta) \quad \text{for } j = 2, \dots, N, \end{aligned}$$

namely only the largest eigenvalue is unbounded in N . In the next section we will see that processes with a covariance matrix like (29) can be well represented by the means of common factor models.

5 Common factor models

Originally proposed in the psychometric literature (Spearman, 1904), factor models are extensively used in macroeconomics and finance to represent the evolution of large cross sectional samples with strong co-movements. Panels with common factors have been applied to characterize the dynamic of stock and bond returns (Chamberlain and Rothschild 1983; Connor and Korajczyk, 1993; Kapetanios and Pesaran, 2007), and in macroeconomics to summarize the empirical content

of a large number of variables by a small set of factors (Stone, 1947; Forni and Reichlin, 1998; Gregory and Head, 1999; Stock and Watson, 2002).

Factor models are typically used in the literature to represent CSD processes. However, according to our definition there are special cases where the factor representation could yield a CWD process. To see this, consider the following multifactor process for $\{z_{it}\}$

$$z_{it} = u_{it} + e_{it}, \quad i = 1, \dots, N, \quad (30)$$

$$u_{it} = \gamma_{i1}f_{1t} + \gamma_{i2}f_{2t} + \dots + \gamma_{im}f_{mt}, \quad (31)$$

where the number of factors, m , is assumed to be smaller than the number of cross sectional units, N . The factors, $f_{\ell t}$, and their loadings, $\gamma_{i\ell}$, satisfy Assumptions 1 and 3, respectively. Without loss of generality we assume that $E(f_{\ell t}^2 | \mathcal{I}_{t-1}) = 1$ and $E(f_{\ell t}f_{pt} | \mathcal{I}_{t-1}) = 0$, for $\ell \neq p = 1, 2, \dots, m$. We also make the following standard assumption to distinguish the common factors from the individual-specific component, $\mathbf{e}_t = (e_{1t}, e_{2t}, \dots, e_{Nt})'$, with $Var(\mathbf{e}_t | \mathcal{I}_{t-1}) = \mathbf{\Sigma}_{et}$.

ASSUMPTION 7: For a given $t \in \mathcal{T}$, $\lambda_1(\mathbf{\Sigma}_{et})$ is bounded, namely $\{e_{it}\}$ is a CWD process.

Theorem 16 Consider the factor model (30)-(31), and suppose that assumption 1, 3 and 7 hold. The process $\{z_{it}\}$ is cross sectionally weakly dependent at a given point in time $t \in \mathcal{T}$, if and only if $\lim_{N \rightarrow \infty} \bar{\gamma}_{w_{t-1}, \ell} = 0$, for $\ell = 1, \dots, m$, and for all weights $\{w_{i,t-1}\}$ satisfying (13)-(14).

Proof. From Definition 1, z_{it} is CWD for at a given point in time $t \in \mathcal{T}$, if

$$\lim_{N \rightarrow \infty} Var(\bar{z}_{wt} | \mathcal{I}_{t-1}) = \lim_{N \rightarrow \infty} [Var(\bar{u}_{wt} | \mathcal{I}_{t-1}) + Var(\bar{e}_{wt} | \mathcal{I}_{t-1})] = 0.$$

Since e_{it} is CWD it immediately follows that $\lim_{N \rightarrow \infty} Var(\bar{e}_{wt} | \mathcal{I}_{t-1}) = 0$. Consider now the first component of z_{it} and note that under our assumptions

$$\begin{aligned} \lim_{N \rightarrow \infty} Var(\bar{u}_{wt} | \mathcal{I}_{t-1}) &= \lim_{N \rightarrow \infty} \sum_{i=1}^N \sum_{j=1}^N w_{i,t-1} w_{j,t-1} (\gamma_{i1} \gamma_{j1} + \dots + \gamma_{im} \gamma_{jm}) \\ &= \lim_{N \rightarrow \infty} \left(\bar{\gamma}_{w_{t-1}, 1}^2 + \dots + \bar{\gamma}_{w_{t-1}, m}^2 \right) \end{aligned}$$

where $\bar{\gamma}_{w_{t-1}, \ell} = \sum_{i=1}^N w_{i,t-1} \gamma_{i\ell}$. This establishes that weak cross section dependence of z_{it} is equivalent to $\lim_{N \rightarrow \infty} \bar{\gamma}_{w_{t-1}, \ell} = 0$, for $\ell = 1, 2, \dots, m$, for all sets of weights satisfying (13)-(14). ■

As a consequence, if z_{it} is CWD then weighted averages of the factor loadings converge to zero as $N \rightarrow \infty$. Conversely, for a strongly dependent process, there must exist a set of weights satisfying (13)-(14) such that

$$\lim_{N \rightarrow \infty} Var(\bar{u}_{wt} | \mathcal{I}_{t-1}) = \bar{\gamma}_{w_{t-1}, 1}^2 + \dots + \bar{\gamma}_{w_{t-1}, m}^2 > 0, \quad (32)$$

which implies that there must at least be one factor $f_{\ell t}$ such that its loadings satisfy the condition,

$$\lim_{N \rightarrow \infty} \bar{\gamma}_{w_{t-1}, \ell} \neq 0.$$

In the literature on factor models, it is quite common to impose conditions on the loadings or on the eigenvalues of the conditional covariance matrix, $\mathbf{\Sigma}_{ut}$, of $\mathbf{u}_t = (u_{1t}, \dots, u_{Nt})'$ that constrain the

form of cross section dependence carried by the factor structure. For example, Bai (2005) assumes that the factor loadings satisfy $\sum_{i=1}^N \gamma_{i\ell}^2 = O(N)$, for $\ell = 1, \dots, m$. But this condition does not necessarily imply $\lim_{N \rightarrow \infty} \bar{\gamma}_{w_{t-1}, \ell} \neq 0$. It is, therefore, possible for a factor structure to exhibit weak cross section dependence in the sense of Definition 1. This, for example, could arise in the case where the factor loadings are independent draws from random a coefficient model with *zero means* and finite variances.

In the literature on asset pricing models, one common assumption is that $\lambda_m(\boldsymbol{\Sigma}_{ut})$ is bounded away from zero at rate N (Chamberlain, 1983; Forni and Lippi, 2001). Under this assumption, since $\text{rank}(\boldsymbol{\Sigma}_{ut}) = m$, and $\lambda_i(\boldsymbol{\Sigma}_{ut}) > 0$, for $i = 1, 2, \dots, m$, and $\lambda_i(\boldsymbol{\Sigma}_{ut}) = 0$, for $i = m + 1, m + 2, \dots, N$. Noting that under Assumption 7 $\lambda_1(\boldsymbol{\Sigma}_{et}) = O(1)$, we have

$$\lambda_m(\boldsymbol{\Sigma}_t) \geq \lambda_m(\boldsymbol{\Sigma}_{ut}),$$

and

$$\lambda_{m+1}(\boldsymbol{\Sigma}_t) \leq \lambda_{m+1}(\boldsymbol{\Sigma}_{ut}) + \lambda_1(\boldsymbol{\Sigma}_{et}) = \lambda_1(\boldsymbol{\Sigma}_{et}).$$

Hence, it follows that $\lambda_1(\boldsymbol{\Sigma}_t), \dots, \lambda_m(\boldsymbol{\Sigma}_t)$ increase without bound as $N \rightarrow \infty$, while $\lambda_{m+1}(\boldsymbol{\Sigma}_t), \dots, \lambda_N(\boldsymbol{\Sigma}_t)$ satisfy the bounded eigenvalue condition. Such a model is generally referred to as an approximate factor model. Most factor structures yield eigenvalues that increase at rate N . But as shown by Kapetanios and Marcellino (2006), it is possible to devise factor models that generate eigenvalues that rise at rate N^d , for $0 < d < 1$.

6 Weak cross section dependence in spatial models

When a metric of distance is available, spatial techniques can be used to allow for interactions and cross dependencies in regression model. Important applications of spatial models in economics can be found in the literature on regional income growth (Rivera-Batiz and Romer, 1991), in the field of microeconomics that analyses the diffusion processes of certain behaviours across a population (Brock and Durlauf, 2001), and in a recent strand of public economics that studies the role of interaction among policy makers (Brueckner, 2003).

In the space-time process $\{z_{it}, i \in \mathcal{S}, t \in \mathcal{T}\}$, the index i denotes the i^{th} spatial unit whose location s_i is known and varies over a fixed subset of the k -dimensional Euclidean space \mathbb{R}^k . The spatial approach to cross section dependence assumes that the covariance between any two random variables depends on their relative distance. The ordering or neighbour relation among the variables (z_{1t}, \dots, z_{Nt}) is expressed by means of the spatial weights matrix \mathbf{S}_t (Anselin, 1988, 2002). The specification of \mathbf{S}_t is in general arbitrary, typically based on some measures of distance between units. A range of suggestions have been offered in the literature, based on geographic distance (Cliff and Ord, 1981; Anselin, 1988), as well as more general metrics, such as the ones based on economic (Conley, 1999; Pesaran, Schuermann and Weiner, 2004), or social proximity (Conley and Topa, 2002). When dealing with geographic distance, it is common to define weights using the

notion of contiguity between units, and assigning a nonzero weight only when units i and j are contiguous. There exists a number of different ways to build a matrix that contains contiguity information for the same spatial layout; the essential point is deciding whether units that only share a common vertex should be considered neighbours or not. In the example of a regular square lattice, these alternatives are denoted as the *rook* formation under which two units are considered as neighbours if they share common boundaries; the *bishop* formation, if they share common vertices, and the *queen* formation if they share both boundaries and vertices. Alternatively, weights can have continuous values, in general a declining function of the distance between spatial observations, such as $s_{ij,t} = e^{-\varphi d_{ij,t}}$, or $s_{ij,t} = 1/d_{ij,t}^\varphi$, where $d_{ij,t}$ is the distance between units at time t , and φ is a parameter that can be set *a priori*, or estimated jointly (Haining, 2003). Finally, some authors have suggested to use a block structure for the weights matrix, by dividing the population into groups, and considering as neighbours only those units that belong to the same group (Case, 1991).

In this section we show that, under certain regularity conditions, cross section correlations arising from most widely applied spatial models can be considered as particular examples of CWD processes. We prove that, given certain conditions on the weights matrix, the largest eigenvalue of the covariance matrix of our sequence of variables is bounded as the number of cross section units is increased.

Consider the variables e_{it} for $i = 1, \dots, N$ collected in the vector $\mathbf{e}_t = (e_{1t}, \dots, e_{Nt})'$, for $t = 1, 2, \dots, T$. The process $\{e_{it}\}$ follows a Spatial Moving Average (SMA) process (Haining, 1978) if

$$\mathbf{e}_t = \delta \mathbf{S}_t \boldsymbol{\varepsilon}_t + \boldsymbol{\varepsilon}_t, \quad (33)$$

where δ is a scalar parameter and $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \dots, \varepsilon_{Nt})'$. In the spatial econometric literature where the focus is often on a single cross section, it is typically assumed that $\varepsilon_{it} \sim IID(0, \sigma^2)$. But in what follows we consider a less restrictive set up and allow the idiosyncratic errors, ε_{it} , to be serially correlated with mean zero and a finite variance $0 < \sigma_i^2 < K < \infty$. However, we continue to assume that ε_{it} and $\varepsilon_{jt'}$ are independently distributed for all t, t' ; and for all $i \neq j$.

For the spatial weight matrix, \mathbf{S}_t , we make the following assumption:

ASSUMPTION 8: The main diagonal elements of \mathbf{S}_t are zero,

- (a) $\delta \in \left(-\frac{1}{\rho_{\mathbf{S}}}, \frac{1}{\rho_{\mathbf{S}}}\right)$, where $\rho_{\mathbf{S}} = \max_{1 \leq t \leq T} \{\rho(\mathbf{S}_t)\}$,
- (b) $\|\mathbf{S}_t\|_r$ is bounded for all t ,
- (c) $\|\mathbf{S}_t\|_c$ is bounded for all t .

For the SMA model the covariance matrix of \mathbf{e}_t is

$$\boldsymbol{\Sigma}_t = (\mathbf{I}_N + \delta \mathbf{S}_t) \boldsymbol{\Lambda}_\varepsilon (\mathbf{I}_N + \delta \mathbf{S}_t'),$$

where $\boldsymbol{\Lambda}_\varepsilon$ is a diagonal matrix with σ_i^2 on its diagonal. We note that the SMA specification induces a short-range correlation across section, since the only off-diagonal non-zero elements of

the covariance matrix are those corresponding to non-zero elements in \mathbf{S}_t . Given the properties of matrix norms, the largest eigenvalue of $\boldsymbol{\Sigma}_t$ satisfies the following inequality

$$\lambda_1(\boldsymbol{\Sigma}_t) = \lambda_1[(\mathbf{I}_N + \delta\mathbf{S}_t)\boldsymbol{\Lambda}_\varepsilon(\mathbf{I}_N + \delta\mathbf{S}'_t)] \leq \sigma_{\max}^2 \|\mathbf{I}_N + \delta\mathbf{S}_t\|_c \|\mathbf{I}_N + \delta\mathbf{S}'_t\|_c,$$

where $\sigma_{\max}^2 = \max_{1 \leq i \leq N} \{\sigma_i^2\} < K$. Hence, under Assumptions 8(b) and 8(c), e_{it} is a weakly cross sectionally correlated process. As we will see later on in this section, bounded row and column norms of the weights matrix is a very common assumption in the spatial econometric literature. Note that this assumption rules out the case where, for some t , \mathbf{S}_t has dominant units.

We say that $\{e_{it}\}$ follows a Spatial Error Component (SEC) process (Kelejian and Robinson, 1995) if

$$\mathbf{e}_t = \delta\mathbf{S}_t\boldsymbol{\psi}_t + \boldsymbol{\varepsilon}_t, \quad (34)$$

where $\boldsymbol{\psi}_t$ and $\boldsymbol{\varepsilon}_t$ are two uncorrelated vectors of random variables, with zero mean and diagonal variance matrices $\boldsymbol{\Lambda}_\psi$ and $\boldsymbol{\Lambda}_\varepsilon$, with elements ω_i^2 and σ_i^2 , respectively. In this case \mathbf{e}_t has covariance matrix

$$\boldsymbol{\Sigma}_t = \delta^2\mathbf{S}_t\boldsymbol{\Lambda}_\psi\mathbf{S}'_t + \boldsymbol{\Lambda}_\varepsilon.$$

The range of spatial correlation induced by the SEC is very similar to that in the SMA, since it is constrained to close neighbours, for which the off-diagonal elements of the weighting matrix are non-zero. The largest eigenvalue of $\boldsymbol{\Sigma}_t$ satisfies

$$\lambda_1(\boldsymbol{\Sigma}_t) \leq \delta^2\omega_{\max}^2 \|\mathbf{S}_t\|_c \|\mathbf{S}_t\|_r + \sigma_{\max}^2.$$

where $\omega_{\max}^2 = \max_{1 \leq i \leq N} \{\omega_i^2\} < K$. As for the SMA process, Assumptions 8(b) and 8(c) ensure that the SEC model is cross sectionally weakly dependent.

Consider now the Spatial Autoregressive (SAR) process introduced at the end of Section 2:

$$\mathbf{e}_t = \delta\mathbf{S}_t\mathbf{e}_t + \boldsymbol{\varepsilon}_t. \quad (35)$$

Under Assumption 8(a), this is an invertible process and can be expanded according to the geometric series (Horn and Johnson, 1985, p. 301)

$$\mathbf{e}_t = \sum_{h=0}^{\infty} (\delta\mathbf{S}_t)^h \boldsymbol{\varepsilon}_t.$$

The covariance matrix of the SAR process is

$$\boldsymbol{\Sigma}_t = (\mathbf{I}_N - \delta\mathbf{S}_t)^{-1}\boldsymbol{\Lambda}_\varepsilon(\mathbf{I}_N - \delta\mathbf{S}'_t)^{-1} = \left[\sum_{h=0}^{\infty} (\delta\mathbf{S}_t)^h \right] \boldsymbol{\Lambda}_\varepsilon \left[\sum_{h=0}^{\infty} (\delta\mathbf{S}'_t)^h \right].$$

As $\boldsymbol{\Sigma}_t$ contains matrix powers and products of \mathbf{S}_t , even if \mathbf{S}_t contains few non-zero values, the covariance structure induced by the SAR model is not sparse. This argument has been used in spatial econometrics literature to claim that the SAR process describes a *global* pattern of spatial

correlation, since it links all units in the system to each other (Anselin, 2002). As a result, the cross section correlation carried by a Spatial Autoregressive representation can be much more extensive than that implied by a SMA or SEC processes. However, we note that the SAR model, under some regularity conditions, is weakly cross sectionally correlated. Indeed, the largest eigenvalue of Σ_t satisfies the following inequality

$$\begin{aligned}\lambda_1(\Sigma_t) &\leq \sigma_{\max}^2 \sum_{h=0}^{\infty} |\delta|^h \|\mathbf{S}_t^h\|_c \sum_{h=0}^{\infty} |\delta|^h \|\mathbf{S}_t^h\|_r, \\ &\leq \sigma_{\max}^2 \sum_{h=0}^{\infty} |\delta|^h \|\mathbf{S}_t\|_c^h \sum_{h=0}^{\infty} |\delta|^h \|\mathbf{S}_t\|_r^h.\end{aligned}$$

Hence, under Assumptions 8(b) and 8(c), and under the condition

$$\delta < 1/\|\mathbf{S}_t\|_c, \text{ and } \delta < 1/\|\mathbf{S}_t\|_r, \quad (36)$$

we have

$$\lambda_1(\Sigma_t) < \|\Sigma_t\|_c \leq \sigma_{\max}^2 \left(\frac{1}{1 - |\delta| \|\mathbf{S}_t\|_c} \right) \left(\frac{1}{1 - |\delta| \|\mathbf{S}_t\|_r} \right) < K,$$

namely \mathbf{e}_t is a CWD process. Note that (36) also implies that the spatial autoregressive process is invertible, hence, under (36), Assumption 8(a) is not needed.

The conditions on bounded row and column norms of \mathbf{S}_t implies that each unit has a finite number of neighbours that does not increase with N , namely \mathbf{S}_t has no dominant units (columns/rows). This way of expanding the spatial weight matrix as the number of cross section units gets larger corresponds to the notion of increasing-domain asymptotics, which is the asymptotic based on a growing spatial domain, rather than on increasingly dense observations in a fixed and bounded region (Cressie, 1993).

We conclude this section noting that the SMA and the SAR processes, under the assumption of invertibility (i.e. Assumption 8(a) for the SAR process), are particular cases of the following general process

$$\mathbf{e}_t = \mathbf{R}_t \boldsymbol{\varepsilon}_t, \quad (37)$$

where \mathbf{R}_t is a given $N \times N$ matrix. Specifically, for a SMA process $\mathbf{R}_t = (\mathbf{I}_N + \delta \mathbf{S}_t)$, and for a SAR process $\mathbf{R}_t = (\mathbf{I}_N - \delta \mathbf{S}_t)^{-1}$. This matrix plays a crucial role in the analysis of spatial models, and as we shall see below typically satisfies the following assumption:

ASSUMPTION 9: \mathbf{R}_t has bounded row and column norms for all t .

The key result for the cross section averages of \mathbf{e}_t is summarized in the following theorem.

Theorem 17 Consider the process (37), where \mathbf{R}_t satisfies Assumption 9, and for each t , ε_{it} is distributed independently across i with mean zero, finite variances $0 < \sigma_i^2 < \sigma_{\max}^2 < \infty$, and finite fourth-order moments, $E(\varepsilon_{it}^4) = \mu'_{i4} < K < \infty$. Then for all sets of weights satisfying (13)-(14),

$$E(\bar{e}_{wt}^2 | \mathcal{I}_{t-1}) = O(N^{-1}), \text{ and } \text{Var}(\bar{e}_{wt}^2 | \mathcal{I}_{t-1}) = O(N^{-2}).$$

Proof. First note that

$$E(\bar{e}_{wt}^2 | \mathcal{I}_{t-1}) = \mathbf{w}'_{t-1} \mathbf{R}_t \boldsymbol{\Lambda}_\varepsilon \mathbf{R}'_t \mathbf{w}_{t-1} \leq \sigma_{\max}^2(\mathbf{w}'_{t-1} \mathbf{w}_{t-1}) \lambda_1(\mathbf{R}_t \mathbf{R}'_t).$$

But since \mathbf{R}_t has bounded row and column norms, $\lambda_1(\mathbf{R}_t \mathbf{R}'_t)$ is bounded, and for all sets of weights satisfying (13)-(14) we have

$$E(\bar{e}_{wt}^2 | \mathcal{I}_{t-1}) = O(N^{-1}).$$

To prove the second part of the theorem, let

$$z_t = (\boldsymbol{\kappa}'_t \boldsymbol{\kappa}_t)^{-1/2} \boldsymbol{\kappa}'_t \boldsymbol{\eta}_t,$$

where

$$\boldsymbol{\kappa}_t = \boldsymbol{\Lambda}_\varepsilon^{1/2} \mathbf{R}'_t \mathbf{w}_{t-1}, \text{ and } \boldsymbol{\eta}_t = \boldsymbol{\Lambda}_\varepsilon^{-1/2} \boldsymbol{\varepsilon}_t. \quad (38)$$

Then

$$\bar{e}_{wt}^2 = (\mathbf{w}'_{t-1} \mathbf{R}_t \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t \mathbf{R}'_t \mathbf{w}_{t-1})^2 = (\boldsymbol{\kappa}'_t \boldsymbol{\kappa}_t) z_t^2,$$

and

$$\text{Var}(\bar{e}_{wt}^2 | \mathcal{I}_{t-1}) = (\boldsymbol{\kappa}'_t \boldsymbol{\kappa}_t)^2 \text{Var}(z_t^2).$$

But for each t , z_t is a scalar random variable with mean zero, a unit variance, and finite fourth order moment (since it is a linear function of ε_t). Note also that since ε_{it} are cross sectionally independently distributed then by standard central limit theorems $z_t^2 \stackrel{a}{\rightsquigarrow} \chi_1^2$ as $N \rightarrow \infty$. Hence $\text{Var}(z_t^2)$ is finite for all N . As a result

$$\text{Var}(\bar{e}_{wt}^2 | \mathcal{I}_{t-1}) \leq K(\boldsymbol{\kappa}'_t \boldsymbol{\kappa}_t)^2.$$

Now substituting for $\boldsymbol{\kappa}_t$ from (38) we have

$$\text{Var}(\bar{e}_{wt}^2 | \mathcal{I}_{t-1}) \leq \sigma_{\max}^4(\mathbf{w}'_{t-1} \mathbf{w}_{t-1})^2 [\lambda_1(\mathbf{R}_t \mathbf{R}'_t)]^2.$$

As $\lambda_1(\mathbf{R}_t \mathbf{R}'_t)$ and σ_{\max}^2 are bounded, then in view of (13)-(14), it follows that

$$\text{Var}(\bar{e}_{wt}^2 | \mathcal{I}_{t-1}) = O(N^{-2}).$$

■

Therefore, for any invertible spatial process $\{e_{it}\}$, where the spatial weight matrix \mathbf{S}_t has bounded row and column norms, \bar{e}_{wt}^2 converges to zero in quadratic mean as $N \rightarrow \infty$, for all sets of predetermined weights satisfying the granularity conditions. It is worth emphasizing that this result holds even if the idiosyncratic errors, ε_{it} , follow heterogeneous processes that are serially correlated over time. But to identify the spatial model we would need some restrictions on their cross section correlations. For the purpose of proving the consistency of the CCE estimator it is sufficient that $\bar{e}_{wt}^2 \xrightarrow{q.m.} 0$ as $N \rightarrow \infty$. (see Section 7).

The assumptions required for spatial correlation to be weak are often adopted in spatial econometric literature as regularity conditions, since they ensure consistency and asymptotic normality of maximum likelihood estimators (Lee, 2004; Kelejian and Prucha, 1999; Mardia and Marshall, 1984). Kelejian and Prucha (1999) prove consistency of their GMM estimator in the context of a single cross section under the assumption that the matrices \mathbf{S} and $(\mathbf{I}_N - \delta\mathbf{S})^{-1}$, for $|\delta| < 1$, have bounded row and column norms. The proofs of consistency of GMM estimator (and also of ML) are based on a central limit theorem for quadratic forms in ε (see Kelejian and Prucha, 1999, 2001). Lee (2004) has studied the properties of ML and quasi-ML estimators of a spatial process where each unit is influenced by few neighboring units. The author shows that ML and quasi-ML estimators are consistent and asymptotic normal under the assumption that the matrix \mathbf{S} has bounded row and column norms, and that $(\mathbf{I}_N - \delta\mathbf{S})^{-1}$ has bounded row and column norms at the true value of δ (say δ_0), and is uniformly bounded in either row or column sums for any other value of δ in a compact set around δ_0 .

Mardia and Marshall (1984) have demonstrated that in a regression equation with correlated errors, bounds on the largest eigenvalues of the error covariance matrix and of its first and second order derivatives (with respect to the unknown parameters) are sufficient for weak consistency and asymptotic normality of the ML estimator of the spatial parameters.

7 Estimating panels with common factors and spatial correlation

Consider the panel data model with common factors set out in equations (1)-(2), where e_{it} is an invertible spatial process of the type (37). From the discussion on strong and weak cross section correlation, it follows that Assumption 9 guarantees that (8) holds. Hence, the unobservable common factors can be well approximated by the cross section averages of the dependent variable and individual specific regressors. We now focus on estimation and inference on the means of the slope parameters β_i . To this end, we assume for ease of exposition that the rank condition (6) is satisfied, though this assumption can be relaxed. Let $\bar{\mathbf{M}}_w$ be defined by

$$\bar{\mathbf{M}}_w = \mathbf{I}_T - \bar{\mathbf{H}}_w(\bar{\mathbf{H}}_w' \bar{\mathbf{H}}_w)^{-1} \bar{\mathbf{H}}_w',$$

$\bar{\mathbf{H}}_w = (\mathbf{D}, \bar{\mathbf{Z}}_w)$, where \mathbf{D} and $\bar{\mathbf{Z}}_w$ are, respectively, the matrices of observations on \mathbf{d}_t and $\bar{\mathbf{z}}_{wt} = (\bar{y}_{wt}, \bar{\mathbf{x}}_{wt})$. We make the following additional assumption:

ASSUMPTION 10: Consider the cross section averages of the individual specific variables, $\mathbf{z}_{it} = (y_{it}, \mathbf{x}'_{it})'$ defined by $\bar{\mathbf{z}}_{wt} = \sum_{i=1}^N w_i \mathbf{z}_{it}$, with time-invariant weights $\{w_j\}$ satisfying Assumption 5. Then the following conditions hold:

- (a) The $k \times k$ matrices $\hat{\Psi}_{iT} = T^{-1}(\mathbf{X}'_i \bar{\mathbf{M}}_w \mathbf{X}_i)$ and $\hat{\Psi}_{ig} = T^{-1}(\mathbf{X}'_i \mathbf{M}_g \mathbf{X}_i)$, where $\bar{\mathbf{M}}_g = \mathbf{I}_T - \mathbf{G}(\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}'$ and $\mathbf{G} = (\mathbf{D}, \mathbf{F})$, are non-singular and $\hat{\Psi}_{iT}^{-1}$ and $\hat{\Psi}_{ig}^{-1}$ have finite second order moments, for all i

(b) The matrix $\hat{\Psi}_{NT} = \sum_{i=1}^N w_i \frac{\mathbf{X}_i' \bar{\mathbf{M}}_w \mathbf{X}_i}{T}$ is non-singular.

Pesaran (2006) has suggested two alternative estimators for the means β of the individual slope coefficients. The Common Correlated Effects Mean Group (CCEMG) estimator and the Common Correlated Effects Pooled (CCEP) estimator. The latter is defined by

$$\hat{\mathbf{b}}_{MG} = N^{-1} \sum_{i=1}^N \hat{\mathbf{b}}_i, \quad (39)$$

where

$$\hat{\mathbf{b}}_i = (\mathbf{X}_i' \bar{\mathbf{M}}_w \mathbf{X}_i)^{-1} \mathbf{X}_i' \bar{\mathbf{M}}_w \mathbf{y}_i, \quad (40)$$

and $\mathbf{X}_i = (\mathbf{x}_{i1}, \dots, \mathbf{x}_{iT})'$, $\mathbf{y}_i = (y_{i1}, \dots, y_{iT})'$. The following theorem applies to this estimator.

Theorem 18 *Consider the panel data model (1)-(2) with errors e_{it} given by (37). Suppose that Assumptions 1-5, 9 and 10(a) hold. Then the common correlated effects mean group estimator $\hat{\mathbf{b}}_{MG}$ given by (39) is asymptotically (for fixed T and $N \rightarrow \infty$) unbiased for β , and as $(N, T) \xrightarrow{j} \infty$ we have*

$$\sqrt{N} (\hat{\mathbf{b}}_{MG} - \beta) \rightarrow N(0, \Sigma_{MG}),$$

where

$$\Sigma_{MG} = \Omega_v.$$

A proof is provided in the Appendix. The covariance matrix, Σ_{MG} , can be consistently estimated by

$$\hat{\Sigma}_{MG} = \frac{1}{N-1} \sum_{i=1}^N (\hat{\mathbf{b}}_i - \hat{\mathbf{b}}_{MG}) (\hat{\mathbf{b}}_i - \hat{\mathbf{b}}_{MG})'. \quad (41)$$

The CCEP estimator, which gains efficiency from pooling observations if $\beta_i = \beta$, is defined by

$$\hat{\mathbf{b}}_P = \left(\sum_{i=1}^N \mathbf{X}_i' \bar{\mathbf{M}}_w \mathbf{X}_i \right)^{-1} \sum_{i=1}^N \mathbf{X}_i' \bar{\mathbf{M}}_w \mathbf{y}_i. \quad (42)$$

The following theorem applies to this estimator.

Theorem 19 *Consider the panel data model (1)-(2) with errors e_{it} given by (37). Suppose that Assumptions 1-5, 9 and 10(b) hold. Then the common correlated effects pooled estimator $\hat{\mathbf{b}}_P$ given by (42) is asymptotically unbiased for β and as $(N, T) \xrightarrow{j} \infty$ we have*

$$\left(\sum_{i=1}^N w_i^2 \right)^{-1/2} (\hat{\mathbf{b}}_P - \beta) \rightarrow N(0, \Sigma_P),$$

where

$$\Sigma_P = \Psi^{*-1} \mathbf{R}^* \Psi^{*-1},$$

with

$$\begin{aligned}\Psi^* &= \lim_{N \rightarrow \infty} \left(\sum_{i=1}^N w_i \Sigma_i \right), \\ \mathbf{R}^* &= \lim_{N \rightarrow \infty} \left[N^{-1} \sum_{i=1}^N \tilde{w}_i^2 (\Sigma_i \Omega_{\mathbf{v}} \Sigma_i) \right], \\ \tilde{w}_i &= \frac{w_i}{\sqrt{N^{-1} \sum_{i=1}^N w_i^2}}.\end{aligned}$$

See the Appendix for a proof. A consistent estimator for Σ_P is (Pesaran, 2006)

$$\hat{\Sigma}_P = \left(\sum_{i=1}^N w_i^2 \right) \hat{\Psi}^{*-1} \hat{\mathbf{R}}^* \hat{\Psi}^{*-1}, \quad (43)$$

with

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

In the rest of the paper, we study the small sample properties of CCE estimators in a model with common factors and/or spatial correlations.

8 Monte Carlo experiments

8.1 Monte Carlo design

This section provides Monte Carlo evidence on the small sample properties of CCE estimators, under different assumptions on the stochastic process generating the error terms. The study is comprised of four sets of experiments. In the first set to be used as a benchmark, we replicate the Monte Carlo study in Pesaran (2006), where the dependent variable and the regressors are assumed to depend on a linear combination of unobserved common factors. In a second set of experiments, we consider a panel where the error terms are generated by a spatial autoregressive (SAR) process. In the third set, we combine the two sources of cross section dependence, assuming that the error term is the orthogonal sum of a factors structure and a spatial process. In the fourth set of experiments, we allow the source of cross section dependence to vary over time, assuming that in a certain interval of time the error term follows a factor structure, and in the remaining time periods is generated by a pure spatial process.

For all experiments we considered the following data generating process

$$\begin{aligned} y_{it} &= \alpha_i d_{1t} + \beta_{i1} x_{1it} + \beta_{i2} x_{2it} + \gamma_{i1} f_{1t} + \gamma_{i2} f_{2t} + e_{it}, \\ x_{jit} &= a_{ij1} d_{1t} + a_{ij2} d_{2t} + \gamma_{ij1} f_{1t} + \gamma_{ij2} f_{3t} + v_{ijt}, \quad j = 1, 2, \end{aligned}$$

for $i = 1, \dots, N$ and $t = 1, \dots, T$, where we assume homogeneous slopes with $\beta_i = \beta = (1, 1)'$. In the above equation, d_{1t} and d_{2t} are observed common factors, f_{1t} , f_{2t} , and f_{3t} are unobserved common effects, and e_{it} are idiosyncratic errors. We adopt the following data generating processes:

$$\begin{aligned} d_{1t} &= 1, \quad d_{2t} = \delta_d d_{2t-1} + v_{dt}, \quad t = 1, \dots, T, \\ v_{dt} &\sim IIDN(0, 1 - \delta_d^2), \quad \delta_d = 0.5, \quad d_{2,0} = 0, \end{aligned}$$

$$\begin{aligned} f_{jt} &= \delta_{f_j} f_{jt-1} + v_{f_{jt}}, \quad j = 1, 2, 3; \quad t = 1, \dots, T, \\ v_{f_{jt}} &\sim IIDN(0, 1 - \delta_{f_j}^2), \quad \delta_{f_j} = 0.5, \quad f_{j0} = 0, \end{aligned}$$

and

$$\begin{aligned} v_{ijt} &= \delta_{v_{ijt}} v_{ijt-1} + v_{ijt}, \quad t = 1, \dots, T, \\ v_{ijt} &\sim IIDN(0, 1 - \delta_{v_{ijt}}^2), \quad v_{ij0} = 0, \\ \delta_{v_{ijt}} &\sim IIDU(0.05, 0.95) \text{ for } j = 1, 2. \end{aligned}$$

The factor loadings of the observed common effects do not change across replications and are generated as

$$\begin{aligned} \alpha_i &\sim IIDN(1, 1), \quad i = 1, \dots, N, \\ (a_{i11}, a_{i21}, a_{i12}, a_{i22}) &\sim IIDN(0.5\boldsymbol{\tau}_4, 0.5\mathbf{I}_4), \end{aligned}$$

where $\boldsymbol{\tau}_4 = (1, 1, 1, 1)'$ and \mathbf{I}_4 is a 4×4 identity matrix.

The various experiments involve different hypothesis on the loadings of the unobserved common factors, and on the way the errors e_{it} are generated. In particular, we have carried out the following sets of experiments:

Experiment A

The parameters of the unobserved common effects in the \mathbf{x}_{it} equation are generated as

$$\begin{pmatrix} \gamma_{i11} & 0 & \gamma_{i13} \\ \gamma_{i21} & 0 & \gamma_{i23} \end{pmatrix} \sim IID \begin{pmatrix} N(0.5, 0.5) & 0 & N(0, 0.5) \\ N(0, 0.5) & 0 & N(0.5, 0.5) \end{pmatrix},$$

the parameters of the unobserved common effects in the y_{it} equation are⁷

$$\gamma_{i1} \sim IIDN(1, 0.2), \quad \gamma_{i2} \sim IIDN(1, 0.2).$$

Finally, in this experiment the errors are independently distributed across i , namely

$$e_{it} \sim IIDN(0, \sigma_i^2), \quad \sigma_i^2 \sim IIDU[0.5, 1.5], \quad \text{for } i = 1, \dots, N.$$

Experiment B

The factor loadings of the unobserved common effects are set to zero, $\gamma_{i11} = \gamma_{i13} = \gamma_{i21} = \gamma_{i23} = \gamma_{i1} = \gamma_{i2} = 0$. In other words, the regressors and the dependent variable do not depend on the unobserved factors \mathbf{f}_t . The individual specific errors e_{it} are generated according to a spatial autoregressive process

$$e_{it} = \delta \sum_{j=1}^N s_{ij} e_{jt} + \varepsilon_{it}, \quad \text{for } i = 1, \dots, N, t = 1, \dots, T$$

$$\varepsilon_{it} \sim IIDN(0, \sigma^2), \sigma^2 = 1,$$

where δ is the spatial autoregressive coefficient, s_{ij} , for $i, j = 1, \dots, N$, are elements of a spatial weight matrix \mathbf{S} , assumed to be time-invariant.

In this experiment we consider a spatial weight matrix \mathbf{S} with elements $s_{ij} = 1$ if units i and j are adjacent and $s_{ij} = 0$ otherwise. Further, we assume that cross section units are ordered so that the p^{th} order neighbours of the i^{th} cross section unit can be defined as $i - p$ and $i + p$ cross section units. The spatial weight matrix is defined in a circular world, where the first observation is adjacent to the last observation. Finally the weighting matrix is row-standardized, so that all its non-zero elements are equal $1/2p$. We note that such spatial weight matrix has bounded row and column sums.

We have experimented with two different values of the spatial coefficient, $\delta = 0.8$, which corresponds to a sizeable level of cross section error correlation, and $\delta = 0.4$ which correspond to a more moderate level of spatial error correlation. Further, we have experimented with a spatial weight matrix of order $p = 2$, and a sparser matrix of order $p = 1$. In total, Experiment B consists of four different sub-experiments.

This experiment is intended to illustrate the extent to which the inclusion of cross section averages of the dependent variable and regressors as in the CCE approach is able to capture spatial correlations.

⁷These parameters are drawn such that the rank condition (6) is satisfied.

Experiment C

The parameters of the unobserved common effects in the \mathbf{x}_{it} and in the y_{it} equations are generated as in Experiment A

$$\begin{pmatrix} \gamma_{i11} & 0 & \gamma_{i13} \\ \gamma_{i21} & 0 & \gamma_{i23} \end{pmatrix} \sim IID \begin{pmatrix} N(0.5, 0.5) & 0 & N(0, 0.5) \\ N(0, 0.5) & 0 & N(0.5, 0.5) \end{pmatrix},$$

$$\gamma_{i1} \sim IIDN(1, 0.2), \gamma_{i2} \sim IIDN(1, 0.2).$$

The individual specific errors e_{it} are generated as in Experiment B, that is

$$e_{it} = \delta \sum_{j=1}^N s_{ij} e_{jt} + \varepsilon_{it}, \quad \text{for } i = 1, \dots, N, t = 1, \dots, T$$

$$\varepsilon_{it} \sim IIDN(0, \sigma^2), \quad \sigma^2 = 1,$$

where δ is the spatial autoregressive coefficient, s_{ij} are the spatial weights.

Again, we have experimented the case of high spatial correlation $\delta = 0.8$, and the case of moderate spatial dependence $\delta = 0.4$, and changed the order of the weights matrix from $p = 1$ to $p = 2$. This set of experiments aims at verifying to what extent CCE estimators capture both local and global cross section dependence.

Experiment D

In these experiments the source of cross section dependence is allowed to vary over time. In particular, we assume that common factors have nonzero loadings in the equations for y_{it} and \mathbf{x}_{it} only over a certain interval of time, and assume that cross section dependence is generated by a spatial process for the remaining sample period. We consider the following two sub-sets of experiments, where the cross dependence changes from strong to weak and back to strong; and a second sub-set under which the cross dependence begins to be weak, turns into strong before reverting back to the weak form.

Experiment D1

- For $t = 1, \dots, \lfloor \frac{T}{3} \rfloor$ parameters of the unobserved common effects in the \mathbf{x}_{it} and in the y_{it} equations are generated as in Experiment A

$$\begin{pmatrix} \gamma_{i11} & 0 & \gamma_{i13} \\ \gamma_{i21} & 0 & \gamma_{i23} \end{pmatrix} \sim IID \begin{pmatrix} N(0.5, 0.5) & 0 & N(0, 0.5) \\ N(0, 0.5) & 0 & N(0.5, 0.5) \end{pmatrix},$$

$$\gamma_{i1} \sim IIDN(1, 0.2), \gamma_{i2} \sim IIDN(1, 0.2).$$

- $t = \lfloor \frac{T}{3} \rfloor + 1, \dots, \lfloor \frac{2T}{3} \rfloor$, we set $\gamma_{i11} = \gamma_{i13} = \gamma_{i21} = \gamma_{i23} = \gamma_{i1} = \gamma_{i2} = 0$ (as in Experiment B),

and assume

$$e_{it} = \delta \sum_{j=1}^N s_{ij} e_{jt} + \varepsilon_{it}, \quad \text{for } i = 1, \dots, N, t = 1, \dots, T$$

$$\varepsilon_{it} \sim IIDN(0, \sigma^2), \sigma^2 = 1,$$

with $\delta = 0.4$.

- For $t = \lfloor \frac{2T}{3} \rfloor + 1, \dots, T$ parameters of the unobserved common effects in the \mathbf{x}_{it} and in the y_{it} equations are generated as in Experiment A

$$\begin{pmatrix} \gamma_{i11} & 0 & \gamma_{i13} \\ \gamma_{i21} & 0 & \gamma_{i23} \end{pmatrix} \sim IID \begin{pmatrix} N(0.5, 0.5) & 0 & N(0, 0.5) \\ N(0, 0.5) & 0 & N(0.5, 0.5) \end{pmatrix},$$

$$\gamma_{i1} \sim IIDN(1, 0.2), \gamma_{i2} \sim IIDN(1, 0.2).$$

Experiment D2

- For $t = 1, \dots, \lfloor \frac{T}{3} \rfloor$ we set $\gamma_{i11} = \gamma_{i13} = \gamma_{i21} = \gamma_{i23} = \gamma_{i1} = \gamma_{i2} = 0$ (as in Experiment B), and assume

$$e_{it} = \delta \sum_{j=1}^N s_{ij} e_{jt} + \varepsilon_{it}, \quad \text{for } i = 1, \dots, N, t = 1, \dots, T$$

$$\varepsilon_{it} \sim IIDN(0, \sigma^2), \sigma^2 = 1,$$

with $\delta = 0.4$

- $t = \lfloor \frac{T}{3} \rfloor + 1, \dots, \lfloor \frac{2T}{3} \rfloor$ the parameters of the unobserved common effects in the \mathbf{x}_{it} and in the y_{it} equations are generated as in Experiment A

$$\begin{pmatrix} \gamma_{i11} & 0 & \gamma_{i13} \\ \gamma_{i21} & 0 & \gamma_{i23} \end{pmatrix} \sim IID \begin{pmatrix} N(0.5, 0.5) & 0 & N(0, 0.5) \\ N(0, 0.5) & 0 & N(0.5, 0.5) \end{pmatrix},$$

$$\gamma_{i1} \sim IIDN(1, 0.2), \gamma_{i2} \sim IIDN(1, 0.2).$$

- For $t = \lfloor \frac{2T}{3} \rfloor + 1, \dots, T$ we set $\gamma_{i11} = \gamma_{i13} = \gamma_{i21} = \gamma_{i23} = \gamma_{i1} = \gamma_{i2} = 0$ (as in Experiment B), and assume

$$e_{it} = \delta \sum_{j=1}^N s_{ij} e_{jt} + \varepsilon_{it}, \quad \text{for } i = 1, \dots, N, t = 1, \dots, T$$

$$\varepsilon_{it} \sim IIDN(0, \sigma^2), \sigma^2 = 1,$$

with $\delta = 0.4$.

The aim of this set of experiments is to investigate the robustness of CCE estimators to the possible time variations in the nature of cross section dependence.

Each experiment was replicated 2,000 times for the (N, T) pairs with $N, T = 20, 30, 50, 100, 200$. In each experiment we computed the CCE Mean Group and the CCE Pooled estimator provided by formula (39) and (42), assuming equal weights $w_i = \frac{1}{N}$, $i = 1, \dots, N$. We further considered a misspecified structure that ignores the presence of common factors and/or spatial correlations, i.e. the fixed effects estimator

$$\hat{\mathbf{b}}_{FE} = \left(\sum_{i=1}^N \mathbf{X}_i' \mathbf{M}_\tau \mathbf{X}_i \right)^{-1} \sum_{i=1}^N \mathbf{X}_i' \mathbf{M}_\tau \mathbf{y}_i, \quad (44)$$

where $\mathbf{M}_d = \mathbf{I}_T - \boldsymbol{\tau}(\boldsymbol{\tau}'\boldsymbol{\tau})^{-1}\boldsymbol{\tau}'$, and $\boldsymbol{\tau}$ is a vector of ones.

To facilitate the interpretation of results, in each experiment we computed a statistic of cross section dependence, the CD test (Pesaran, 2004), a statistic of local cross section correlation, the $CD(p)$, and the simple average of pair-wise cross section correlation coefficients of the residuals, $\bar{\hat{r}}$. We have chosen these tests because they do not require the specification of a generating process for the error term. The CD statistic is

$$CD = \sqrt{\frac{2T}{N(N-1)}} \left(\sum_{i=1}^{N-1} \sum_{j=i+1}^N \hat{r}_{ij} \right),$$

where \hat{r}_{ij} is the sample estimate of the pair-wise correlation of the residuals, specifically

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

and \hat{u}_{it} is an estimate of the regression residuals $u_{it} = y_{it} - \alpha_i d_{1t} - \boldsymbol{\beta}' \mathbf{x}_{it}$, using the pooled estimator $\hat{\mathbf{b}}_P$ of $\boldsymbol{\beta}$. Pesaran (2004) has shown that the CD test is suitable under global alternatives such as the multi-factor residual models. However, when the cross section units can be ordered, it is more appropriate to compute the following $CD(p)$ test statistic

$$CD(p) = \sqrt{\frac{2T}{p(2N-p-1)}} \left(\sum_{s=1}^p \sum_{i=s+1}^N \hat{r}_{i,i-s} \right),$$

where p is the order of the spatial weight matrix. Finally, the average of pair-wise cross section correlation coefficients is

$$\bar{\hat{r}} = \frac{2}{N(N-1)} \left(\sum_{i=1}^{N-1} \sum_{j=i+1}^N \hat{r}_{ij} \right).$$

This Monte Carlo study is intended to investigate the relationship between the small sample properties of a number of estimators and the source of cross section dependence. In addition, this analysis provides interesting results for a number of issues. First, the performance of the fixed effects estima-

tor in Experiment B highlights the consequences of ignoring spatial correlation in panels. Further, we can verify whether various assumptions of weak dependence in the error term have an impact on the asymptotic behaviour of the statistics of global and local cross section dependence. Finally, we can explore whether there is a link between the range and intensity of cross section dependence and the performance of estimators.

8.2 Monte Carlo results

Results on the estimation of the slope parameters for Experiments A, B, C and D are summarized in Tables 1, 2, 3, and 4, respectively. In what follows we focus on the estimation of β_1 ; results for β_2 are very similar and are not reported. Further, we only provide results for the spatial weight matrix of order $p = 2$, since a more sparse weights matrix (case $p = 1$) leads to very similar results. Finally, to save space, we only provide Size and Power of estimators for Experiments B and C, while we do not report their Bias and RMSE, since these are very close to Bias and RMSE values in Experiments A. Note that the power of the various tests are computed under the alternative $H_1 : \beta_1 = 0.95$.

Table 1 shows results that are very similar to those in Pesaran (2006). First, the fixed effects estimator, ignoring the factor structure, performs very poorly, is substantially biased, and is subject to large size distortions. Further, as expected in the case of homogeneous slopes, the RMSE of the CCE Pooled estimators is lower than the RMSE of the CCE Mean Group estimator, although the difference between them becomes very small as N and T increase. Finally, tests using the CCE Mean Group and the CCE Pooled estimators are correctly sized, reflecting the fact that the estimated variance is a consistent estimate of the true variance.

Moving to Experiments B, Table 2 refers to the case where the error term is generated by a spatial autoregressive process. Monte Carlo results show that the fixed effects estimator, ignoring the spatial correlation, over-rejects the null hypothesis. The over-rejection tendency is due to the use of inappropriate standard errors, and appears to be substantial only in the case of the large value of the spatial parameter (namely when $\delta = 0.8$). Note that in the case of these experiments the FE estimator is consistent.

Focusing on the tests that use CCE Mean Group and CCE Pooled estimators, their empirical sizes are very close to the nominal size, for both values of the spatial parameter. Similarly, in Experiments C (Table 3), the combination of common factors and spatial correlation in the error term does not affect the empirical size of CCE estimators, which is close to the level of 5%. These Monte Carlo results clearly show that the CCE approach seems to work well not only in the case of global cross section dependence but also in the presence of spatial correlation. Finally, results from Experiments D (Table 4) suggest that CCE estimators are also robust to time variations in the source of cross section dependence.

Table 5 reports some statistics of cross section dependence for the above experiments, based on residuals from the CCE Pooled estimation. The average cross section correlation in Experiments A is high and persistent as N increases, for all values of T . Conversely, in Experiments B, \bar{r} declines rapidly to zero as the sample size is increased. In fact, local correlation implies that units have a

limited number of neighbours regardless of the sample size. Therefore, the average of local cross section correlation computed over all units decreases as N increases. This also affects the CD statistic, which slightly diminishes as N is increased, for all values of T .

9 Concluding remarks

Cross section dependence is a rapidly growing field of study in panel data analysis. In the case of panels with large cross section dimension, unobserved factor models, or when a measure of distance among units is available, spatial processes are used. The essential aim of this paper has been to discuss the estimation of a panel regression model with common factors and spatial dependence. To this end, we introduced the notions of strong and weak cross section dependence, and show that these notions are critical to the distinction that prevails in the literature between the two approaches advanced for modelling of cross section dependence. An important extension of this paper would be to study of the properties of classical estimation techniques in regression models with cross sectionally weakly correlated errors. It might also be worth considering spatial patterns that are not weak and are dominated by one or more cross section units. This is likely to be a theoretically more meaningful approach to investigate common effects as compared to the purely statistical factor models the results of which are often difficult to interpret.

Appendix

We now sketch the proofs of Theorems 18 and 19. We refer to Pesaran (2006) for further details.

Lemma 20 *Let $\mathbf{e}_t = \mathbf{R}_t \boldsymbol{\varepsilon}_t$, where \mathbf{R}_t satisfies Assumption 9, and for each t , ε_{it} is distributed independently across i and t with mean zero, finite variances $0 < \sigma_i^2 < K < \infty$, and finite fourth-order moments, $E(\varepsilon_{it}^4) = \mu_{i4} < K < \infty$. Then under Assumptions 1, and 2 we have for all sets of weights $\{w_i\}$ satisfying (13)-(14)*

$$\frac{\bar{\mathbf{e}}_w' \bar{\mathbf{e}}_w}{T} = O_p \left(\frac{1}{N} \right), \quad (45)$$

$$\frac{\mathbf{F}' \bar{\mathbf{e}}_w}{T} = O_p \left(\frac{1}{\sqrt{NT}} \right), \quad \frac{\mathbf{D}' \bar{\mathbf{e}}_w}{T} = O_p \left(\frac{1}{\sqrt{NT}} \right), \quad (46)$$

$$\frac{\mathbf{V}_i' \bar{\mathbf{e}}_w}{T} = O_p \left(\frac{1}{\sqrt{NT}} \right), \quad \frac{\mathbf{e}_i' \bar{\mathbf{e}}_w}{T} = O_p \left(\frac{1}{\sqrt{NT}} \right), \quad (47)$$

where $\bar{\mathbf{e}}_w = (\bar{e}_{w1}, \dots, \bar{e}_{wT})'$, $\bar{e}_{wt} = \sum_{i=1}^N w_i \varepsilon_{it}$, \mathbf{D} and \mathbf{F} are $T \times n$ and $T \times m$ matrices on observed and unobserved common factors, and $\mathbf{V}_i = (\mathbf{v}_{i1}, \dots, \mathbf{v}_{iT})'$.

Proof. Note that $T^{-1} \bar{\mathbf{e}}_w' \bar{\mathbf{e}}_w = T^{-1} \sum_{t=1}^T \bar{e}_{wt}^2$, by Theorem 17

$$E(T^{-1} \bar{\mathbf{e}}_w' \bar{\mathbf{e}}_w) = T^{-1} \sum_{t=1}^T E(\bar{e}_{wt}^2) = O \left(\frac{1}{N} \right). \quad (48)$$

Similarly, since ε_{it} are serially independent

$$\text{Var}(T^{-1} \bar{\mathbf{e}}_w' \bar{\mathbf{e}}_w) = T^{-2} \sum_{t=1}^T \text{Var}(\bar{e}_{wt}^2),$$

and by Theorem 17

$$\text{Var}(T^{-1} \bar{\mathbf{e}}_w' \bar{\mathbf{e}}_w) = O \left(\frac{1}{N^2 T} \right).$$

which proves (45). As for (46), consider the ℓ th row of $T^{-1} \mathbf{F}' \bar{\mathbf{e}}_w$ and note that it can be written as $T^{-1} \sum_{t=1}^T f_{\ell t} \bar{e}_{wt}$, where $f_{\ell t}$ and \bar{e}_{wt} are distributed independently of each other. Then $T^{-1} \sum_{t=1}^T f_{\ell t} \bar{e}_{wt}$ has zero mean and variance

$$\text{Var} \left(T^{-1} \sum_{t=1}^T f_{\ell t} \bar{e}_{wt} \right) = T^{-2} \sum_{t=1}^T E(f_{\ell t}^2) E(\bar{e}_{wt}^2) \leq \sigma_{\max}^2 O \left(\frac{1}{N} \right) \left\{ T^{-2} \sum_{t=1}^T E(f_{\ell t}^2) \right\} = O \left(\frac{1}{NT} \right).$$

This establishes (46). The second result in (46) and the first result in (47) follow similarly. As for the second result in (47), note that, since $w_i = O(1/N)$ and given that e_{it} is CWD

$$\begin{aligned} E(T^{-1} \mathbf{e}_i' \bar{\mathbf{e}}_w) &= T^{-1} \sum_{t=1}^T E(e_{it} \bar{e}_{wt}) \\ &= T^{-1} \sum_{t=1}^T [w_1 E(e_{it} e_{1t}) + \dots + w_N E(e_{it} e_{Nt})] = O\left(\frac{1}{N}\right) \end{aligned}$$

Denote the i^{th} row of \mathbf{R}_t by \mathbf{r}_{it} and note that⁸

$$\mathbf{r}_{it}' \mathbf{r}_{it} = \sum_{j=1}^N r_{ij,t}^2 \leq \|\mathbf{r}_{it}\|_1^2 = O(1).$$

Hence we have

$$\begin{aligned} E(e_{it}^4) &= E\left[(\mathbf{r}_{it}' \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t' \mathbf{r}_{it})^2\right] = E\left(\sum_{j=1}^N \sum_{\ell=1}^N r_{ij,t} r_{i\ell,t} \varepsilon_{jt} \varepsilon_{\ell t}\right)^2 \\ &= E\left(\sum_{j=1}^N \sum_{\ell=1}^N \sum_{k=1}^N \sum_{h=1}^N r_{ij,t} r_{i\ell,t} r_{ik,t} r_{ih,t} \varepsilon_{jt} \varepsilon_{\ell t} \varepsilon_{kt} \varepsilon_{ht}\right) \\ &= \sum_{j=1}^N \sum_{k=1}^N r_{ij,t}^2 r_{ik,t}^2 E(\varepsilon_{jt}^2 \varepsilon_{kt}^2) \leq K \left(\sum_{j=1}^N r_{ij,t}^2\right)^2 = O(1), \end{aligned}$$

and the second moment of $T^{-1} \mathbf{e}_i' \bar{\mathbf{e}}_w$ satisfies (recall that $E(\bar{e}_{wt}^4) = O(N^{-2})$)

$$\begin{aligned} E\left[(T^{-1} \mathbf{e}_i' \bar{\mathbf{e}}_w)^2\right] &= T^{-2} \sum_{t=1}^T E(e_{it}^2 \bar{e}_{wt}^2) \leq T^{-2} \sum_{t=1}^T [E(e_{it}^4) E(\bar{e}_{wt}^4)]^{1/2} \\ &= O\left(\frac{1}{N}\right) T^{-2} \sum_{t=1}^T [E(e_{it}^4)]^{1/2} = O\left(\frac{1}{NT}\right). \end{aligned}$$

■

The above results can be used to prove further results that are helpful in deriving the asymptotic distribution of CCE estimators. Let

$$\boldsymbol{\xi}_{it} = \begin{pmatrix} e_{it} + \boldsymbol{\beta}_i' \mathbf{v}_{it} \\ \mathbf{v}_{it} \end{pmatrix}.$$

⁸Let $\mathbf{x} = (x_1, \dots, x_n)'$, we define the vector norm $\|\cdot\|_1$ of \mathbf{x} as $\|\mathbf{x}\|_1 = |x_1| + \dots + |x_n|$.

From Lemma 20 it follows that (see also Lemma 2 and 3 in Pesaran (2006))

$$\begin{aligned}\frac{\bar{\boldsymbol{\xi}}_w' \bar{\boldsymbol{\xi}}_w}{T} &= O_p\left(\frac{1}{N}\right), \\ \frac{\mathbf{F}' \bar{\boldsymbol{\xi}}_w}{T} &= O_p\left(\frac{1}{\sqrt{NT}}\right), \quad \frac{\mathbf{D}' \bar{\boldsymbol{\xi}}_w}{T} = O_p\left(\frac{1}{\sqrt{NT}}\right), \\ \frac{\mathbf{V}'_i \bar{\boldsymbol{\xi}}_w}{T} &= O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right), \quad \frac{\mathbf{e}'_i \bar{\boldsymbol{\xi}}_w}{T} = O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right), \\ \frac{\mathbf{X}'_i \bar{\boldsymbol{\xi}}_w}{T} &= O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right).\end{aligned}$$

The above results prove the fundamental relations (under the assumption that the rank condition is satisfied)

$$\frac{\mathbf{X}'_i \bar{\mathbf{M}}_w \mathbf{F}}{T} = O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right), \quad (49)$$

$$\frac{\mathbf{X}'_i \bar{\mathbf{M}}_w \mathbf{X}_i}{T} = \frac{\mathbf{X}'_i \bar{\mathbf{M}}_g \mathbf{X}_i}{T} + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right), \quad (50)$$

$$\frac{\mathbf{X}'_i \bar{\mathbf{M}}_w \mathbf{e}_i}{T} = \frac{\mathbf{X}'_i \bar{\mathbf{M}}_g \mathbf{e}_i}{T} + O_p\left(\frac{1}{N}\right), \quad (51)$$

where $\bar{\mathbf{M}}_g = \mathbf{I}_T - \mathbf{G}(\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}'$. Note that (49)-(51) are identical to relations (40), (43) and (44) in Pesaran (2006), and will be used to derive the asymptotic distribution of CCE Mean Group and CCE Pooled estimators.

Proof of Theorem 18. Consider

$$\sqrt{N}(\mathbf{b}_{MG} - \boldsymbol{\beta}) = \frac{1}{\sqrt{N}} \sum_{i=1}^N \mathbf{v}_i + \frac{1}{N} \sum_{i=1}^N \hat{\boldsymbol{\Psi}}_{iT}^{-1} \left(\frac{\sqrt{N} \mathbf{X}'_i \bar{\mathbf{M}}_w \mathbf{F}}{T} \right) \gamma_i + \frac{1}{N} \sum_{i=1}^N \hat{\boldsymbol{\Psi}}_{iT}^{-1} \left(\frac{\sqrt{N} \mathbf{X}'_i \bar{\mathbf{M}}_w \mathbf{e}_i}{T} \right),$$

where by assumption $\hat{\boldsymbol{\Psi}}_{iT}^{-1} = (T^{-1} \mathbf{X}'_i \bar{\mathbf{M}}_w \mathbf{X}_i)^{-1}$ has second-order moments. Given (49) it follows that

$$\frac{\sqrt{N} \mathbf{X}'_i \bar{\mathbf{M}}_w \mathbf{F}}{T} = O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right), \quad (52)$$

and hence, since by assumption factor loadings are bounded, we have (Pesaran, 2006, p. 983)

$$\frac{1}{N} \sum_{i=1}^N \hat{\boldsymbol{\Psi}}_{iT}^{-1} \left(\frac{\sqrt{N} \mathbf{X}'_i \bar{\mathbf{M}}_w \mathbf{F}}{T} \right) \gamma_i \xrightarrow{p} \mathbf{0}, \quad \text{as } (N, T) \xrightarrow{j} \infty.$$

From (50)-(51) we have

$$\begin{aligned}\frac{1}{N} \sum_{i=1}^N \hat{\boldsymbol{\Psi}}_{iT}^{-1} \left(\frac{\sqrt{N} \mathbf{X}'_i \bar{\mathbf{M}}_w \mathbf{e}_i}{T} \right) &= \frac{1}{N} \sum_{i=1}^N \left(\frac{\mathbf{X}'_i \bar{\mathbf{M}}_g \mathbf{X}_i}{T} \right)^{-1} \left(\frac{\sqrt{N} \mathbf{X}'_i \bar{\mathbf{M}}_g \mathbf{e}_i}{T} \right) + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right) \\ &= \boldsymbol{\Delta}_{NT} + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right),\end{aligned}$$

where

$$\begin{aligned}\Delta_{NT} &= \frac{1}{\sqrt{N}} \sum_{i=1}^N \left(\frac{\mathbf{X}'_i \mathbf{M}_g \mathbf{X}_i}{T} \right)^{-1} \left(\frac{\mathbf{X}'_i \mathbf{M}_g \mathbf{e}_i}{T} \right) \\ &= \frac{1}{\sqrt{NT}} \sum_{i=1}^N \left(\frac{\mathbf{V}'_i \mathbf{M}_g \mathbf{V}_i}{T} \right)^{-1} \left(\frac{\mathbf{V}'_i \mathbf{M}_g \mathbf{e}_i}{\sqrt{T}} \right).\end{aligned}$$

Δ_{NT} has zero mean and variance

$$\begin{aligned}\text{Var}(\Delta_{NT}) &= \frac{1}{NT} \sum_{i=1}^N \sum_{j=1}^N E \left[\left(\frac{\mathbf{V}'_i \mathbf{M}_g \mathbf{V}_i}{T} \right)^{-1} \left(\frac{\mathbf{V}_i \mathbf{M}_g \mathbf{e}_i \mathbf{e}'_j \mathbf{M}_g \mathbf{V}_j}{T} \right) \left(\frac{\mathbf{V}'_j \mathbf{M}_g \mathbf{V}_j}{T} \right)^{-1} \right] \\ &= \frac{1}{NT} \sum_{i=1}^N \sum_{j=1}^N \Sigma_i^{-1} E \left(\frac{\mathbf{V}'_i \mathbf{M}_g \mathbf{e}_i \mathbf{e}'_j \mathbf{M}_g \mathbf{V}_j}{T} \right) \Sigma_j^{-1}.\end{aligned}$$

Consider the $(t, s)^{th}$ element of the $T \times T$ matrix $E(\mathbf{e}_i \mathbf{e}'_j)$ and note that

$$E(e_{it} e_{js}) = E(\mathbf{r}'_{it} \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_s \mathbf{r}_{jt}) = \begin{cases} \mathbf{r}'_{it} \boldsymbol{\Lambda}_\varepsilon \mathbf{r}_{jt}, & \text{for } t = s \\ 0, & \text{otherwise} \end{cases}$$

where \mathbf{r}_{it} is the i th row of the matrix \mathbf{R}_t . Given that the matrix \mathbf{R}_t has bounded absolute column and row sums, it follows that the diagonal elements of the matrix $E(\mathbf{e}_i \mathbf{e}'_j)$ satisfy

$$\mathbf{r}'_{it} \boldsymbol{\Lambda}_\varepsilon \mathbf{r}_{jt} \leq \sigma_{\max}^2 \|\mathbf{r}_{it}\|_1 \|\mathbf{r}_{jt}\|_1 = O(1), \quad \text{for } t = 1, 2, \dots, T.$$

Hence, since \mathbf{V}_i are independent across i , we have

$$\text{Var}(\Delta_{NT}) \leq \frac{1}{NT} K \sum_{i=1}^N \sum_{j=1}^N \Sigma_i^{-1} E \left(\frac{\mathbf{V}'_i \mathbf{M}_g \mathbf{V}_j}{T} \right) \Sigma_j^{-1} = \frac{1}{NT} K \sum_{i=1}^N \Sigma_i^{-1} = O\left(\frac{1}{T}\right).$$

Therefore

$$\sqrt{N}(\mathbf{b}_{MG} - \boldsymbol{\beta}) = \frac{1}{\sqrt{N}} \sum_{i=1}^N \mathbf{v}_i + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right),$$

and

$$\sqrt{N}(\mathbf{b}_{MG} - \boldsymbol{\beta}) \xrightarrow{d} N(0, \boldsymbol{\Sigma}_{MG}) \quad \text{as } (N, T) \xrightarrow{j} \infty.$$

■

Proof of Theorem 19. Consider

$$\left(\sum_{i=1}^N w_i^2 \right)^{-1/2} (\hat{\mathbf{b}}_P - \boldsymbol{\beta}) = \left(\sum_{i=1}^N w_i \frac{\mathbf{X}'_i \bar{\mathbf{M}}_w \mathbf{X}_i}{T} \right)^{-1} \left(\frac{1}{\sqrt{N}} \sum_{i=1}^N \tilde{w}_i \frac{\mathbf{X}'_i \bar{\mathbf{M}}_w (\mathbf{X}_i \mathbf{v}_i + \mathbf{e}_i)}{T} + \mathbf{q}_{NT} \right),$$

where

$$\begin{aligned}\tilde{w}_i &= \left(N^{-1} \sum_{i=1}^N w_i^2 \right)^{-1/2} \quad w_i = O(1), \\ \mathbf{q}_{NT} &= \frac{1}{\sqrt{N}} \sum_{i=1}^N \tilde{w}_i \frac{\mathbf{X}'_i \bar{\mathbf{M}}_w \mathbf{F} \gamma_i}{T}.\end{aligned}$$

Using (49) and (52)

$$\mathbf{q}_{NT} = \frac{1}{N} \sum_{i=1}^N \tilde{w}_i \frac{\sqrt{N} \mathbf{X}'_i \bar{\mathbf{M}}_w \mathbf{F} \gamma_i}{T} \xrightarrow{p} \mathbf{0} \quad \text{as } (N, T) \xrightarrow{j} \infty.$$

Hence,

$$\begin{aligned}\left(\sum_{i=1}^N w_i^2 \right)^{-1/2} (\hat{\mathbf{b}}_P - \boldsymbol{\beta}) &= \left(\sum_{i=1}^N w_i \frac{\mathbf{X}'_i \bar{\mathbf{M}}_g \mathbf{X}_i}{T} \right)^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^N \tilde{w}_i \frac{\mathbf{X}'_i \bar{\mathbf{M}}_g (\mathbf{X}_i \mathbf{v}_i + \mathbf{e}_i)}{T} \\ &\quad + O_p \left(\frac{1}{\sqrt{N}} \right) + O_p \left(\frac{1}{\sqrt{T}} \right).\end{aligned}$$

Note that the element

$$\left(\sum_{i=1}^N w_i \frac{\mathbf{X}'_i \bar{\mathbf{M}}_g \mathbf{X}_i}{T} \right)^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^N \frac{\mathbf{X}'_i \bar{\mathbf{M}}_g \mathbf{e}_i}{T},$$

has mean zero and variance

$$\begin{aligned}&\left(\sum_{i=1}^N w_i \frac{\mathbf{X}'_i \bar{\mathbf{M}}_g \mathbf{X}_i}{T} \right)^{-1} \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^N \tilde{w}_i \tilde{w}_j E \left(\frac{\mathbf{V}'_i \bar{\mathbf{M}}_g \mathbf{e}_i \mathbf{e}'_j \bar{\mathbf{M}}_g \mathbf{V}_j}{T} \right) \left(\sum_{i=1}^N w_i \frac{\mathbf{X}'_i \bar{\mathbf{M}}_g \mathbf{X}_i}{T} \right)^{-1} \\ &= O \left(\frac{1}{T} \right),\end{aligned}$$

considering that $E(\mathbf{e}_i \mathbf{e}'_j)$ is a diagonal matrix with finite elements, and \mathbf{V}_i is independently distributed across i . It then follows that

$$\left(\sum_{i=1}^N w_i^2 \right)^{-1/2} (\hat{\mathbf{b}}_P - \boldsymbol{\beta}) \rightarrow N(0, \boldsymbol{\Sigma}_P).$$

■

References

- [1] Anselin, L. (1988), *Spatial Econometrics: Methods and Models*, Kluwer Academic Publishers, Dordrecht, the Netherlands.
- [2] Anselin, L. (2002), Under the hood. Issues in the specification and interpretation of spatial regression models, *Agricultural Economics*, 27, 247-267.
- [3] Bai, J., and Ng, S. (2004), A PANIC Attack on Unit Roots and Cointegration, *Econometrica*, 72, 1127-1177.
- [4] Bai, J. (2005), Panel Data Models with Interactive Fixed Effects, Unpublished Manuscript, New York University.
- [5] Bernstein, D.S. (2005), *Matrix Mathematics*, Princeton University Press.
- [6] Breitung, J., and Pesaran M.H. (2006), Unit Roots and Cointegration in Panels, in L. Matyas and Sevestre P. (eds.) *The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice*, Kluwer Academic Publishers, forthcoming.
- [7] Brock, W., and Durlauf, S. (2001), Interactions-Based Models, *Handbook of Econometrics*, Volume V, Heckman, J., Leamer, E., (eds), North Holland, Amsterdam.
- [8] Brueckner, J.K. (2003), Strategic Interaction Among Governments: an Overview of Empirical Studies, *International Regional Science Review*, 26, 175-188.
- [9] Case, A.C. (1991), Spatial Pattern in Household Demand, *Econometrica*, 59, 953-965.
- [10] Chamberlain, G. (1983), Funds, Factors, and Diversification in Arbitrage Pricing Models, *Econometrica*, 51, 1281-1304.
- [11] Chamberlain, G., and Rothschild, M. (1983), Arbitrage, Factor Structure and Mean-variance Analysis in Large Asset Market, *Econometrica*, 51, 1305-1324.
- [12] Chang, Y. (2004), Bootstrap Unit Root Test in Panels with Cross-Sectional Dependency, *Journal of Econometrics*, 120, 263-293.
- [13] Choi, I. (2002), Combination Unit Root Tests for Cross-sectionally Correlated Panels, in *Econometric Theory and Practice: Frontiers of Analysis and Applied Research, Essays in Honor of Peter C.,B., Phillips*, Cambridge University Press.
- [14] Cliff, A., and Ord, J.K. (1973), *Spatial Autocorrelation*, Pion, London.
- [15] Cliff, A., and Ord, J.K. (1981), *Spatial Processes: Models and Applications*, Pion, London.
- [16] Coakley, J., Fuertes, A.M., and Smith, R. (2002), A Principal Components Approach to Cross-Section Dependence in Panels, Birkbeck College Discussion Paper 01/2002.

- [17] Coakley, J., Fuertes, A.M., and Smith, R. (2006), Unobserved Heterogeneity in Panel Time Series., *Computational Statistics and Data Analysis*, 50, 2361-2380.
- [18] Conley, T.G. (1999), GMM Estimation with Cross Sectional Dependence, *Journal of Econometrics*, 92, 1-45.
- [19] Conley, T.G., and Topa, G. (2002), Socio-economic Distance and Spatial Patterns in Unemployment, *Journal of Applied Econometrics*, 17, 303-327.
- [20] Connor, G., and Korajczyk, R.A. (1993), A Test for the Number of Factors in an Approximate Factor Structure, *The Journal of Finance*, 48, 1263-1291.
- [21] Cressie, N. (1993), *Statistics for Spatial Data*, Wiley, New York.
- [22] Forni, M., and Lippi, M. (2001), The Generalized Factor Model: Representation Theory, *Econometric Theory*, 17, 1113-1141.
- [23] Forni, M., and Reichlin, L. (1998), Let's Get Real: a Factor Analytical Approach to Disaggregated Business Cycle Dynamics, *Review of Economic Studies*, 65, 453-473.
- [24] Gregory, A., and Head, A. (1999), Common and Country Specific Fluctuations in Productivity, Investment, and the Current Account, *Journal of Monetary Economics*, 44, 423-452.
- [25] Haining, R.P. (2003), *Spatial Data Analysis: Theory and Practice*, University Press, Cambridge.
- [26] Haining, R.P. (1978), The Moving Average Model for Spatial Interaction, *Transactions of the Institute of British Geographers*, 3, 202-225.
- [27] Horn, R.A., and Johnson, C.A. (1985), *Matrix Analysis*, Cambridge University Press.
- [28] Im, K.S., Pesaran, M.H., and Shin, Y. (2003), Testing for Unit Roots in Heterogeneous Panels, *Journal of Econometrics*, 115, 53-74.
- [29] Kapetanios, G. and Marcellino, M. (2006), Factor-GMM Estimation with Large Sets of Possibly Weak Instruments, Unpublished Manuscript, Queen Mary College, University of London.
- [30] Kapetanios, G., and Pesaran, M.H. (2007), Alternative Approaches to Estimation and Inference in Large Multifactor Panels: Small Sample Results with an Application to Modelling of Asset Returns, in Phillips, G., Tzavalis, E. (Eds.), *The Refinement of Econometric Estimation and Test Procedures: Finite Sample and Asymptotic Analysis*, Cambridge University Press, Cambridge.
- [31] Kapetanios, G., Pesaran, M.H., and Yagamata, T. (2006), Panels with Nonstationary Multifactor Error Structures, *CESifo Working Paper* No. 1788.
- [32] Kelejian, H.H., and Prucha, I. (2001), On the Asymptotic Distribution of the Moran I Test Statistic with Applications, *Journal of Econometrics*, 104, 219-257.
- [33] Kelejian, H.H., and Prucha, I. (1999), A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Models, *International Economic Review*, 40, 509-533.

- [34] Kelejian, H.H., and Robinson, D.P. (1995), Spatial correlation: A suggested Alternative to the Autoregressive Model, in Anselin, L., and Florax, R.J. (eds), *New Directions in Spatial Econometrics*, 75-95, Springer-Verlag, Berlin.
- [35] Kelejian, H., Kapoor, M., and Prucha, I. (2006), Panel Data Models with Spatially Correlated Error Components, *Journal of Econometrics*, in Press.
- [36] Lee, L.F. (2002), Consistency and Efficiency of Least Squares Estimation for Mixed Regressive, Spatial Autoregressive Models, *Econometric Theory*, 18, 252-277.
- [37] Lee, L.F. (2004), Asymptotic Distributions of Quasi-Maximum Likelihood Estimators for Spatial Autoregressive Models, *Econometrica*, 72, 1899-1925.
- [38] Levin, A., Lin, C.F., and Chu, J. (2002), Unit Root Tests in Panel Data: Asymptotic and Finite Sample Properties, *Journal of Econometrics*, 108, 1-24.
- [39] Mardia, K.V., and Marshall, R.J. (1984), Maximum Likelihood Estimation of Models for Residual Covariance in Spatial Regression, *Biometrika*, 71, 135-146.
- [40] Moon, H.R., and Perron B. (2004), Testing for a Unit Root in Panels with Dynamic Factors, *Journal of Econometrics*, 122, 81-126.
- [41] O'Connell, P.G.J. (1998), The Overvaluation of Purchasing Power Parity, *Journal of International Economics*, 44, 1-19.
- [42] Pesaran, M.H. (2004), General Diagnostic Tests for Cross Section Dependence in Panels, *CEPrifo Working Paper Series* n. 1229, *IZA Discussion Paper* No. 1240.
- [43] Pesaran, M.H. (2006), Estimation and Inference in Large Heterogenous Panels with Multifactor Error Structure, *Econometrica*, 74, 967-1012.
- [44] Pesaran, M.H. (2007), A Simple Panel Unit Root Test in the Presence of Cross Section Dependence, *Journal of Applied Econometrics*, 22, 265-312.
- [45] Pesaran, M.H., and Chudik, A. (2007), Infinite Dimensional VARs, Paper under preparation, University of Cambridge.
- [46] Pesaran, M.,H., Schuermann, T., and Weiner, S. (2004), Modelling Regional Interdependencies using a Global Error-Correcting Macroeconometric Model, *Journal of Business and Economics Statistics*, 22, 129-162.
- [47] Phillips, P.C.B., and Sul, D. (2003), Dynamic Panel Estimation and Homogeneity Testing Under Cross Section Dependence, *The Econometrics Journal*, 6, 217-259.
- [48] Pinkse, J., Slade, M.E., and Brett, C. (2002), Spatial Price Competition, *Econometrica*, 70, 1111-1153.
- [49] Rivera-Batiz, L.A., and Romer, P.M. (1991), Economic Integration and Endogenous Growth, *The Quarterly Journal of Economics*, 106, 531-555.

- [50] Robertson, D., and Symons, J. (2000), Factor Residuals in SUR Regressions: Estimating Panels allowing for Cross Sectional Correlation, Unpublished Manuscript, Faculty of Economics and Politics, University of Cambridge.
- [51] Robinson, P.M. (2003), Long Memory Time Series. In *Time Series with Long Memory*. Edited by Robinson, P. M.. Oxford University Press, pp. 4-32.
- [52] Spearman, C. (1904), General Intelligence Objectively Determined and Measured, *American Journal of Psychology*, 15, 201-293.
- [53] Stock, J.H., and Watson, M.W. (2002), Macroeconomic Forecasting Using Diffusion Indexes, *Journal of Business & Economic Statistics*, 20, 147-162.
- [54] Stone, R.N. (1947), On the Interdependence of Blocks of Transactions, *Supplement to the Journal of the Royal Statistical Society*, 9, 1-45.
- [55] Whittle, P. (1954), On Stationary Processes on the Plane, *Biometrika*, 41, 434-449.
- [56] 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.

Table 1: Small Sample Properties of CCE Type Estimators in Experiments A

N	Bias ($\times 100$)						RMSE ($\times 100$)					
	T						T					
	10	20	30	50	100	200	10	20	30	50	100	200
CCE Mean Group												
20	0.68	0.00	-0.07	-0.08	0.04	0.02	21.54	8.15	5.86	4.29	2.96	2.21
30	0.12	-0.05	0.05	0.03	-0.02	0.04	18.40	6.69	4.81	3.30	2.22	1.66
50	-0.26	0.00	0.12	-0.02	0.03	0.06	14.35	5.17	3.61	2.51	1.62	1.17
100	-0.19	-0.02	0.02	-0.02	0.02	0.01	10.39	3.65	2.56	1.77	1.11	0.74
200	0.13	0.06	-0.00	-0.03	0.01	-0.00	7.55	2.70	1.85	1.21	0.78	0.53
CCE Pooled												
20	0.19	-0.01	-0.02	-0.06	-0.00	0.03	12.06	6.92	5.14	4.03	2.96	2.28
30	0.05	-0.06	-0.04	0.06	-0.03	0.03	10.14	5.48	4.15	3.06	2.16	1.68
50	-0.36	-0.02	0.09	-0.04	0.03	0.07	7.91	4.24	3.15	2.26	1.58	1.16
100	0.02	0.06	0.07	-0.01	0.01	0.01	5.70	2.95	2.23	1.63	1.06	0.72
200	0.09	0.02	0.006	-0.01	0.01	-0.00	4.01	2.09	1.59	1.11	0.74	0.52
Fixed Effects Estimator												
20	17.43	17.23	17.13	17.52	17.20	17.48	23.44	20.98	20.03	19.62	18.70	18.67
30	16.62	17.00	17.08	17.41	17.55	17.51	23.34	20.90	20.03	19.36	18.83	18.53
50	17.50	18.14	18.41	18.34	18.33	18.10	21.65	20.62	20.17	19.67	19.17	18.75
100	16.57	17.26	17.09	17.59	17.25	17.11	21.08	19.86	18.90	18.79	17.95	17.56
200	16.33	17.24	17.28	17.35	17.37	17.35	20.31	19.65	18.91	18.41	17.97	17.69

N	Size (5% level) ($\times 100$)($H_0: \beta_1 = 1$)						Power (5% level) ($\times 100$)($H_1: \beta_1 = 0.95$)					
	T						T					
	10	20	30	50	100	200	10	20	30	50	100	200
CCE Mean Group												
20	5.90	6.45	6.45	8.25	7.25	7.25	6.65	12.50	17.25	26.85	45.90	67.30
30	6.05	6.25	6.95	6.85	5.85	6.80	7.15	13.15	20.80	35.70	62.95	87.15
50	4.80	6.15	5.50	5.05	5.70	6.20	6.90	17.15	31.30	52.70	86.60	98.85
100	5.35	5.10	5.60	5.85	5.35	3.85	7.50	27.80	50.65	81.90	99.25	100.00
200	5.15	5.25	5.75	4.65	5.15	4.75	10.75	48.50	78.80	98.30	100.00	100.00
CCE Pooled												
20	5.30	6.60	6.80	7.70	7.90	7.45	6.75	14.30	19.15	30.05	46.95	66.25
30	5.30	6.30	6.65	6.45	6.30	6.50	7.90	16.60	24.05	39.60	65.90	85.70
50	4.95	5.85	5.95	5.40	5.95	5.95	9.80	23.90	38.20	58.75	88.55	98.85
100	5.50	5.00	5.45	5.95	5.40	4.55	15.35	39.40	63.15	86.80	99.55	100.00
200	5.60	4.95	5.35	5.20	5.00	5.10	25.85	65.85	88.65	99.35	100.00	100.00
Fixed Effects estimator												
20	50.25	68.20	79.00	89.70	96.20	98.75	63.40	82.80	91.50	97.05	99.10	99.80
30	59.95	75.25	83.65	93.05	98.15	99.45	70.50	86.55	92.70	98.05	99.80	99.75
50	70.55	87.80	95.65	98.70	99.90	100.00	84.40	96.70	99.40	99.80	100.00	100.00
100	80.00	92.05	96.20	99.50	99.90	100.00	89.25	97.75	99.50	100.00	100.00	100.00
200	86.25	95.45	98.20	99.70	100.00	100.00	93.80	99.00	99.90	100.00	100.00	100.00

Note: The model has common factors but no spatial dependence. CCE Mean Group, CCE Pooled and the fixed effects estimator are defined by (39), (42) and (44).

Variances of CCE Mean Group and CCE Pooled are, respectively, (41) and (43).

Table 2: Size and Power of CCE Type Estimators in Experiments B

N	Size (5% level) ($\times 100$)($H_0: \beta_1 = 1$)						Power (5% level) ($\times 100$)($H_1: \beta_1 = 0.95$)					
	T						T					
	$\delta = 0.8$											
	10	20	30	50	100	200	10	20	30	50	100	200
CCE Mean Group												
20	6.55	7.00	6.55	5.55	6.65	6.55	6.40	9.35	11.40	16.85	31.60	53.65
30	6.15	5.75	6.00	6.05	5.70	6.80	6.10	9.25	11.50	18.60	22.00	64.45
50	4.60	4.65	5.30	6.40	4.45	6.45	5.90	9.25	12.75	23.15	48.20	80.95
100	5.15	4.90	5.70	4.70	5.20	6.05	5.65	12.60	22.00	40.50	73.10	97.40
200	4.75	5.30	5.50	4.60	5.05	5.00	6.20	20.65	35.55	61.80	95.60	100.00
CCE Pooled												
20	4.60	6.70	7.05	6.35	7.40	7.15	5.55	9.15	11.95	17.65	33.85	54.45
30	5.20	5.80	6.75	5.70	5.30	7.30	6.65	10.05	13.05	20.30	26.55	66.15
50	5.15	5.50	5.15	6.85	5.05	6.10	6.85	11.15	15.70	26.90	52.00	81.70
100	5.45	5.05	5.30	5.65	4.95	5.55	8.45	15.95	26.55	45.40	76.85	97.60
200	5.30	5.60	5.20	4.75	4.75	4.90	11.55	29.00	44.25	70.80	96.60	100.00
Fixed Effects estimator												
20	10.65	11.05	11.00	9.65	10.85	10.60	13.40	15.90	18.60	22.65	38.15	55.70
30	16.50	17.15	17.45	17.15	13.35	18.80	18.55	22.55	25.55	33.55	46.15	68.80
50	8.55	8.30	8.25	9.55	8.05	9.60	14.35	18.75	24.25	37.55	61.80	87.65
100	13.60	14.35	13.35	13.10	14.75	14.30	23.55	35.00	46.15	65.30	87.30	98.25
200	12.90	13.70	12.35	13.20	12.80	13.85	30.85	52.35	68.30	85.70	98.75	100.00

N	Size (5% level) ($\times 100$)($H_0: \beta_1 = 1$)						Power (5% level) ($\times 100$)($H_1: \beta_1 = 0.95$)					
	T						T					
	$\delta = 0.4$											
	10	20	30	50	100	200	10	20	30	50	100	200
CCE Mean Group												
20	5.55	7.35	6.50	6.35	5.75	6.90	6.75	12.00	15.90	25.90	51.80	80.45
30	5.80	6.05	5.90	5.45	6.90	6.50	7.15	13.50	19.40	34.60	64.30	93.40
50	4.80	4.50	4.55	6.15	5.50	5.90	6.15	14.80	25.15	47.75	84.50	99.00
100	4.65	5.00	5.45	5.20	5.65	5.25	7.75	26.25	44.80	77.10	98.45	100.00
200	5.15	5.60	5.80	4.65	5.60	4.75	9.75	44.15	73.65	96.85	100.00	100.00
CCE Pooled												
20	4.60	6.80	6.85	6.50	6.55	7.05	6.40	12.30	17.80	26.90	53.90	81.00
30	5.25	5.75	6.00	5.60	7.05	6.60	7.75	14.60	22.30	38.50	67.70	94.15
50	5.00	5.55	5.95	6.25	5.55	5.95	9.05	21.05	32.05	54.45	87.30	99.20
100	4.50	5.20	5.50	4.85	5.30	5.25	13.80	34.20	55.10	83.90	99.00	100.00
200	5.30	4.95	5.45	4.65	5.10	4.60	21.55	60.90	85.40	98.75	100.00	100.00
Fixed Effects estimator												
20	6.85	6.70	7.40	6.15	6.60	5.95	13.30	21.45	27.00	41.85	72.45	94.20
30	8.40	7.05	7.10	6.85	7.90	7.95	15.65	27.50	37.45	57.75	84.35	99.25
50	6.30	5.35	5.45	6.55	5.70	6.25	20.90	37.10	50.25	75.55	97.00	100.00
100	6.95	6.90	7.35	6.65	7.10	7.35	34.50	65.15	83.40	96.85	99.95	100.00
200	6.70	6.70	6.30	6.45	7.00	7.15	56.05	90.35	98.65	99.90	100.00	100.00

Note: The model has common factors but no spatial dependence. CCE Mean Group, CCE Pooled and the fixed effects estimator are defined by (39), (42) and (44).

Variances of CCE Mean Group and CCE Pooled are, respectively, (41) and (43).

Table 3: Size and Power of CCE Type Estimators in Experiments C

N	Size (5% level) ($\times 100$)($H_0: \beta_1 = 1$)						Power (5% level) ($\times 100$)($H_1: \beta_1 = 0.95$)					
	T						T					
	$\delta = 0.8$											
	10	20	30	50	100	200	20	30	50	100	200	
CCE Mean Group												
20	5.95	6.30	7.80	7.85	6.30	6.65	6.85	9.30	13.25	17.40	30.20	46.90
30	5.70	5.60	6.05	6.50	5.75	6.65	6.50	8.90	12.25	19.35	35.00	62.50
50	5.60	5.75	5.25	5.80	5.30	5.80	6.65	9.25	15.30	24.25	49.05	79.05
100	5.55	5.20	5.25	5.35	4.60	4.95	5.45	12.25	22.00	37.50	73.75	96.80
200	5.55	5.20	5.50	5.05	5.60	4.80	7.45	19.00	33.70	62.55	95.50	99.95
CCE Pooled												
20	4.85	6.85	7.05	7.50	6.60	7.05	5.95	10.60	14.10	19.55	31.40	48.40
30	5.35	6.35	5.85	6.80	6.15	6.80	6.60	10.95	13.85	20.90	36.95	63.00
50	5.80	4.95	5.60	5.45	5.55	5.75	7.40	12.10	17.80	27.20	52.40	80.90
100	4.65	5.20	5.55	5.55	4.60	4.80	8.00	17.00	27.40	43.30	77.90	97.85
200	5.50	5.25	4.15	4.50	5.65	5.35	12.75	27.70	44.25	70.55	96.00	99.95
Fixed Effects estimator												
20	40.30	56.15	67.10	80.75	92.60	97.75	48.65	68.95	81.35	92.50	98.55	99.45
30	48.55	64.80	74.70	86.20	96.00	99.05	57.70	77.10	86.25	95.35	99.35	99.70
50	58.80	78.65	88.80	95.80	99.70	100.00	70.10	90.05	97.40	99.50	99.95	100.00
100	70.95	86.65	93.05	98.95	99.80	100.00	81.50	95.20	98.80	99.85	100.00	100.00
200	80.05	92.05	96.70	99.50	100.00	100.00	90.00	98.10	99.75	100.00	100.00	100.00

N	Size (5% level) ($\times 100$)($H_0: \beta_1 = 1$)						Power (5% level) ($\times 100$)($H_1: \beta_1 = 0.95$)					
	T						T					
	$\delta = 0.4$											
	10	20	30	50	100	200	10	20	30	50	100	200
CCE Mean Group												
20	5.90	6.85	7.15	8.40	7.05	7.35	7.05	12.05	17.15	25.80	44.45	65.30
30	5.00	5.95	7.20	6.75	6.00	6.75	6.85	12.55	19.15	33.20	59.80	85.55
50	5.90	5.75	5.05	5.20	5.80	6.10	7.35	15.40	27.80	47.90	83.40	98.25
100	5.50	5.20	5.30	5.95	4.90	4.25	6.85	24.25	45.25	76.25	98.75	100.00
200	5.65	5.45	5.85	4.70	5.60	5.05	10.35	42.45	71.55	96.70	100.00	100.00
CCE Pooled												
20	5.05	6.85	7.10	7.80	7.25	7.10	7.05	13.75	19.10	28.20	45.15	63.80
30	5.70	6.80	6.60	6.65	5.85	7.00	7.65	15.90	22.95	37.25	62.30	83.90
50	5.75	5.40	6.00	5.45	5.65	5.95	10.00	21.05	33.85	54.60	85.80	98.35
100	5.15	5.05	5.45	5.90	5.40	4.30	14.45	35.70	56.95	82.65	99.00	100.00
200	5.90	5.05	5.30	4.45	5.60	5.05	23.40	58.50	83.85	98.35	100.00	100.00
Fixed Effects estimator												
20	49.25	67.45	78.15	89.55	96.05	98.65	61.50	81.25	91.05	96.90	99.25	99.75
30	57.70	74.05	82.70	92.60	97.90	99.45	69.45	86.60	92.45	98.00	99.80	99.75
50	69.25	87.15	95.40	98.50	99.90	100.00	82.15	95.95	99.30	99.85	100.00	100.00
100	79.15	91.60	96.10	99.45	99.90	100.00	88.70	97.60	99.45	100.00	100.00	100.00
200	85.80	95.40	98.00	99.70	100.00	100.00	93.80	98.95	99.90	100.00	100.00	100.00

Note: The model has common factors but no spatial dependence. CCE Mean Group, CCE Pooled and the fixed effects estimator are defined by (39), (42) and (44).

Variances of CCE Mean Group and CCE Pooled are, respectively, (41) and (43).

Table 4: Size and Power of CCE Type Estimators in Experiments D

N	Size (5% level) ($\times 100$)($H_0: \beta_1 = 1$)						Power (5% level) ($\times 100$)($H_1: \beta_1 = 0.95$)					
	T						T					
	Experiment D1											
	10	20	30	50	100	200	10	20	30	50	100	200
CCE Mean Group												
20	6.80	6.00	6.60	7.60	6.95	6.94	7.00	13.75	17.85	27.10	48.30	68.37
30	6.40	5.60	6.50	6.05	6.00	5.60	7.05	14.20	21.30	35.75	65.05	87.60
50	5.70	5.55	5.55	5.85	5.25	6.00	7.40	16.95	29.75	53.30	85.15	98.90
100	4.25	5.55	5.60	5.00	5.50	4.65	7.85	27.90	48.10	79.05	99.05	100.00
200	4.31	4.91	5.30	5.80	5.29	4.90	11.23	47.81	77.65	96.30	100.00	100.00
CCE Pooled												
20	6.30	6.80	7.00	7.25	7.70	7.76	8.00	15.05	19.75	28.80	50.35	69.08
30	5.65	5.90	6.40	6.65	6.00	5.85	6.65	18.05	24.20	40.80	67.25	88.40
50	4.15	6.00	5.70	5.60	4.80	6.40	10.30	22.80	36.20	58.80	87.80	0.98
100	5.10	6.05	4.95	5.60	5.55	4.55	15.05	37.95	59.60	85.60	99.50	100.00
200	5.57	4.70	4.70	5.75	4.97	4.72	23.48	64.37	87.95	98.15	100.00	100.00
Fixed Effects estimator												
20	36.90	71.00	80.70	81.85	94.05	97.35	46.65	88.65	89.75	95.15	99.10	99.74
30	49.20	73.05	75.55	87.55	94.85	99.05	62.75	78.00	90.05	96.45	99.20	99.95
50	60.55	80.45	89.40	96.10	99.70	100.00	77.15	94.10	97.75	99.55	100.00	100.00
100	71.40	85.05	91.95	97.65	99.55	99.95	84.95	95.90	98.25	99.90	100.00	100.00
200	79.05	90.76	95.45	97.70	100.00	100.00	89.95	97.81	99.70	99.85	100.00	100.00

N	Size (5% level) ($\times 100$)($H_0: \beta_1 = 1$)						Power (5% level) ($\times 100$)($H_1: \beta_1 = 0.95$)					
	T						T					
	Experiment D2											
	10	20	30	50	100	200	10	20	30	50	100	200
CCE Mean Group												
20	6.55	6.70	7.10	7.50	7.65	7.15	7.60	11.85	17.75	27.90	49.30	77.15
30	5.30	6.30	6.20	5.95	6.55	6.00	6.15	13.55	21.05	34.05	66.50	92.00
50	5.15	5.55	5.35	5.95	5.70	5.55	7.10	16.55	27.25	50.45	84.75	99.40
100	4.95	4.70	5.25	5.05	5.95	5.19	7.90	25.10	45.60	77.95	99.15	100.00
200	4.90	5.35	6.10	5.10	5.20	5.35	10.85	44.90	74.45	97.21	100.00	100.00
CCE Pooled												
20	6.35	7.05	6.85	7.15	7.60	7.05	8.10	14.15	19.35	30.35	51.70	76.85
30	4.95	5.55	6.50	6.90	6.45	5.80	7.55	17.15	25.10	36.80	68.00	92.55
50	5.85	5.95	6.20	5.95	5.75	6.00	9.10	22.25	34.35	56.60	86.35	99.45
100	4.85	4.90	5.55	5.30	5.90	5.43	13.95	37.05	56.25	84.10	99.35	100.00
200	5.15	6.15	5.35	5.24	5.45	4.95	23.30	62.00	87.25	99.18	100.00	100.00
Fixed Effects estimator												
20	29.20	35.10	49.00	57.30	78.90	91.40	44.45	58.25	74.40	86.55	97.25	99.55
30	40.20	45.50	57.10	66.15	85.35	94.10	52.20	67.60	80.15	89.40	98.80	99.90
50	47.30	54.75	68.90	81.65	93.70	98.90	66.55	80.35	92.15	98.35	99.90	100.00
100	59.85	66.00	76.00	86.20	96.30	99.64	78.40	88.75	95.70	98.90	99.95	100.00
200	68.60	73.55	85.05	92.93	98.80	100.00	87.25	94.40	97.80	99.73	100.00	100.00

Note: The model has common factors but no spatial dependence. CCE Mean Group, CCE Pooled and the fixed effects estimator are defined by (39), (42) and (44).

Variances of CCE Mean Group and CCE Pooled are, respectively, (41) and (43).

Table 5: CD, CD(p), and Average Correlation Coefficient in Experiments A, B and C

N	CD					CD(p)					\bar{r} ($\times 100$)							
	T					T					T							
	10	20	30	50	100	200	10	20	30	50	100	200	10	20	30	50	100	200
Experiment A: common factors																		
20	22.36	33.63	42.11	55.37	79.14	112.54	7.04	10.27	13.12	17.51	49.80	70.91	51.29	54.55	55.77	56.81	57.41	57.73
30	33.81	50.65	63.69	83.12	119.58	170.80	11.43	25.50	32.24	42.90	61.41	87.95	51.27	54.31	55.75	56.36	57.34	57.91
50	56.06	85.02	106.74	140.20	201.07	285.95	14.18	33.43	41.89	55.65	80.40	114.98	50.65	54.32	55.68	56.65	57.45	57.77
100	113.54	172.06	214.41	282.65	404.78	576.21	24.71	47.47	59.63	79.50	114.41	162.33	51.03	54.68	55.64	56.82	57.53	57.91
200	227.64	346.92	431.96	565.41	811.91	1155.43	34.40	67.00	85.15	112.55	162.08	231.04	51.03	54.99	55.91	56.68	57.56	57.92
Experiment B: SAR errors with $\delta=0.8$																		
20	14.55	21.03	25.99	33.79	47.93	67.87	13.11	19.07	23.56	30.55	43.28	61.37	33.00	34.12	34.43	34.67	34.77	34.82
30	14.58	20.97	26.10	33.49	47.77	67.40	16.29	23.65	29.20	37.85	53.67	76.01	22.10	22.49	22.85	22.71	22.90	22.85
50	14.45	21.02	25.45	33.40	47.17	66.73	21.30	30.89	38.03	49.34	69.96	99.08	13.06	13.43	13.28	13.50	13.48	13.48
100	14.36	20.79	25.65	33.31	47.11	66.90	30.47	44.03	54.22	70.31	99.72	141.20	6.45	6.61	6.66	6.69	6.70	6.72
200	14.49	20.60	25.37	32.99	46.69	66.46	43.42	62.56	77.01	99.82	141.55	200.41	3.25	3.27	3.28	3.31	3.31	3.33
Experiment B: SAR errors with $\delta=0.4$																		
20	3.12	4.37	5.42	7.02	9.89	14.03	4.73	6.88	8.57	11.11	15.67	22.34	7.15	7.09	7.18	7.20	7.17	7.20
30	3.18	4.39	5.47	6.92	9.87	13.88	5.88	8.53	10.61	13.74	19.54	27.69	4.82	4.71	4.79	4.69	4.73	4.71
50	3.11	4.41	5.23	6.91	9.69	13.88	7.64	11.18	13.77	17.91	25.44	27.68	2.81	2.82	2.73	2.79	2.77	4.71
100	3.04	4.33	5.31	6.88	9.70	13.78	10.88	15.89	19.60	25.50	36.27	51.38	1.37	1.38	1.38	1.38	1.38	1.39
200	3.12	4.29	5.22	6.79	9.59	13.67	15.52	22.57	27.87	36.25	51.50	72.97	0.70	0.68	0.68	0.68	0.68	0.69

N	CD					CD(p)					\bar{r} ($\times 100$)							
	T					T					T							
	10	20	30	50	100	200	10	20	30	50	100	200	10	20	30	50	100	200
Experiment C: common factors + SAR errors with $\delta=0.8$																		
20	22.09	32.30	40.45	52.80	75.09	106.83	14.14	20.45	25.34	32.88	46.64	66.10	50.67	52.39	53.57	54.17	54.47	54.80
30	28.39	41.85	52.65	68.87	98.65	140.71	17.48	25.37	31.36	40.69	57.83	82.01	43.04	44.87	46.09	46.70	47.30	47.71
50	40.80	61.50	77.49	101.98	146.07	207.80	22.81	33.20	40.95	53.13	75.43	106.86	36.87	39.29	40.42	41.20	41.74	41.98
100	71.86	111.36	138.74	184.60	265.09	378.08	32.70	47.36	58.35	75.80	107.58	152.44	32.30	35.39	36.00	37.11	37.68	38.00
200	134.73	211.06	264.90	348.84	503.07	718.34	46.47	67.34	82.99	107.59	152.73	216.39	30.20	33.46	34.28	34.97	35.66	36.01
Experiment C: common factors + SAR errors with $\delta=0.4$																		
20	22.32	33.41	41.86	54.99	78.53	111.72	11.30	16.73	20.89	27.30	38.90	55.24	51.20	54.19	55.45	56.42	56.97	57.31
30	33.06	49.42	62.21	81.22	116.81	166.83	14.01	20.70	25.86	33.68	48.22	68.70	50.13	52.98	54.46	55.08	56.01	56.56
50	54.40	81.88	102.90	135.19	193.90	275.81	18.24	27.07	33.72	44.10	62.97	89.44	49.15	52.31	53.68	54.63	55.40	55.72
100	107.82	164.03	204.42	269.89	386.73	550.72	26.02	38.73	48.04	62.98	89.88	127.70	48.16	52.13	53.05	54.25	54.97	55.35
200	214.86	328.99	410.12	537.32	772.10	1099.50	36.93	55.18	68.46	89.25	127.63	181.34	48.16	52.15	53.08	53.87	54.73	55.11

Notes: the CD, CD(p) and \bar{r} have been computed on residuals $\hat{u}_{it} = y_{it} - \hat{\alpha}_i d_{1t} - \hat{\delta}' x_{it}$, using the CCE Pooled estimator (42)