Weak and Strong Cross Section Dependence and Estimation of Large Panels*

Alexander Chudik†
European Central Bank and CIMF

M. Hashem Pesaran‡
Cambridge University, CIMF and USC

Elisa Tosetti§
Cambridge University

October 5, 2009

Abstract

This paper introduces the concepts of time-specific weak and strong cross section dependence. A double-indexed process is said to be cross sectionally weakly dependent at a given point in time, t, if its weighted average along the cross section dimension (N) converges to its expectation in quadratic mean, as N is increased without bounds for all weights that satisfy certain 'granularity' conditions. Relationship with the notions of weak and strong common factors is investigated and an application to the estimation of panel data models with an infinite number of weak factors and a finite number of strong factors is also considered. The paper concludes with a set of Monte Carlo experiments where the small sample properties of estimators based on principal components and CCE estimators are investigated and compared under various assumptions on the nature of the unobserved common effects.

Keywords: Panels, Strong and Weak Cross Section Dependence, Weak and Strong Factors.
JEL Classification: C10, C31, C33.

*A preliminary version of this paper was presented at The Econometrics Journal Special Session, Royal Economic Society Annual Conference, Surrey, April 2009, and at the Institute for Advanced Studies, Vienna, May 2009. We have benefited from comments by participants, and particularly from our discussions with Manfred Deistler.

†European Central Bank, Kaiserstrasse 29, 60311 Frankfurt am Main, Germany; e-mail: alexander.chudik@ecb.int.

‡Faculty of Economics, Austin Robinson Building, Sidgwick Avenue, Cambridge, CB3 9DD, UK. E-mail: mhp1@econ.cam.ac.uk; http://www.econ.cam.ac.uk/faculty/pesaran/.

§Faculty of Economics, Austin Robinson Building, Sidgwick Avenue, Cambridge, CB3 9DD, UK. E-mail: et268@cam.ac.uk. Elisa Tosetti acknowledges financial support from ESRC (Ref.no. RES-061-25-0317).

1 Introduction

There exists a growing literature on econometric methods for representing and measuring cross section dependence in panel data regression models. Conditioning on variables specific to the cross section units alone typically does not deliver cross section error independence and it is well known that neglecting cross section dependence can lead to biased estimates and spurious inference.

How to account for contemporaneous error correlations depends on the number of cross section units, N, relative to the time series dimension, T, and in most cases on the nature and the degree of cross section dependencies observed. When N is small relative to T, the nature of cross section dependence is unimportant as long as the errors are not correlated with the regressors, in which case the Seemingly Unrelated Regression Equations (SURE) approach can be used (Zellner (1962)). But when N is large relative to T, the SURE procedure is not applicable and the nature of cross section dependence needs to be taken into account. In such cases there are two main approaches to modelling cross section dependence in panels : (i) spatial processes pioneered by Whittle (1954) and developed further by Anselin (1988), Kelejian and Prucha (1999), and Lee (2004); and (ii) factor models introduced by Hotelling (1933) and first applied in economics by Stone (1947). Factor models have been used extensively in finance (Chamberlain and Rothschild (1983), Connor and Korajzcyk (1993); Stock and Watson (1998); Kapetanios and Pesaran (2007)), and in macroeconomics (Forni and Reichlin (1998); Stock and Watson (2002)). While in principle, as we shall see, cross sectionally dependent processes, including spatial and network processes, can be set up as an unobserved factor structure with possibly infinite number of factors, the original idea for using latent factors is to characterize co-movements of individual cross section units by a small number of latent factors plus a white noise, in order to overcome the curse of dimensionality.

The aim of this paper is to characterize the correlation pattern over the cross sectional dimension for a general class of processes, regardless of whether they are represented by factor or spatial models or any other process featuring cross section dimension proposed in the literature. Unlike in the case of time series, data along the cross sectional dimension do not typically have a natural ordering. One way to characterize the correlation structure of double index processes has been proposed in the factor literature. The idiosyncratic (or weak dependence) property, advanced by Forni and Lippi (2001), applies to both dimensions and requires that the weighted average of a stationary process, computed both over time and across sections, converges to zero in quadratic mean for all sets of weights satisfying a certain condition. This notion is used by the authors to characterize dynamic factor models. Their framework is a generalization of the static model for asset markets by Chamberlain (1983) and Chamberlain and Rothschild (1983), and extends some of the results presented by Forni and Reichlin (1998). Forni and Lippi (2001) show that a necessary and sufficient condition for a process to be idiosyncratic (or weakly dependent over time and across the units) is the boundedness of the largest eigenvalue of its spectral density matrix at all frequencies. Using this result, Anderson et al. (2009) (see their Definition 4) formally define a double index stochastic process as weakly dependent if the largest eigenvalues of its spectral density is bounded in N (at all frequencies), as opposed to a strongly dependent process, for which a finite, nonzero number of eigenvalues diverge to infinity as N goes to infinity. We remark that these assumptions on the asymptotic behaviour of eigenvalues of the spectral density are needed for identification of common factors and their loadings, and their estimation by principal components analysis. Further, to ensure the existence of the spectral density, this literature assumes that the underlying time series processes are stationary with absolutely summable autocovariances.

This paper proposes a new characterization of cross section dependence into weak and strong, which are more widely applicable than the definitions introduced by Anderson et al. (2009). We consider the asymptotic behaviour of weighted averages at each point in time, which does not require any stationarity assumptions to be imposed on the underlying time series processes. 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 weights that satisfy certain 'granularity' conditions. If this requirement does not hold, then the process is said to be cross sectionally strongly dependent. Convergence properties of weighted averages is of great importance for the asymptotic theory of various estimators and tests commonly used in panel data econometrics, as well as for arbitrage pricing theory and portfolio optimization with a large number of assets. It is clear that the underlying time series processes in either of the two literature need not be stationary, and concepts of weak and strong dependence that are more generally applicable are needed.

In this paper we focus on the econometric literature and consider the problem of estimating the slope coefficients of large panels, where cross section units are subject to a number of unobserved common factors that may rise with N. It is established that Common Correlated Effects (CCE) estimator introduced by Pesaran (2006) remains asymptotically normal under certain conditions on the loadings of the infinite factor structure, including cases where methods relying on principal components fail. A Monte Carlo study documents these theoretical findings by investigating the small sample performance of estimators based on principal components and the CCE estimators under alternative assumptions on the nature of unobserved common effects. In particular, we examine and compare the performance of these estimator when the errors are subject to a finite number of unobserved strong factors and an infinite number of weak and/or semi-weak unobserved common factors.

The plan of the remainder of the paper is as follows. Section 2 introduces the concepts of strong and weak cross section dependence, and explores the relationship between the dependence structure of processes. Section 3 focuses on cross section dependence in dynamic panels. Section 4 presents common factor models and discusses the notions of weak, semi-strong and strong factors. Section 5 introduces the CCE estimators in the context of panels with an infinite number of common factors. Section 6 describes the Monte Carlo design and discusses the results. Finally, Section 7 provides some concluding remarks.

Notation: $|\lambda_1(\mathbf{A})| \geq |\lambda_2(\mathbf{A})| \geq ... \geq |\lambda_n(\mathbf{A})|$ are the eigenvalues of a matrix $\mathbf{A} \in \mathbb{M}^{n\times n}$, where

$\mathbb{M}^{n\times n}$ is the space of $n\times n$ complex valued matrices. $\mathbf{A}^{-}$ denotes a generalized inverse of $\mathbf{A}$. The spectral radius of $\mathbf{A}\in\mathbb{M}^{n\times n}$ is $\rho(\mathbf{A})=\max_{1\leq j\leq n}[|\lambda_i(\mathbf{A})|]$, and its column norm is $\|\mathbf{A}\|_1=\max_{1\leq j\leq n}\sum_{i=1}^{n}|a_{ij}|$. The row norm of $\mathbf{A}$ is $\|\mathbf{A}\|_\infty=\max_{1\leq i\leq n}\sum_{j=1}^{n}|a_{ij}|$. The spectral norm of $\mathbf{A}$ is $\|\mathbf{A}\|=[\rho(\mathbf{A}\mathbf{A}')]^{1/2}$, and $\|\mathbf{A}\|_2=[Tr\left(\mathbf{A}\mathbf{A}'\right)]^{1/2}$. K is used for a fixed positive constant that does not depend on N.

2 Cross section dependence in large panels

In this section, we study the structure of correlation of the double index process $\{z_{it}, i\in\mathbb{N}, t\in\mathbb{Z}\}$ where z_{it} are random variables defined on a probability space $(\Omega,\mathcal{F},P)$; 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 structure of the double index process $\{z_{it}\}$ over the cross sectional dimension. We start by reviewing definitions provided in the existing literature to characterize the correlation pattern of $\{z_{it}\}$; and next we introduce our general notions of weakly and strongly cross sectionally dependent processes.

2.1 Weak and strong dependence

Forni and Lippi (2001) introduce the notion of idiosyncratic process to characterize a weak form of dependence that involves both time series and cross sectional dimensions under the following assumption:

Assumption 1 (Forni and Lippi, 2001, Assumption 1) *For each $N\in\mathbb{N}$, the process $\mathbf{z}_{Nt}=(z_{1t},...,z_{Nt})'$ is covariance stationary and the spectral measure of $\mathbf{z}_{Nt}$ is absolutely continuous.*

Notice that Assumption 1 guarantees the spectral density for the vector $\mathbf{z}_{Nt}$ to exist. Consider any sequence of weights vectors $\mathbf{w}_N=(w_1,w_2,...,w_N)'$ such that

$$\lim_{N\to\infty}\|\mathbf{w}_N\|=0. \tag{1}$$

Let $\mathbf{F}_{zN}(\omega)$ denote the spectral density matrix for $\mathbf{z}_{Nt}$ and define the norm $\|\mathbf{w}_N\|_{\mathbf{F}_{zN}}$ as

$$\|\mathbf{w}_N\|_{\mathbf{F}_{zN}}=\frac{1}{2\pi}\int_{-\pi}^{\pi}\mathbf{w}_N'\mathbf{F}_{zN}(\theta)\,\mathbf{w}_N d\theta.$$

Forni and Lippi (2001) define the process $\{z_{it}\}$ as *idiosyncratic* if, for all weights $\mathbf{w}_N$ satisfying condition (1), we have

$$\lim_{N\to\infty}\|\mathbf{w}_N\|_{\mathbf{F}_{zN}}=0.$$

The idiosyncratic property implies that the variance of the weighted average of $\{z_{it}\}$, computed both over time and across sections, vanishes to zero as N tends to infinity. The authors show that the sequence $\{z_{it}\}$ is idiosyncratic if and only if the largest eigenvalue of $\mathbf{F}_{zN}(\omega)$, $\lambda^{z}_{N,1}(\omega)$, is bounded in ω and N. Further, a process $\{z_{it}\}$ for which the $(m+1)$th eigenvalue of $\mathbf{F}_{zN}(\omega)$ is bounded in ω and N, and the mth eigenvalue diverges in N for all frequencies ω, can be represented by the so-called generalized factor structure, namely a linear combination of m dynamic factors, plus an idiosyncratic process (see their Theorems 1 and 2). This is an extension to the dynamic case of the static factor model used in arbitrage pricing theory as advanced by Ross (1976) and further developed by Chamberlain (1983), Chamberlain and Rothschild (1983), and Ingersoll (1984).

Based on the above results, Anderson et al. (2009) define the concepts of weak and strong dependence for processes $\{z_{it}\}$ satisfying Assumption 1, on the basis of the asymptotic behaviour of the eigenvalues of $\mathbf{F}_{zN}(\omega)$.

Definition 1 *(Weak and strong dependence) The double index processes $\{z_{it}, i \in \mathbb{N}, t \in \mathbb{Z}\}$ is weakly dependent if $\lambda^{z}_{N,1}(\omega)$ is uniformly bounded in ω and N. The process $\{z_{it}\}$ is strongly dependent if the first $m \geq 1$ $(m < K)$ eigenvalues $(\lambda^{z}_{N,1}(\omega), ..., \lambda^{z}_{N,m}(\omega))$ diverge to infinity as $N \to \infty$, for all frequencies.*

For further details on the above definitions we refer to Forni and Lippi (2001) (see their Assumption 1, Definitions 1, 6 and 9; Theorems 1 and 2), and Anderson et al. (2009) (see their Assumptions 4 and 5).

We note that the stationarity of the time series processes in $\mathbf{z}_{Nt}$ set in Assumption 1 is needed for estimation by (dynamic) principal components analysis of common factors and their loadings in the generalized factor structure. However, this assumption is likely to be quite restrictive and is unlikely to hold in many applications, especially in finance where time series often exhibit time-varying volatility.

2.2 Weak and strong cross section dependence

We now present our definitions of weak and strong cross section dependence at a given point in time. For ease of exposition, in the following we omit the subscript N where not necessary. We make the following assumptions:

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

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

and

$$\frac{w_{jt}}{\|\mathbf{w}_{Nt}\|} = O\left(N^{-\frac{1}{2}}\right) \quad \text{for any } j \in \mathbb{N}. \tag{3}$$

Assumption 3 *Let $\mathcal{I}_t$ be the information set available at time t. For each $t \in \mathcal{T}$, $\mathbf{z}_{Nt} = (z_{1t}, ..., z_{Nt})'$ has conditional mean and variance*

$$E\left(\mathbf{z}_{Nt} \left| \mathcal{I}_{t-1} \right.\right) = \mathbf{0}, \tag{4}$$

$$Var\left(\mathbf{z}_{Nt} \left| \mathcal{I}_{t-1} \right.\right) = \mathbf{\Sigma}_{Nt}, \tag{5}$$

where $\mathbf{\Sigma}_{Nt}$ 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 $i = 1, ..., N$, where K is a finite constant independent of N.

Assumption 2, known in finance as the granularity condition, ensures that the weights $\{w_{it}\}$ are not dominated by a few of the cross section units. Although we have assumed the weights to be non-stochastic, this is done for expositional convenience and can be relaxed by requiring that conditional on the information set the weights, $\mathbf{w}_{Nt}$, are distributed independently of $\mathbf{z}_{Nt}$. In Assumption 3 we impose some regularity conditions on the time series properties of $\{z_{it}\}$. Assumption 3 is also standard in finance and specifies that $\mathbf{z}_{Nt}$ has conditional means and variances. The first part, (4), can be relaxed to $E\left(\mathbf{z}_{Nt} \left| \mathcal{I}_{t-1} \right.\right) = \boldsymbol{\mu}_{N,t-1}$, with $\boldsymbol{\mu}_{N,t-1}$ 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 $\boldsymbol{\mu}_{N,t-1} = \mathbf{0}$.

To simplify the notations we suppress the explicit dependence of $\mathbf{z}_{Nt}$, $\mathbf{w}_{Nt}$ and other vectors and matrices on N, unless this is needed to avoid possible confusions.

Consider now the weighted averages, $\bar{z}_{wt} = \sum_{i=1}^{N} w_{it} z_{it} = \mathbf{w}_t' \mathbf{z}_t$, for $t \in \mathcal{T}$, where $\mathbf{z}_t$ and $\mathbf{w}_t$ satisfy Assumptions 2 and 3. We are interested in the limiting behavior of $\bar{z}_{wt}$ at a given point in time $t \in \mathcal{T}$ as $N \rightarrow \infty$.

Definition 2 *(Weak and strong cross section dependence) The process $\{z_{it}\}$ is said to be cross sectionally weakly dependent (CWD) at a given point in time $t \in \mathcal{T}$ conditional on information set $\mathcal{I}_{t-1}$, if for any sequence of weight vectors $\{\mathbf{w}_t\}$ satisfying the granularity conditions (2)-(3) we have*

$$\lim_{N \rightarrow \infty} Var(\mathbf{w}_t' \mathbf{z}_t \left| \mathcal{I}_{t-1} \right.) = 0. \tag{6}$$

$\{z_{it}\}$ is said to be cross sectionally strongly dependent (CSD) at a given point in time $t \in \mathcal{T}$ conditional on information set $\mathcal{I}_{t-1}$, if there exists a sequence of weights vectors $\{\mathbf{w}_t\}$ satisfying (2)-(3) and a constant K independent of N such that for any N sufficiently large

$$Var(\mathbf{w}_t' \mathbf{z}_t \left| \mathcal{I}_{t-1} \right.) \geq K > 0. \tag{7}$$

The concepts of weak and strong cross section dependence proposed here are defined conditional on an information set, namely the set $\mathcal{I}_{t-1}$ in the definition above. In this way we are able to consider cross section dependence properties of $\{z_{it}\}$ without having to limit the time series features of the process. Various information sets could be considered in practise, depending on applications. One example is the set containing lagged realizations of the process $\{z_{it}\}$, that is $\mathcal{I}_{t-1} = \{\mathbf{z}_{t-1}, \mathbf{z}_{t-2}, ...\}$.

In the context of dynamic models, it is useful to condition on the initialization of the dynamic process (i.e. starting values) only. In stationary panels, unconditional variances of cross section averages could be considered. In the remainder of the paper, if not stated explicitly, the concepts of CWD and CSD are always defined on the information set $\mathcal{I}_{t-1}$.

Remark 1 *In contrast to the notions of weak and strong dependence advanced by Forni and Lippi (2001) and Anderson et al. (2009), our concepts of CWD and CSD do no require the underlying processes to be covariance stationary and have spectral density at all frequencies.*

Remark 2 *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. However, we observe that the notion of weak dependence does not necessarily involve an ordering of the observations or the specification of a distance metric.*

2.3 Properties of weakly and strongly cross sectionally dependent processes

The following proposition establishes the relationship between weak cross section dependence and the asymptotic behaviour of the spectral radius of $\mathbf{\Sigma}_t$ (namely, $\lambda_1(\mathbf{\Sigma}_t)$).

Proposition 1 *The following statements hold:*

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

(ii) The process $\{z_{it}\}$ is CSD at a point in time $t \in \mathcal{T}$ if and only if $\lim_{N\to\infty} \frac{1}{N}\lambda_1(\mathbf{\Sigma}_t) = K > 0$.

Proof. First, suppose $\lambda_1(\mathbf{\Sigma}_t)$ is bounded in N. We have

$$Var(\mathbf{w}_t'\mathbf{z}_t \,|\mathcal{I}_{t-1}) = \mathbf{w}_t'\mathbf{\Sigma}_t\mathbf{w}_t \leq \left(\mathbf{w}_t'\mathbf{w}_t\right)\lambda_1(\mathbf{\Sigma}_t), \tag{8}$$

and under the granularity conditions (2)-(3) it follows that

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

namely that $\{z_{it}\}$ is CWD, which proves (i). Now suppose that $\{z_{it}\}$ is CSD at time t. Then, from (8), it follows that $\lambda_1(\mathbf{\Sigma}_t)$ tends to infinity at least at the rate N. Noting that $\lambda_1(\mathbf{\Sigma}_t) \leq \sum_{i=1}^{N}\sigma_{ii,t}$ where, under Assumption 3, $\sigma_{ii,t}$ are finite, $\lambda_1(\mathbf{\Sigma}_t)$ cannot diverge to infinity at a rate larger than N, and hence it follows that under CSD $\lim_{N\to\infty} \frac{1}{N}\lambda_1(\mathbf{\Sigma}_t) = K > 0$. To prove the reverse relation, first note that, from the Rayleigh-Ritz theorem[1],

$$\lambda_1(\mathbf{\Sigma}_t) = \max_{\mathbf{v}_t'\mathbf{v}_t=1} \mathbf{v}_t'\mathbf{\Sigma}_t\mathbf{v}_t = \mathbf{v}_t^{*\prime}\mathbf{\Sigma}_t\mathbf{v}_t^{*}. \tag{9}$$

[1] See Horn and Johnson (1985), p.176.

Let $\mathbf{w}_t^* = \frac{1}{\sqrt{N}}\mathbf{v}_t^*$ and notice that $\mathbf{w}_t^*$ satisfies (2)-(3). Hence, we can rewrite $\lambda_1(\mathbf{\Sigma}_t)$ as

$$\lambda_1(\mathbf{\Sigma}_t) = N \cdot Var(\mathbf{w}_t^{*\prime}\mathbf{z}_t \left| \mathcal{I}_{t-1} \right.). \tag{10}$$

It follows that if $\lim_{N\to\infty} \frac{1}{N}\lambda_1(\mathbf{\Sigma}_t) = K > 0$, then $\lim_{N\to\infty} Var(\mathbf{w}_t^{*\prime}\mathbf{z}_t \left| \mathcal{I}_{t-1} \right.) > 0$, i.e. the process is CSD, which proves (ii). ■

Since [2]

$$\lambda_1(\mathbf{\Sigma}_t) \le \|\mathbf{\Sigma}_t\|_1 ,$$

it follows from (8) that if $\lim_{N\to\infty} \frac{1}{N}\lambda_1(\mathbf{\Sigma}_t) > 0$ then also $\lim_{N\to\infty} \frac{1}{N}\|\mathbf{\Sigma}_t\|_1 > 0$. Hence, 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 temporal dependence (Robinson (2003)).

A number of remarks concerning the above concepts of CWD and CSD are in order.

Remark 3 *The definition of idiosyncratic process by Forni and Lippi (2001) and our definition of CWD differ in the way weights used to build weighted averages are defined. While Forni and Lippi assume* $\lim_{N\to\infty}\|\mathbf{w}\| = 0$*, our granularity conditions (2)-(3) imply that, for any* $t \in \mathcal{T}$*,* $\lim_{N\to\infty} N^{\frac{1}{2}-\epsilon}\|\mathbf{w}_t\| = 0$ *for any* $\epsilon > 0$*. This difference in the definition of weights has some implications on the properties of our processes. In particular, under (1), it is possible to show that the idiosyncratic process (and hence also the definition of weak dependence à la Anderson et al. (2009)) imply bounded eigenvalues of the spectral density matrix. Conversely, under (2)-(3), it is clear that if* $\lambda_1(\mathbf{\Sigma}_t) = O(N^{1-\epsilon})$ *for any* $\epsilon > 0$*, then, using (8),*

$$\lim_{N\to\infty} \left(\mathbf{w}_t'\mathbf{w}_t\right)\lambda_1(\mathbf{\Sigma}_t) = 0,$$

and the underlying process will be CWD. Hence, the bounded eigenvalue condition 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. In Section 3, we investigate the relation between bounded eigenvalues of the spectral density matrix, and bounded eigenvalues of the covariance matrix,* $\mathbf{\Sigma}_t$*, in the case of dynamic panels.*

One rationale for characterizing processes with increasing largest eigenvalues at the slower pace than N as weakly dependent is that bounded eigenvalues is not a necessary condition for consistent estimation in general, although in some cases, such as the method of principal components, this condition is necessary. More on this below in Section 5, where we consider estimation of slope coefficients in panels with an infinite factor structure.

We conclude this section with two results concerning the relationship between strongly and weakly cross sectionally correlated variables. Following Definition 2, we say that two processes $\{z_{it,a}\}$ and

[2] See Horn and Johnson (1985), pp. 297-298.

$\{z_{it,b}\}$ are weakly correlated at time t if $\lim_{N\rightarrow\infty} E(\bar{z}_{wt,a}\bar{z}_{wt,b}\,|\mathcal{I}_{t-1}) = 0$, for all sets of weights 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 2 *Suppose that $\{z_{it,a}\}$ and $\{z_{it,b}\}$ are CSD and CWD processes, respectively. Then for all sets of weights $\{w_{it}^{a}\}$ and $\{w_{it}^{b}\}$ satisfying conditions (2)-(3), we have*

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

Proof. Let $\left\{w_{i,t-1}^{a}\right\}$ and $\left\{w_{i,t-1}^{b}\right\}$ be two sets of weights satisfying conditions (2)-(3). For $t \in \mathcal{T}$, we have

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

Further, under Assumption 3 the process $z_{it,a}$ satisfies

$$E(\bar{z}_{wt,a}^2\,|\mathcal{I}_{t-1}) < K,$$

where K is a finite constant. Also from (6), 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 (2)-(3), we obtain

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

■

Proposition 3 *Consider two independent processes $\{z_{it,a}\}$ and $\{z_{it,b}\}$, and their linear combinations defined by*

$$z_{it,c} = \beta_a z_{it,a} + \beta_b z_{it,b}, \tag{11}$$

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

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

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

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

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

$$Var(\mathbf{w}_t'\mathbf{z}_{t,c}\,|\mathcal{I}_{t-1}) = Var(\mathbf{w}_t'\mathbf{z}_{t,a}\,|\mathcal{I}_{t-1}) + Var(\mathbf{w}_t'\mathbf{z}_{t,b}\,|\mathcal{I}_{t-1}).$$

Noting that $Var(\mathbf{w}_t'\mathbf{z}_{t,a}\,|\mathcal{I}_{t-1})$ and $Var(\mathbf{w}_t'\mathbf{z}_{t,b}\,|\mathcal{I}_{t-1})$ satisfy (6), then $\lim_{N\to\infty} Var(\mathbf{w}_t'\mathbf{z}_{t,c}\,|\mathcal{I}_{t-1}) = 0$, 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 4.

3 Dynamic panels

Suppose that for each $N \in \mathbb{N}$, cross section units collected into the vector $\mathbf{z}_t = (z_{1t}, z_{2t}, ..., z_{Nt})'$ are generated from the following VAR model,

$$\mathbf{z}_t = \mathbf{\Phi}_t \mathbf{z}_{t-1} + \mathbf{u}_t, \tag{12}$$

where $\mathbf{\Phi}_t$ is a $N \times N$ dimensional matrix of unknown coefficients, which could be time-varying, the vector $\mathbf{u}_t$ of reduced-form errors has mean and variance

$$E\left(\mathbf{u}_t\right) = \mathbf{0},\; E\left(\mathbf{u}_t\mathbf{u}_t'\right) = \mathbf{\Sigma}_t, \tag{13}$$

where $\mathbf{\Sigma}_t$, $t = 1, ..., T$, are $N \times N$ symmetric, nonnegative definite matrix, and $\mathbf{u}_t$ is independently distributed of $\mathbf{u}_{t'}$ for any $t \neq t'$. The initialization of the dynamic process could be from a finite past, $t \in \mathcal{T} \equiv \{-M+1, .., 0, ..\} \subseteq \mathbb{Z}$, M being a fixed positive integer; or we can let $M \to \infty$, as in Chudik and Pesaran (2009). The infinite-dimensional spatio-temporal model (12) can also be viewed more generally as a 'dynamic network', with $\mathbf{\Sigma}_t$ and $\mathbf{\Phi}_t$ capturing the static and dynamic forms of inter-connections that might exist in the network. All linear dynamic panel data models existing in the literature could be written as special cases of (12). Sequence of models (12) of growing dimension ($N \to \infty$) is non-nested since the dependence between unit i and j could change with the inclusion of new unit(s). For this reason, the process $\{z_{it}, N \in \mathbb{N}, i \in \{1, .., N\}, t \in \mathcal{T}\}$ given by (12) is a triple index process, but we continue to omit subscript N (were not necessary) to simplify the exposition.

Object of this section is to investigate the correlation pattern of $\{z_{it}\}$ across the cross sectional units in the dynamic setting given by (12). In our analysis, we set $\mathcal{I}_t$ to contain only the starting values, $\mathbf{z}_{-M}$, i.e. $\mathcal{I}_t = \mathcal{I} = \{\mathbf{z}_{-M}\}$. Consider the following assumptions on the coefficient matrices, $\mathbf{\Phi}_t$, and the error vector, $\mathbf{u}_t$:

Assumption 4 *There exist a constant $K < \infty$ and an arbitrarily small positive constant $\epsilon > 0$ such that for any fixed $t \in \mathcal{T}$ and any $N \in \mathbb{N}$, we have*

$$\|\mathbf{\Phi}_t\| < K, \tag{14}$$

and

$$\|\mathbf{\Sigma}_t\| < K \cdot N^{1-\epsilon}. \tag{15}$$

Remark 4 *Equation (15) of Assumption 4 implies that $\{u_{it}\}$ is CWD. The initialization of a dynamic process could be from a non-stochastic point or could have been from a stochastic point, possibly generated from a process different from the DGP of $\{u_{it}\}$.*

Proposition 4 *Consider model (12) and suppose Assumption 4 holds. Then for any sequence of weight vectors $\{\mathbf{w}_t\}$ satisfying condition (2), and for a fixed M and a fixed $t \in \mathcal{T}$,*

$$\lim_{N\to\infty} Var\left(\mathbf{w}_t'\mathbf{z}_t \mid \mathbf{z}_{-M}\right) = 0. \tag{16}$$

Proof. The vector difference equation (12) can be solved backwards, taking $\mathbf{z}_{-M}$ as given:

$$\mathbf{z}_t = \left(\prod_{s=0}^{t+M-1} \mathbf{\Phi}_{t-s}\right)\mathbf{z}_{-M} + \sum_{\ell=0}^{t+M-1}\left(\prod_{s=0}^{\ell-1}\mathbf{\Phi}_{t-s}\right)\mathbf{u}_{t-\ell}.$$

The variance of $\mathbf{z}_t$ (conditional on initial values) is

$$\mathbf{\Omega}_{t,-M} = Var\left(\mathbf{z}_t \mid \mathbf{z}_{-M}\right) = \sum_{\ell=0}^{t+M-1}\left(\prod_{s=0}^{\ell-1}\mathbf{\Phi}_{t-s}\right)\mathbf{\Sigma}_{t-\ell}\left(\prod_{s=0}^{\ell-1}\mathbf{\Phi}_{t-s}\right)'.$$

For any $t \in \mathcal{T}$, $\|\mathbf{\Omega}_{t,-M}\|$ is under Assumption 4 bounded by

$$\|\mathbf{\Omega}_{t,-M}\| \leq \sum_{\ell=0}^{t+M-1}\left(\prod_{s=0}^{\ell-1}\|\mathbf{\Phi}_{t-s}\|^2\right)\|\mathbf{\Sigma}_{t-\ell}\| = O\left(N^{1-\epsilon}\right).$$

It follows that for any arbitrary vector of weights satisfying (2),

$$Var\left(\mathbf{w}_t'\mathbf{z}_t \mid \mathbf{z}_{-M}\right) = \mathbf{w}_t'\mathbf{\Omega}_{t,-M}\mathbf{w}_t \leq \rho\left(\mathbf{\Omega}_{t,-M}\right)\left(\mathbf{w}_t'\mathbf{w}_t\right) = o(1), \tag{17}$$

where $\rho(\mathbf{\Omega}_{t,-M}) \leq \|\mathbf{\Omega}_{t,-M}\| = O\left(N^{1-\epsilon}\right)$, and $\mathbf{w}_t'\mathbf{w}_t = \|\mathbf{w}_t\|^2 = O\left(N^{-1}\right)$. ■

Hence, the dynamic process $\{z_{it}\}$ given by (12) under Assumption 4 is CWD at any point in time $t \in \mathcal{T}$, conditional on starting values $\mathbf{z}_{-M}$. The result of the above proposition can be readily extended to situations where M and/or $t \to \infty$. In such cases we need the stronger requirement that $\|\mathbf{\Phi}_t\| < 1-\epsilon$, for all $t \in \mathcal{T}$. It is then easily seen that the VAR(1) model, (12), yields a cross sectionally weakly dependent process if for all t and N, $\|\mathbf{\Sigma}_t\| < K \cdot N^{1-\epsilon}$, and $\|\mathbf{\Phi}_t\| < 1-\epsilon$, irrespective of the values of t and M.[3] There are several interesting implications of this finding. Consider the following additional assumption on the coefficients matrix $\mathbf{\Phi}_t$, which states that for some units the off-diagonal elements of the matrix $\mathbf{\Phi}_t$ are small.

Assumption 5 *Let $\mathcal{K} \subseteq \mathbb{N}$ be a non-empty index set. Define vector $\boldsymbol{\phi}_{t,-i} = \left(\phi_{ti1}, ..., \phi_{t,i,i-1}, 0, \phi_{t,i,i+1}, ..., \phi_{t,iN}\right)'$ where ϕ_{tij} for $i,j \in \{1,2,...,N\}$ is the (i,j) element of matrix $\mathbf{\Phi}_t$. For any $i \in \mathcal{K}$ and any $t \in \mathcal{T}$, vector $\boldsymbol{\phi}_{t,-i}$ satisfies*

$$\left\|\boldsymbol{\phi}_{t,-i}\right\| = \left(\sum_{j=1,j\neq i}^{N} \phi_{tij}^2\right)^{1/2} = O\left(N^{-\frac{1}{2}}\right). \tag{18}$$

Remark 5 *Assumption 5 implies that for $i \in \mathcal{K}$, $\sum_{i=1,i\neq j}^{N} \phi_{tij} \leq \left\|\boldsymbol{\phi}_{t,-i}\right\|_1 = O(1)$.[4] Therefore, it is possible for the dependence of each individual unit on the rest of the units in the system to be large. However, as we shall see below, in the case where $\{z_{it}\}$ is a CWD process, the model for the i^{th} cross section unit de-couples from the rest of the system as $N \to \infty$.*

Corollary 1 *Consider model (12) and suppose Assumptions 4 and 5 hold. Then, a fixed M, a fixed $t \in \mathcal{T}$, and any $i \in \mathcal{K}$,*

$$\lim_{N\to\infty} Var\left(z_{it} - \phi_{tii} z_{i,t-1} - u_{it} \mid \mathbf{z}_{-M}\right) = 0. \tag{19}$$

If, in addition to Assumptions 4 and 5, $\|\mathbf{\Phi}_t\| < 1-\epsilon$ and $M \to \infty$, we have

$$\lim_{N\to\infty} Var\left(z_{it} - \phi_{tii} z_{i,t-1} - u_{it}\right) = 0 \text{ for any } i \in \mathcal{K} \text{ and any } t \in \mathcal{T}. \tag{20}$$

[3] Under these assumptions the unconditional variance of $\mathbf{z}_t$ is bounded by

$$\begin{aligned} \|Var(\mathbf{z}_t)\| &= \|\mathbf{\Omega}_t\| \leq \sum_{\ell=0}^{\infty}\left(\prod_{s=0}^{\ell-1}\|\mathbf{\Phi}_{t-s}\|^2\right)\|\mathbf{\Sigma}_{t-\ell}\| \\ &< \sup_{t\in\mathcal{T}}\|\mathbf{\Sigma}_t\| \cdot \sum_{\ell=0}^{\infty}(1-\epsilon)^{2\ell} = O\left(N^{1-\epsilon}\right). \end{aligned}$$

[4] Note that $\left\|\boldsymbol{\phi}_{t,-i}\right\|_1 \leq \sqrt{N}\left\|\boldsymbol{\phi}_{t,-i}\right\|$. See Horn and Johnson (1985, p. 314). An example of vector $\boldsymbol{\phi}_{t,-i}$ for which $\lim_{N\to\infty}\sum_{i=1,i\neq j}^{N} \phi_{tij} \neq 0$ is when $\phi_{tij} = k/N$ for $i \neq j$ and any fixed non-zero constant k.

Proof. Assumption 5 implies that for $i \in \mathcal{K}$, vector $\boldsymbol{\phi}_{t,-i}$ satisfies condition (2). It follows from Proposition 4 that

$$\lim_{N\to\infty} Var\left(\boldsymbol{\phi}_{t,-i}'\mathbf{z}_t \mid \mathbf{z}_{-M}\right) = 0 \text{ for any } i \in \mathcal{K} \text{ and any } t \in \mathcal{T}. \tag{21}$$

Similarly, under the assumption $\|\boldsymbol{\Phi}_t\| < 1-\epsilon$ and $M \to \infty$, we have $\|Var(\mathbf{z}_t)\| = O\left(N^{1-\epsilon}\right)$ (see Footnote 3), which implies

$$\lim_{N\to\infty} Var\left(\boldsymbol{\phi}_{t,-i}'\mathbf{z}_t\right) = 0 \text{ for any } i \in \mathcal{K} \text{ and any } t \in \mathcal{T}. \tag{22}$$

System (12) implies

$$z_{it} - \phi_{tii} z_{i,t-1} - u_{it} = \boldsymbol{\phi}_{t,-i}'\mathbf{z}_t, \text{ for any } i \in \{1,..,N\} \text{ and any } t \in \mathcal{T}. \tag{23}$$

Taking conditional variance of (23) and using (21)-(22) now yields (19)-(20). ■

Strong dependence in infinite-dimensional VAR models could arise as a result of CSD errors $\{u_{it}\}$, or could be due to dominant patterns in the coefficients of $\boldsymbol{\Phi}_t$, or both. An example of the former is the residual common factor model where the weighted averages of factor loadings do not converge to zero. Further examples of CSD IVAR models, featuring also dominant unit, are provided in Chudik and Pesaran (2009).

The following proposition presents sufficient conditions for the VAR(1) process to be weakly dependent in the sense of Anderson et al. (2009). Since the concept of weak dependence by Anderson et al. (2009) is defined only for stationary processes, we have to assume that $\boldsymbol{\Phi}_t$ and $\boldsymbol{\Sigma}_t$ are time invariant.

Proposition 5 *Consider model (12) with time invariant coefficient matrix $\boldsymbol{\Phi}_t = \boldsymbol{\Phi}$, and suppose that for each $t \in \mathcal{T}$, $\mathbf{u}_t$ satisfies $E(\mathbf{u}_t) = \mathbf{0}$, $E(\mathbf{u}_t\mathbf{u}_t') = \boldsymbol{\Sigma}$, where $\boldsymbol{\Sigma}$ is a time invariant $N \times N$ symmetric, nonnegative definite matrix, $\mathbf{u}_t$ is independently distributed of $\mathbf{u}_{t'}$ for any $t \neq t'$, and $\rho(\boldsymbol{\Phi}) < 1$, so that $\mathbf{z}_t$ is a covariance stationary process. Then $\mathbf{z}_t$ is weakly dependent, in the sense of Anderson et al. (2009), if $\rho(\boldsymbol{\Sigma}) \leq K < \infty$ and $\|\boldsymbol{\Phi}\| < 1-\epsilon$.*

Proof. The spectral density of $\mathbf{z}_t$ is given by ($i = \sqrt{-1}$)

$$\mathbf{F}_z(\omega) = \frac{1}{2\pi}\left(\mathbf{I}_N - e^{-i\omega}\boldsymbol{\Phi}\right)^{-1}\boldsymbol{\Sigma}\left(\mathbf{I}_N - e^{i\omega}\boldsymbol{\Phi}'\right)^{-1}.$$

For each $N \in \mathbb{N}$, we have

$$\rho\left[\mathbf{F}_z(\omega)\right] = \|\mathbf{F}_z(\omega)\|,$$

and

$$\|\mathbf{F}_z(\omega)\| \leq \frac{1}{2\pi}\left\|\left(\mathbf{I}_N - e^{-i\omega}\boldsymbol{\Phi}\right)^{-1}\right\| \|\boldsymbol{\Sigma}\| \left\|\left(\mathbf{I}_N - e^{i\omega}\boldsymbol{\Phi}'\right)^{-1}\right\|.$$

Under the assumption that $\rho(\mathbf{\Phi}) < 1$,

$$\left(\mathbf{I}_N - e^{-i\omega}\mathbf{\Phi}\right)^{-1} = \mathbf{I}_N + e^{-i\omega}\mathbf{\Phi} + e^{-2i\omega}\mathbf{\Phi}^2 + ...$$

Now we assume $\|\mathbf{\Phi}\| < 1$, and since $\left|e^{-ij\omega}\right| = 1$, it follows

$$\begin{aligned}\left\|\left(\mathbf{I}_N - e^{-i\omega}\mathbf{\Phi}\right)^{-1}\right\| &\leq 1 + \|\mathbf{\Phi}\| + \|\mathbf{\Phi}\|^2 + ... \\ &= \frac{1}{1-\|\mathbf{\Phi}\|}.\end{aligned}$$

Similarly

$$\left\|\left(\mathbf{I}_N - e^{i\omega}\mathbf{\Phi}'\right)^{-1}\right\| \leq \frac{1}{1-\|\mathbf{\Phi}'\|} = \frac{1}{1-\|\mathbf{\Phi}\|},$$

If, in addition, $\rho(\mathbf{\Sigma}) \leq K < \infty$ we have

$$\begin{aligned}\rho\left[\mathbf{F}_z(\omega)\right] &\leq \frac{1}{2\pi}\left\|\left(\mathbf{I}_N - e^{-i\omega}\mathbf{\Phi}\right)^{-1}\right\|\|\mathbf{\Sigma}\|\left\|\left(\mathbf{I}_N - e^{i\omega}\mathbf{\Phi}'\right)^{-1}\right\| \\ &= \frac{1}{2\pi}\rho(\mathbf{\Sigma})\left\|\left(\mathbf{I}_N - e^{-i\omega}\mathbf{\Phi}\right)^{-1}\right\|\left\|\left(\mathbf{I}_N - e^{i\omega}\mathbf{\Phi}'\right)^{-1}\right\| \\ &\leq \frac{1}{2\pi}\rho(\mathbf{\Sigma})\left(\frac{1}{1-\|\mathbf{\Phi}\|}\right)^2 = O(1),\end{aligned}$$

which is bounded in N since both $\rho(\mathbf{\Sigma})$ and $\frac{1}{1-\|\mathbf{\Phi}\|}$ are bounded. This completes the proof. ■

Remark 6 *Notice that under the assumption that* $\|\mathbf{\Phi}\| < 1-\epsilon$ *and if, for at least one frequency* ω_0, *the matrix* $(\mathbf{I}_N - e^{-i\omega_0}\mathbf{\Phi})^{-1}(\mathbf{I}_N - e^{i\omega_0}\mathbf{\Phi}')^{-1}$ *is non-singular, it is possible to show that weak dependence in the sense of Anderson et al. (2009) implies* $\rho(\mathbf{\Sigma}) \leq K < \infty$. *To prove this, first notice that if* $\mathbf{A}, \mathbf{B}$ *are two* $n \times n$ *complex valued matrices then*[5]

$$\|\mathbf{AB}\| \geq \|\mathbf{A}\|\lambda_{\min}\left(\mathbf{BB}'\right)^{1/2}, \tag{24}$$

$$\|\mathbf{AB}\| \geq \|\mathbf{B}\|\lambda_{\min}\left(\mathbf{AA}'\right)^{1/2}. \tag{25}$$

Applying (24)-(25) to $\rho\left[\mathbf{F}_z(\omega_0)\right]$, *we obtain*

$$\begin{aligned}\rho\left[\mathbf{F}_z(\omega_0)\right] &= \|\mathbf{F}_z(\omega_0)\| = \frac{1}{2\pi}\left\|(\mathbf{I}_N - e^{-i\omega_0}\mathbf{\Phi})^{-1}\mathbf{\Sigma}(\mathbf{I}_N - e^{i\omega_0}\mathbf{\Phi}')^{-1}\right\| \\ &\geq \frac{1}{2\pi}\left\|(\mathbf{I}_N - e^{-i\omega_0}\mathbf{\Phi})^{-1}\mathbf{\Sigma}\right\|\lambda_{\min}\left[(\mathbf{I}_N - e^{i\omega_0}\mathbf{\Phi}')^{-1}(\mathbf{I}_N - e^{-i\omega_0}\mathbf{\Phi})^{-1}\right]^{1/2} \\ &\geq \frac{1}{2\pi}\|\mathbf{\Sigma}\|\lambda_{\min}\left[(\mathbf{I}_N - e^{-i\omega_0}\mathbf{\Phi})^{-1}(\mathbf{I}_N - e^{i\omega_0}\mathbf{\Phi}')^{-1}\right]^{1/2}\lambda_{\min}\left[(\mathbf{I}_N - e^{i\omega_0}\mathbf{\Phi}')^{-1}(\mathbf{I}_N - e^{-i\omega_0}\mathbf{\Phi})^{-1}\right]^{1/2} \\ &= \frac{1}{2\pi}\rho(\mathbf{\Sigma})\lambda_{\min}\left[(\mathbf{I}_N - e^{-i\omega_0}\mathbf{\Phi})^{-1}(\mathbf{I}_N - e^{i\omega_0}\mathbf{\Phi}')^{-1}\right] > 0.\end{aligned}$$

Given that $\rho\left[\mathbf{F}_z(\omega)\right] \leq K < \infty$ *at all frequencies* ω, *it must follow that* $\rho(\mathbf{\Sigma}) \leq K < \infty$.

[5] See Bernstein (2005), page 362.

4 Common factor models

Consider the following infinite factor model for $\{z_{it}\}$:

$$z_{it} = \gamma_{i1} f_{1t} + \gamma_{i2} f_{2t} + ... + \gamma_{iN} f_{Nt} + \varepsilon_{it}, \quad i = 1, ..., N, \tag{26}$$

where the common factors, $f_{\ell t}$, and the idiosyncratic errors, ε_{it}, satisfy the following assumptions:

Assumption 6 *The $N \times 1$ vector $\mathbf{f}_t$ is a covariance stationary process, with absolute summable autocovariances, distributed independently of $\varepsilon_{it'}$ for all i, t, t', and such that $E(f_{\ell t}^2 | \mathcal{I}_{t-1}) = 1$ and $E(f_{\ell t} f_{pt} | \mathcal{I}_{t-1}) = 0$, for $\ell \neq p = 1, 2, ..., N$.*

Assumption 7 *$Var(\varepsilon_{it} | \mathcal{I}_{t-1}) = \sigma_i^2 \leq K < \infty$, and ε_{it}, ε_{jt} are independently distributed for all $i \neq j$ and for all t.*

The process z_{it} in (26) has conditional variance

$$Var(z_{it} | \mathcal{I}_{t-1}) = Var(u_{it} | \mathcal{I}_{t-1}) + Var(\varepsilon_{it} | \mathcal{I}_{t-1}) = \sum_{\ell=1}^{N} \gamma_{i\ell}^2 + \sigma_i^2.$$

Finiteness of the conditional variance of z_{it} as stated in Assumption 3 implies that

$$\sum_{\ell=1}^{N} \gamma_{i\ell}^2 \leq K < \infty, \text{ for } i = 1, ..., N. \tag{27}$$

This could arise if, for example,

$$\gamma_{i\ell} = O(1), \text{ for } \ell = 1, ..., m; \ i = 1, ..., N, \tag{28}$$

$$\gamma_{i\ell} = O\left(\frac{1}{\sqrt{N}}\right), \text{ for } \ell = m+1, ..., N; \ i = 1, ..., N, \tag{29}$$

where $0 \leq m < \infty$ does not depend on N.

We now introduce the definition of weak and strong factors.

Definition 3 *(Weak and strong factors) The factor $f_{\ell t}$ is said to be strong if*

$$\lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} E|\gamma_{i\ell}| = K > 0. \tag{30}$$

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

$$\lim_{N \to \infty} \sum_{i=1}^{N} E|\gamma_{i\ell}| = K < \infty. \tag{31}$$

In the case where the loadings attached to $f_{\ell t}$ do not satisfy either of the above conditions (30)-(31), we refer to the corresponding common factor $f_{\ell t}$ as semi-weak (or semi-strong). For example, a factor is semi-weak when the the absolute sum of its loadings, $\sum_{i=1}^{N} E\left|\gamma_{i\ell}\right|$, increases at a rate slower than N.

There exists a relationship between the notions of CSD and CWD and the definitions of weak and strong factors. This is provided in the following theorem.

Theorem 1 *Consider the factor model (26), and suppose that Assumptions 3-7 hold and factor loadings are non-stochastic. Then under the condition that* $\lim_{N\to\infty}\sum_{\ell=1}^{N}\left|\gamma_{i\ell}\right| = K < \infty$ *(for any* $i \in \mathbb{N}$*), the following statements hold:*

(i) The process $\{z_{it}\}$ *is cross sectionally weakly dependent at a given point in time* $t \in \mathcal{T}$ *if* $f_{\ell t}$ *is weak for* $\ell = 1, ..., N$*.*

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

Proof. In matrix form, the covariance of $\mathbf{z}_t = (z_{1t}, ..., z_{Nt})'$ is

$$\mathbf{\Sigma}_t = \mathbf{\Gamma}\mathbf{\Gamma}' + \mathbf{\Lambda}_{\varepsilon}.$$

where $\mathbf{\Lambda}_{\varepsilon}$ is a diagonal matrix with elements σ_i^2. If $f_{\ell t}$ is weak for $\ell = 1, ..., N$ then $\left\|\mathbf{\Gamma}\right\|_1$ is bounded in N, and

$$\lambda_1\left(\mathbf{\Sigma}_t\right) \leq \left\|\mathbf{\Gamma}\mathbf{\Gamma}' + \mathbf{\Lambda}_{\varepsilon}\right\|_1 \leq \left\|\mathbf{\Gamma}\right\|_1\left\|\mathbf{\Gamma}'\right\|_1 + \sigma_{\max}^2 \leq K, \tag{32}$$

and, from Proposition 1, $\{z_{it}\}$ is CWD, which proves point (i). Now suppose that $\{z_{it}\}$ is CSD. Then

$$0 < \lim_{N\to\infty}\frac{1}{N}\lambda_1\left(\mathbf{\Sigma}_t\right) \leq \lim_{N\to\infty}\frac{1}{N}\left\|\mathbf{\Gamma}\right\|_1\left\|\mathbf{\Gamma}'\right\|_1 + \lim_{N\to\infty}\frac{1}{N}\sigma_{\max}^2$$

Given that, by assumption, $\left\|\mathbf{\Gamma}'\right\|_1$ is bounded in N, it follows that $\lim_{N\to\infty}\frac{1}{N}\left\|\mathbf{\Gamma}\right\|_1 = K > 0$, and there exists at least one strong factor in (26). To prove the reverse relation, assume that there exists at least one strong factor in (26) (i.e., $\lim_{N\to\infty}\frac{1}{N}\left\|\mathbf{\Gamma}\right\|_1 = K > 0$). Noting that[6]

$$\lambda_1^{1/2}\left(\mathbf{\Sigma}_t\right) \geq \lambda_1^{1/2}\left(\mathbf{\Gamma}\mathbf{\Gamma}'\right) \geq \frac{\left\|\mathbf{\Gamma}\right\|_1}{\sqrt{N}}. \tag{33}$$

it follows that $\lim_{N\to\infty}\frac{1}{N}\lambda_1\left(\mathbf{\Sigma}_t\right) = K > 0$ and the process is CSD, which proves point (ii). ■

Under (30)-(31), z_{it} can rewritten as

$$z_{it} = u_{it} + e_{it}, \tag{34}$$

[6] See Bernstein (2005), p.368, eq. xiv.

where

$$u_{it} = \sum_{\ell=1}^{m} \gamma_{i\ell} f_{\ell t}; \quad e_{it} = \sum_{\ell=m+1}^{N} \gamma_{i\ell} f_{\ell t} + \varepsilon_{it}, \tag{35}$$

and $\gamma_{i\ell}$ satisfy conditions (30) for $\ell = 1, ..., m$, and (31) for $\ell = m+1, ..., N$. In the light of Theorem 1, it follows that u_{it} is CSD and e_{it} is CWD. Also, notice that when $m = 0$, we have a model with an infinite number of weak factors.

Remark 7 *Consider the following general spatial process*

$$\mathbf{z}_t = \mathbf{R}\mathbf{v}_t, \tag{36}$$

where $\mathbf{R}$ is an $N \times N$ matrix and $\mathbf{v}_t$ is an $N \times 1$ vector of independently distributed random variables. Pesaran and Tosetti (2009) have shown that spatial processes commonly used in the empirical literature, such as the Spatial Autoregressive (SAR) process, or the Spatial Moving Average (SMA), can be written as special cases of (36). Specifically, for a SMA process $\mathbf{R} = \mathbf{I}_N + \delta\mathbf{S}$, where δ is a scalar parameter ($|\delta| < K$) and $\mathbf{S}$ is $N \times N$ nonnegative matrix that expresses the ordering or network linkages among errors, while in the case of an invertible SAR process, we have $\mathbf{R} = (\mathbf{I}_N - \delta\mathbf{S})^{-1}$. Standard spatial literature assumes that $\mathbf{R}$ has bounded column and row norms. It is easy to see that under these conditions the above process can be represented by a factor process with infinite weak factors (i.e., with $m = 0$), and no idiosyncratic error (i.e., $\varepsilon_{it} = 0$). For example by setting

$$z_{it} = \sum_{\ell=1}^{N} \gamma_{i\ell} f_{\ell t},$$

where $\gamma_{i\ell} = r_{i\ell}$, and $f_{\ell t} = v_{\ell t}$ for $i, \ell = 1, ..., N$. Clearly, under the bounded column and row norms of $\mathbf{R}$, the loadings of the above factor structure satisfy (31) and hence carry weak cross section dependence.

Remark 8 *Consistent estimation of factor models with weak or semi-weak factors may be problematic. To see this, consider the following single factor model where suppose that loadings are known*

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

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

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

If for example $\sum_{i=1}^{N} \gamma_i^2$ is bounded, as in the case of weak factors, then $Var\left(\hat{f}_t\right)$ does not vanish as $N \to \infty$, for each t.

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 $\mathbf{u}_t = (u_{1t}, ..., u_{Nt})'$ that constrain the form of cross section dependence carried by the factor structure. For example, Bai (2009) imposes that factor loadings satisfy $\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\gamma_{i\ell}^{2}>0$, for $\ell=1,...,m$. Onatski (2006) and Paul (2007) consider the case where the idyosyncratic errors are independent with a homogeneous variance, σ^2, and consider the ℓth factor as strong if $\sum_{i=1}^{N}\gamma_{i\ell}^{2}>\sqrt{c}\sigma^2$, and weak if $\sum_{i=1}^{N}\gamma_{i\ell}^{2}\leq\sqrt{c}\sigma^2$, where c is such that $\frac{N}{T}-c=o\left(N^{-1/2}\right)$. In the literature on asset pricing models, one common assumption is that $\lambda_m\left(\mathbf{\Sigma}_{ut}\right)$ is bounded away from zero at rate N (Chamberlain (1983); Forni and Lippi (2001)). Consider now factor model (34)-(35). Since $rank\left(\mathbf{\Sigma}_{ut}\right)=m$, and $\lambda_i\left(\mathbf{\Sigma}_{ut}\right)>0$, for $i=1,2,..,m$, and $\lambda_i\left(\mathbf{\Sigma}_{ut}\right)=0$, for $i=m+1,m+2,...,N$, we have

$$\lambda_m\left(\mathbf{\Sigma}_t\right)\geq\lambda_m\left(\mathbf{\Sigma}_{ut}\right),$$

and

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

Under the assumption that $\lambda_m\left(\mathbf{\Sigma}_{ut}\right)$ is bounded away from zero at rate N, and noting that, under (31), $\lambda_1\left(\mathbf{\Sigma}_{et}\right)=O(1)$, it follows that $\lambda_1\left(\mathbf{\Sigma}_t\right),...,\lambda_m\left(\mathbf{\Sigma}_t\right)$ increase without bound as $N\to\infty$, while $\lambda_{m+1}\left(\mathbf{\Sigma}_t\right),...,\lambda_N\left(\mathbf{\Sigma}_t\right)$ satisfy the bounded eigenvalue condition. Most factor structures yield eigenvalues that increase at rate N. But as shown by Kapetanios and Marcellino (2008), it is possible to devise factor models that generate eigenvalues that rise at rate N^d, for $0<d<1$.

Remark 9 *Our concepts of weak and strong cross section dependence are 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_{it}*, for* $i=1,2,...,N$*, denotes the proportion of investor's wealth allocated to the* i^{th} *asset. From Definition 2 it follows that the part of asset returns that is weakly (or semi-weakly) dependent will be fully diversified by portfolios constructed using* $\mathbf{w}_t$ *as the portfolio weights, and as* $N\to\infty$*. Suppose that the asset returns* $\{r_{it}\}$ *have the factor structure*

$$r_{it}=\mu_{i,t-1}+\boldsymbol{\gamma}_i'\mathbf{f}_t+e_{it},\ \ i=1,2,...,N,$$

where $\mu_{i,t-1}$ *is the conditional mean returns,* $\mathbf{f}_t$ *is an* $m\times1$ *vector of unobserved factors,* $\boldsymbol{\gamma}_i$ *is the associated* $m\times1$ *vector of factor loadings, and* $\{e_{it}\}$ *is a CWD process distributed independently of* $\mathbf{f}_t$ *and* $\boldsymbol{\gamma}_i$*. It is assumed that for each i,* e_{it} *is distributed independently of* $\boldsymbol{\gamma}_i$*, whilst* $\mathbf{f}_t$ *follows a general time series process with the conditional* $m\times m$ *covariance matrix,* $\mathbf{\Omega}_t$*, also distributed independently of* e_{it}*. The return on a portfolio constructed with the granular weights* w_{it} *is given by*

$$\rho_t=\textstyle\sum_{i=1}^{N}w_{it}r_{it}=\mathbf{w}_t'\boldsymbol{\mu}_{t-1}+\mathbf{w}_t'\mathbf{\Gamma}\mathbf{f}_t+\mathbf{w}_t'\mathbf{e}_t,$$

where $\boldsymbol{\mu}_{t-1}=(\mu_{1,t-1},\mu_{2,t-1},...,\mu_{N,t-1})'$, $\mathbf{e}_t=(e_{1t},e_{2t},...,e_{Nt})'$*, and* $\mathbf{\Gamma}=(\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_2,...,\boldsymbol{\gamma}_N)'$*. It is easily seen that*

$$Var\left(\rho_t\left|\mathcal{I}_{t-1}\right.\right)=\mathbf{w}_t'\mathbf{\Gamma}\mathbf{\Omega}_t\mathbf{\Gamma}'\mathbf{w}_t+Var\left(\mathbf{w}_t'\mathbf{e}_t\left|\mathcal{I}_{t-1}\right.\right),$$

and since by assumption $\{e_{it}\}$ is a CWD process, then

$$\lim_{N\to\infty} Var\left(\rho_t \left|\mathcal{I}_{t-1}\right.\right) = \lim_{N\to\infty}\left(\mathbf{w}_t'\boldsymbol{\Gamma}\boldsymbol{\Omega}_t\boldsymbol{\Gamma}'\mathbf{w}_t\right).$$

First consider the case where the factors are weak or semi-weak, and note that

$$\mathbf{w}_t'\boldsymbol{\Gamma}\boldsymbol{\Omega}_t\boldsymbol{\Gamma}'\mathbf{w}_t \leq \left(\mathbf{w}_t'\mathbf{w}_t\right)\lambda_1\left(\boldsymbol{\Gamma}\boldsymbol{\Omega}_t\boldsymbol{\Gamma}'\right) \leq \left(\mathbf{w}_t'\mathbf{w}_t\right)\left\|\boldsymbol{\Gamma}\right\|_1\left\|\boldsymbol{\Omega}_t\right\|_1\left\|\boldsymbol{\Gamma}'\right\|_1.$$

Since m is finite then $\left\|\boldsymbol{\Omega}_t\right\|_1\left\|\boldsymbol{\Gamma}'\right\|_1 \leq K$, and the portfolio is fully diversified for all granular weights if

$$\left(\mathbf{w}_t'\mathbf{w}_t\right)\left\|\boldsymbol{\Gamma}\right\|_1 \to 0.$$

This condition holds if $\left\|\boldsymbol{\Gamma}\right\|_1 = O(N^{1-\varepsilon})$ for some positive fixed ε, namely if the factors are weak or semi-weak. In general, however, the portfolio is not fully diversifiable if there is at least one strong factor (see Theorem 1). In the presence of strong factors full diversification is only possible with portfolio weights that are dependent on the factor loadings. One such portfolio weights is given by

$$\mathbf{w}^* = N^{-1}\left[\mathbf{I}_N - \mathbf{M}_\tau\boldsymbol{\Gamma}(\boldsymbol{\Gamma}'\mathbf{M}_\tau\boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma}'\right]\boldsymbol{\tau}_N,$$

where $\mathbf{M}_\tau = \mathbf{I}_N - \boldsymbol{\tau}_N(\boldsymbol{\tau}_N'\boldsymbol{\tau}_N)^{-1}\boldsymbol{\tau}_N'$, and $\boldsymbol{\tau}_N = (1, 1, ..., 1)'$. It is easily seen that the weights $\mathbf{w}^*$ add up to unity and are granular in the sense that[7]

$$\mathbf{w}^{*\prime}\mathbf{w}^* = \frac{1}{N}\left[1 + \left(\frac{\boldsymbol{\tau}_N'\boldsymbol{\Gamma}}{N}\right)\left(\frac{\boldsymbol{\Gamma}'\mathbf{M}_\tau\boldsymbol{\Gamma}}{N}\right)^{-1}\left(\frac{\boldsymbol{\Gamma}'\boldsymbol{\tau}_N}{N}\right)\right] \to 0, \text{ as } N \to \infty.$$

It is also easily seen that $\boldsymbol{\Gamma}'\mathbf{w}^* = \mathbf{0}$. Hence, $\lim_{N\to\infty} Var\left(\mathbf{w}^{*\prime}\mathbf{r}_t \left|\mathcal{I}_{t-1}\right.\right) = 0$, as required.

5 CCE estimation of panel data models with infinite factors

In this section we focus on consistent estimation of a regression model where the error term has a factor structure with infinite factors.

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

$$y_{it} = \boldsymbol{\alpha}_i'\mathbf{d}_t + \boldsymbol{\beta}_i'\mathbf{x}_{it} + u_{it}, \tag{37}$$

where $\mathbf{d}_t = (d_{1t}, d_{2t}, ..., d_{nt})'$ is a $n \times 1$ vector of observed common effects, and $\mathbf{x}_{it}$ is a $k \times 1$ vector of observed individual specific regressors. The parameter of interest is the mean of individual slope coefficients, $\boldsymbol{\beta} = E(\boldsymbol{\beta}_i)$.[8]

[7] When the factors are strong $N^{-1}\boldsymbol{\Gamma}'\boldsymbol{\tau}_N$ and $N^{-1}\boldsymbol{\Gamma}'\mathbf{M}_\tau\boldsymbol{\Gamma}$ are $O(1)$. If some of the factors are weak the columns of $\boldsymbol{\Gamma}$ associated with the weak factors can be removed when constructing the weights, $\mathbf{w}^*$.

[8] We assume that individual slope coefficients are drawn from common distribution with mean $\boldsymbol{\beta}$. In the case

The error term, u_{it}, is given by the following general factor structure,

$$u_{it} = \sum_{\ell=1}^{m_1} \gamma_{i\ell} f_{\ell t} + \sum_{\ell=1}^{m_2} \lambda_{i\ell} g_{\ell t} + e_{it}, \tag{38}$$

where we distinguish between two types of unobserved common factors, $\mathbf{f}_t = (f_{1t}, ..., f_{m_1 t})'$ and $\mathbf{g}_t = (g_{1t}, ..., g_{m_2 t})'$. The former are factors that are possibly correlated with regressors $\mathbf{x}_{it}$, while the latter are not correlated with the regressors. Define for future reference the vectors of factor loadings $\boldsymbol{\gamma}_i = (\gamma_{i1}, ..., \gamma_{im_1})'$ and $\boldsymbol{\lambda}_i = (\lambda_{i1}, ..., \lambda_{im_2})'$.

To model the correlation between the individual specific regressors, $\mathbf{x}_{it}$, and the innovations u_{it}, we suppose that $\mathbf{x}_{it}$ can be correlated with any of the factors in $\mathbf{f}_t$,

$$\mathbf{x}_{it} = \mathbf{A}_i' \mathbf{d}_t + \boldsymbol{\Gamma}_i' \mathbf{f}_t + \mathbf{v}_{it}, \tag{39}$$

where $\mathbf{A}_i'$ and $\boldsymbol{\Gamma}_i'$ are $n \times k$ and $m_1 \times k$ 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 innovations u_{it}, and of the common factors.

Equations (37) and (39) can be written more compactly as

$$\mathbf{z}_{it} = \begin{pmatrix} y_{it} \\ \mathbf{x}_{it} \end{pmatrix} = \mathbf{B}_i' \mathbf{d}_t + \mathbf{C}_i' \mathbf{f}_t + \boldsymbol{\xi}_{it}, \tag{40}$$

where

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

Similar panel data models have been analyzed by Pesaran (2006), Kapetanios, Pesaran, and Yagamata (2009), and Pesaran and Tosetti (2009). Pesaran (2006) introduced CCE estimators in a panel model where m_1 is fixed and $m_2 = 0$, and $\boldsymbol{\gamma}_i' \mathbf{f}_t$ represents a strong factor structure. Contrary to what Bai (2009) (see page 1231) suggests, CCE estimators are valid even in the rank deficient case where m_1 could be larger than $k+1$. Kapetanios, Pesaran, and Yagamata (2009) extended the results of Pesaran (2006) by allowing unobserved common factors to follow unit root processes. In both papers, innovations $\{e_{it}\}$ are assumed to be cross sectionally independent although possibly serially correlated. This assumption is relaxed by Pesaran and Tosetti (2009) who assume that $\{e_{it}\}$ is a weakly dependent process, which includes spatial MA or AR processes considered in the literature as special cases. In this paper, we focus explicitly on cross-correlations modelled by general factor structures - weak, strong, or

where $\boldsymbol{\beta}_i'$s are assumed to be non-stochastic, the object of interest would be cross section mean of $\boldsymbol{\beta}_i$, defined by $\boldsymbol{\beta} = \lim_{N \to \infty} \left(N^{-1} \sum_{i=1}^{N} \boldsymbol{\beta}_i \right)$.

somewhere in between. Our model is thus an extension of Pesaran (2006) to infinite factor structures.

The special case where both m_1 and m_2 are fixed has already been analyzed in the above cited papers. The case where $f_{1t}, ..., f_{m_1 t}$ are strong factors and $m_1 = m_1(N) \rightarrow \infty$ as $N \rightarrow \infty$, is not that meaningful as the variances of u_{it} rise with N. However, it would be possible to let m_2, the number of the weak factors, to rise with N, whilst keeping m_1 fixed. We show below that the CCE estimators continue to be consistent and asymptotically normal under this type of infinite-factor error structures.

We make the following assumptions on the common factors and their loadings:

Assumption 8 *(Common factors) The $(n+m_1)\times 1$ vector $(\mathbf{d}_t', \mathbf{f}_t')'$ is a covariance stationary processes, with absolute summable autocovariances, distributed independently of $g_{it'}$, $e_{it'}$ and $\mathbf{v}_{it'}$ for all i, t and t'.*[9] *For each i, common factor g_{it} follows a linear stationary process with absolute summable autocovariances, zero mean, unit variance, and finite fourth moments. Individual factors collected in vector $\mathbf{g}_t$ are distributed independently of each other and of $e_{it'}$ and $\mathbf{v}_{it'}$ for all i, t and t'.*

Assumption 9 *(Factor loadings) Factor loadings $\boldsymbol{\gamma}_i$, $\boldsymbol{\Gamma}_i$, and $\boldsymbol{\lambda}_i$ are non-stochastic. In addition, we assume that the following conditions hold.*

(a) The unobserved factor loadings, $\boldsymbol{\gamma}_i$ and $\boldsymbol{\Gamma}_i$ are bounded, i.e. $\|\boldsymbol{\gamma}_i\|_2 < K$ and $\|\boldsymbol{\Gamma}_i\|_2 < K$, for all i.

(b) The unobserved factor loadings $\boldsymbol{\lambda}_i$ satisfy the following absolute summability condition for each individual unit,

$$\lim_{N\rightarrow\infty} \sum_{\ell=1}^{m_2} |\lambda_{i\ell}| < K < \infty, \tag{41}$$

where $m_2 = m_2(N)$ is a nondecreasing function of N and the constant K does not depend on i nor on N.

Remark 10 *Factor structure $\boldsymbol{\gamma}_i'\mathbf{f}_t$ could be strong, weak or neither strong nor weak. Note that the number of strong factors cannot increase with N for variance of u_{it} to exists as $N \rightarrow \infty$. We do not impose that $\sum_{\ell=1}^{N} \lambda_{i\ell} g_{\ell t}$ is a weak factor structure.*

Remark 11 *Condition (41) is required for $Var\left(\boldsymbol{\lambda}_i'\mathbf{g}_t\right)$ to exist as $N \rightarrow \infty$. Note that the matrix of factor loadings $\boldsymbol{\Lambda} = (\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, ..., \boldsymbol{\lambda}_N)'$ is not required to have bounded column norm as $N \rightarrow \infty$.*

Remark 12 *It is straightforward to extend the analysis to stochastic factor loadings distributed independently of the errors e_{it}, $\mathbf{v}_{it}$ and the individual coefficients $\boldsymbol{\beta}_i$. In case where factor loadings are non-stochastic, the following rank condition*

$$rank\left(\overline{\mathbf{C}}\right) = m_1 \text{ for all } N, \tag{42}$$

[9] This assumption can be relaxed to allow for unit roots in the common factors, along the lines shown in Kapetanios, Pesaran, and Yagamata (2009).

where $\overline{\mathbf{C}} = N^{-1}\sum_{i=1}^{N}\mathbf{C}_i$, *would have to hold for the consistent inference about* $\boldsymbol{\beta}$. *Regardless of whether the rank condition (42) holds or not, it is straightforward to show, along the same lines as in Pesaran (2006), that the CCE estimators continues to be valid in the case when the factor loadings* $\boldsymbol{\gamma}_i$, *for* $i = 1, .., N$, *are stochastic and distributed independently from the common factors with mean* $\boldsymbol{\gamma}$. *Also see Kapetanios, Pesaran, and Yagamata (2009).*

The remaining assumptions are similar to Pesaran (2006):

Assumption 10 *(Errors) 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*

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

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. For each* i, *the coefficient matrices* $\boldsymbol{\Pi}_{i\ell}$ *satisfy the condition*

$$Var(\mathbf{v}_{it}) = \sum_{\ell=0}^{\infty}\boldsymbol{\Pi}_{i\ell}\boldsymbol{\Pi}_{i\ell}' = \boldsymbol{\Sigma}_{\mathbf{v}_i},$$

where $\boldsymbol{\Sigma}_{\mathbf{v}_i}$ *is a positive definite matrix, such that* $\sup_i \left\|\boldsymbol{\Sigma}_{\mathbf{v}_i}\right\|_2 < K$. *Errors* e_{it}, *for* $i = 1, .., N$, *follow a linear stationary process with absolute summable autocovariances,*

$$\varepsilon_{it} = \sum_{\ell=0}^{\infty} a_{is}\epsilon_{i,t-\ell},$$

where $\epsilon_{is} \sim IID\left(0, 1\right)$ *with finite fourth moments.*

Assumption 11 *(Random coefficients) The slope coefficients follow the random coefficient model*

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

where $\left\|\boldsymbol{\beta}\right\|_2 < K$, $\left\|\boldsymbol{\Omega}_{\upsilon}\right\|_2 < K$, $\boldsymbol{\Omega}_{\upsilon}$ *is symmetric non-negative definite matrix, and the random deviations* $\boldsymbol{\upsilon}_i$ *are distributed independently of* $\mathbf{x}_{jt}$, $\mathbf{d}_t$ *and* u_{jt} *for all* i, j *and* t.

Assumption 12 *Consider the cross section averages of the individual specific variables* $\mathbf{z}_{it} = (y_{it}, \mathbf{x}_{it})'$, *defined by* $\overline{\mathbf{z}}_t = \frac{1}{N}\sum_{i=1}^{N} z_{it}$ *and let* $\overline{\mathbf{M}} = \mathbf{I}_T - \overline{\mathbf{H}}\left(\overline{\mathbf{H}}'\overline{\mathbf{H}}\right)^{-}\overline{\mathbf{H}}$, $\overline{\mathbf{H}} = \left(\mathbf{D}, \overline{\mathbf{Z}}\right)$, *where* $\mathbf{D}$ *and* $\overline{\mathbf{Z}}$ *are, respectively, the matrices of observations on* $\mathbf{d}_t$ *and* $\overline{\mathbf{z}}_t$. *Then the following conditions hold:*

(a) The matrix $\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Sigma}_{\mathbf{v}_i}$ *is finite and nonsingular.*

(b) There exists T_0 *and* N_0 *such that for all* $T \geq T_0$ *and* $N \geq N_0$, *the* $k \times k$ *matrices* $\left(\frac{\mathbf{X}_i\overline{\mathbf{M}}\mathbf{X}_i}{T}\right)^{-1}$ *and* $\left(\frac{\mathbf{X}_i\mathbf{M}_g\mathbf{X}_i}{T}\right)^{-1}$ *exist for all* i, *where* $\mathbf{M}_g = \mathbf{I}_T - \mathbf{G}\left(\mathbf{G}'\mathbf{G}\right)^{-}\mathbf{G}$, *with* $\mathbf{G} = (\mathbf{D}, \mathbf{F})$, $\mathbf{F}$ *and*

$\mathbf{X}_i$ *are matrices of observations on* $\mathbf{f}_t$ *and* $\mathbf{x}_{it}$. *Furthermore,* $\sup_i \|E(\widetilde{\mathbf{v}}_{it}\widetilde{\mathbf{v}}_{it}')\| < K < \infty$ *and* $\sup_i \|E(\mathbf{w}_{it}\mathbf{w}_{it}')\| < K < \infty$, *where* $\widetilde{\mathbf{v}}_{it}'$ *and* $\mathbf{w}_{it}'$ *are* t*-th rows of the matrices* $\widetilde{\mathbf{V}}_i = \mathbf{M}_g\mathbf{V}_i$ *and* $\mathbf{W}_i = \mathbf{M}_g\mathbf{V}_i\left(\frac{\mathbf{V}_i'\mathbf{M}_g\mathbf{V}_i}{T}\right)^{-1}$, *respectively, and* $\mathbf{V}_i = (\mathbf{v}_{i1}, \mathbf{v}_{i2}, ..., \mathbf{v}_{iT})'$.

Remark 13 *For ease of exposition in this section we consider augmentation by arithmetic cross section averages. However, it is straightforward to relax this assumption along the lines of Pesaran (2006) and consider cross section averages that are constructed using more general weights satisfying granularity conditions (2)-(3).*

The idea underlying the CCE approach is that as far as estimation of the slope coefficients are concerned the unobservable common factors can be well approximated by the cross section averages of the dependent variable and individual specific regressors. The common correlated mean group estimator (CCEMG) is given by

$$\widehat{\boldsymbol{\beta}}_{CCEMG} = \frac{1}{N}\sum_{i=1}^{N}\widehat{\boldsymbol{\beta}}_{CCE,i} \tag{43}$$

where the estimates of the individual slopes are

$$\widehat{\boldsymbol{\beta}}_{CCE,i} = \left(\mathbf{X}_i'\overline{\mathbf{M}}\mathbf{X}_i\right)^{-1}\mathbf{X}_i'\overline{\mathbf{M}}\mathbf{y}_i.$$

The common correlated pooled (CCEP) estimator is defined by

$$\widehat{\boldsymbol{\beta}}_{CCEP} = \left(\sum_{i=1}^{N}\mathbf{X}_i'\overline{\mathbf{M}}\mathbf{X}_i\right)^{-1}\sum_{i=1}^{N}\mathbf{X}_i'\overline{\mathbf{M}}\mathbf{y}_i. \tag{44}$$

Theorem 2 *(CCE estimation) Consider the panel data model (37) and (39) and suppose that Assumptions 8-12 hold,* m_1 *does not vary with* N, *and the rank condition (42) holds. Then for the common correlated mean group estimator* $\widehat{\boldsymbol{\beta}}_{CCEMG}$ *given by (43), as* $m_2, N, T \xrightarrow{j} \infty$, *such that* $N\sum_{\ell=1}^{m_2}\overline{\lambda}_\ell\lambda_i < K < \infty$, *we have*

$$\sqrt{N}\left(\widehat{\boldsymbol{\beta}}_{CCEMG} - \boldsymbol{\beta}\right) \to N\left(0, \boldsymbol{\Sigma}_{CCEMG}\right), \tag{45}$$

where $\boldsymbol{\Sigma}_{CCEMG} = \boldsymbol{\Omega}_v$. *If in addition,* $\|\boldsymbol{\Omega}_v\|_2 > 0$, *then for the common correlated pooled estimator* $\widehat{\boldsymbol{\beta}}_{CCEP}$ *given by (44) we have*

$$\sqrt{N}\left(\widehat{\boldsymbol{\beta}}_{CCEP} - \boldsymbol{\beta}\right) \to N\left(0, \boldsymbol{\Sigma}_{CCEP}\right), \tag{46}$$

where

$$\boldsymbol{\Sigma}_{CCEP} = \boldsymbol{\Psi}^{*-1}\mathbf{R}^*\boldsymbol{\Psi}^{*-1},$$

with

$$\boldsymbol{\Psi}^* = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{\Sigma}_{\mathbf{v}_i},$$

$$\mathbf{R}^* = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{\Sigma}_{\mathbf{v}_i} \boldsymbol{\Omega}_{\upsilon} \boldsymbol{\Sigma}_{\mathbf{v}_i}.$$

Proof. Proof is relegated to Appendix. ∎

Consistent estimators for the variances of $\widehat{\boldsymbol{\beta}}_{CCEMG}$ and $\widehat{\boldsymbol{\beta}}_{CCEP}$ are given equation (58) and (69) of Pesaran (2006), respectively. In case of homogenous slopes, namely $\boldsymbol{\Omega}_{\upsilon} = \mathbf{0}$, $\widehat{\boldsymbol{\beta}}_{CCEP}$ continues to be consistent, but in this case $\widehat{\boldsymbol{\beta}}_{CCEP} - \boldsymbol{\beta}$ should be multiplied by $\sqrt{NT}$ instead of $\sqrt{N}$, to obtain a non-degenerate asymptotic distribution. See Pesaran (2006) for more details.

Remark 14 *Besides the absolute summability condition in Assumption 9.b, additional restriction on factor loadings* $\{\lambda_{i\ell}\}$ *in Theorem 2 is that for each* i*,*

$$N \sum_{\ell=0}^{m_2} \overline{\lambda}_{\ell} \lambda_{i\ell} < K < \infty, \text{ as } m_2, N, T \overset{j}{\to} \infty, \tag{47}$$

where $m_2 = m_2(N)$ *and the constant* K *does not depend on* i *and /or* N*. These conditions rule out strong factor structures, but allow for (possibly) an infinite number of weak of semi-weak factors influencing* y_{it}*. In particular, we do not necessarily require bounded column norm of the factor loading matrix* $\boldsymbol{\Lambda}$*.*[10] *For example,* $\overline{\lambda}_{\ell} = O(N^{-1})$*,* $m_2 = \sqrt{N}$ *and* $\lambda_{i\ell} = O(N^{-1/2})$ *satisfy condition (47) and Assumption 9.b. In Monte Carlo experiments below, we also investigate performance of CCE estimators in case of infinite semi-strong (weak) factor structures where condition (47) is not necessarily satisfied.*

Remark 15 *As mentioned in Remark 12, rank condition (42) can be relaxed, along the lines of Pesaran (2006) or Kapetanios, Pesaran, and Yagamata (2009), at the expense of requiring the factor loadings,* $\boldsymbol{\gamma}_i$*, to be random and distributed independently of the common factors and the individual specific errors. Hence CCE estimators are valid for any finite (fixed) number of possibly strong common factors, which are correlated with regressors, and, in addition, innovations could follow a general infinite weak factor structure, or a certain semi-strong (semi-weak) infinite factor structure, or, as shown in Pesaran and Tosetti (2009), could simply follow a spatial model.*

6 Monte Carlo experiments

We consider the following data generating process

$$y_{it} = \alpha_i d_{1t} + \beta_{i1} x_{i1t} + \beta_{i2} x_{i2t} + u_{it}, \tag{48}$$

[10] The bounded row and column norms of $\boldsymbol{\Lambda}$ are sufficient (but not necessary) for condition (47) and Assumption 9.b to hold.

for $i=1,2,...,N$ and $t=1,2,...,T$. We assume heterogeneous slopes, and set $\beta_{ij}=\beta_j+\eta_{ij}$, with $\eta_{ij}\sim IIDN\left(1,0.04\right)$,for $i=1,2,...,N$ and $j=1,2$, varying across replications.

The errors, u_{it}, are generated as

$$u_{it}=\textstyle\sum_{\ell=1}^{3}\gamma_{i\ell}f_{\ell t}+\sum_{\ell=1}^{m_2}\lambda_{i\ell}g_{\ell t}+\varepsilon_{it},$$

where $\varepsilon_{it}\sim N(0,\sigma_i^2)$, $\sigma_i^2\sim IIDU\left(0.5,1.5\right)$, for $i=1,2,...,N$ (the MC results will be robust to serial correlation in ε_{it}), and unobserved common factors are generated as an independent AR(1) processes with unit variance.

$$\begin{aligned}
f_{\ell t} &= 0.5f_{\ell t-1}+v_{f_{\ell t}},\;\; \ell=1,..,3;\;\; t=-49,...,0,1,..,T,\\
v_{f_{\ell t}} &\sim IIDN(0,1-0.5^2),\;\; f_{\ell,-50}=0,\\
g_{\ell t} &= 0.5g_{\ell t-1}+v_{g_{\ell t}},\;\; \ell=1,..,m_2;\;\; t=-49,...,0,1,..,T,\\
v_{g_{\ell t}} &\sim IIDN(0,1-0.5^2),\;\; g_{\ell,-50}=0.
\end{aligned}$$

The first three factors will be assumed to be strong in the sense that their loadings are unbounded in N and are generated as

$$\gamma_{i\ell}\sim IIDU(0,1),\text{ for } i=1,...,N,\ell=1,2,3.$$

The following two cases are considered for the remaining m_2 factors $g_{\ell t}$:

Experiment A $\{g_{\ell t}\}$ are weak, with their loadings given by

$$\lambda_{i\ell}=\frac{\eta_{i\ell}}{2\sum_{i=1}^{N}\eta_{i\ell}},\;\eta_{i\ell}\sim IIDU(0,1),\text{ for }\ell=1,...,m_2,\text{ and } i=1,2,...,N.$$

It is easily seen that for each ℓ, $\sum_{i=1}^{N}|\lambda_{i\ell}|=O(1)$ and for each i, $\sum_{\ell=1}^{m_2}\lambda_{i\ell}^2=O(m/N^2)$. Therefore, asymptotically as $N\to\infty$, the R_i^2 are only affected by the strong factors, even if $m_2\to\infty$.

Experiment B As an intermediate case we shall also consider semi-strong (weak) factors where the loadings are generated by

$$\lambda_{i\ell}=\frac{\eta_{i\ell}}{\sqrt{3\sum_{i=1}^{N}\eta_{i\ell}^2}},\;for\;\ell=1,...,m_2,\text{ and } i=1,2,...,N.$$

In this case, for each ℓ, $\sum_{i=1}^{N}|\lambda_{i\ell}|=O(N^{1/2})$, and for each i, $\sum_{\ell=1}^{m_2}\lambda_{i\ell}^2=O(m_2/N)$, and the signal-to-noise ratio of the regressions deteriorate as m_2 is increased for any given N. In Section 6.1, we will investigate this issue further, to check if the effect of m_2 on R_i^2 for a given N impacts on the performance of our estimators.

The remaining variables in the panel data model are set out as follows: regressors x_{ijt} are assumed to be correlated with strong unobserved common factors and generated as follows:

$$x_{ijt} = a_{ij1}d_{1t} + a_{ij2}d_{2t} + \sum_{\ell=1}^{3} \gamma_{ij\ell} f_{\ell t} + v_{ijt}, \quad j = 1, 2,$$

where

$$\gamma_{ij\ell} \sim IIDU(0, 1), \text{ for } i = 1, ..., N, \ell = 1, 2, 3; j = 1, 2.$$

$$\begin{aligned} v_{ijt} &= \rho_{v_{ij}} v_{ijt-1} + \vartheta_{ijt}, \quad i = 1, 2, ..., N; t = -49, ..., 0, 1, .., T, \\ \vartheta_{ijt} &\sim IIDN(0, 1 - \rho_{\vartheta_{ij}}^2), \quad v_{ij,-50} = 0, \\ \rho_{\vartheta_{ij}} &\sim IIDU(0.05, 0.95) \text{ for } j = 1, 2. \end{aligned}$$

The observed common effects are generated as

$$\begin{aligned} d_{1t} &= 1; d_{2t} = 0.5 d_{2t-1} + v_{dt}, \quad t = -49, ..., 0, 1, .., T, \\ v_{dt} &\sim IIDN(0, 1 - 0.5^2), \quad d_{2,-50} = 0, \end{aligned}$$

When generating v_{ijt} and the common factors $f_{\ell t}, g_{\ell t}$ and d_{2t} the first 50 observations have been discarded to reduce the effect on estimates of initial values. The factor loadings of the observed common effects do not change across replications and are generated as

$$\begin{aligned} \alpha_i &\sim IIDN(1, 1), \quad i = 1, 2, ..., N, \\ (a_{i11}, a_{i21}, a_{i12}, a_{i22}) &\sim IIDN(0.5\boldsymbol{\tau}_4, 0.5\mathbf{I}_4), \end{aligned}$$

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

Each experiment was replicated 2,000 times for all pairs of N and $T = 20, 30, 50, 100, 200$. For each N we shall consider $m = 0, N/5, 3N/5, N$. For example, for $N = 100$, we consider $m = 0, 20, 60, 100$. We report bias, RMSE, size and power for six estimators: the FE estimator with standard variance, the CCEMG and CCEP estimators given by (43) and (44), respectively, the MGPC and PPC estimators proposed by Kapetanios and Pesaran (2007), and the PC estimator proposed by Bai (2009). The MGPC and PPC estimators are similar to (43) and (44) except that $\bar{\mathbf{z}}_t = (\bar{y}_t, \bar{\mathbf{x}}_t')'$ is replaced by $\hat{\mathbf{F}}$ computed as the $T \times (m + n)$ matrix of observations on $\hat{\mathbf{f}}_t$, the vector of $(m + n)$ principal components extracted from $\mathbf{z}_{it} = (y_{it}, \mathbf{x}_{it}')'$. In the PC iterative estimator by Bai (2009), $\left(\hat{\mathbf{b}}_{PC}, \hat{\mathbf{F}}\right)$ is the solution to the following set of non-linear equations:

$$\hat{\mathbf{b}}_{PC} = \left(\sum_{i=1}^{N} \mathbf{X}_i \mathbf{M}_{\hat{F}} \mathbf{X}_i \right)^{-1} \sum_{i=1}^{N} \mathbf{X}_i \mathbf{M}_{\hat{F}} \mathbf{y}_i,$$

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

where $\mathbf{M}_{\hat{F}} = \mathbf{I}_T - \hat{\mathbf{F}}\left(\hat{\mathbf{F}}\hat{\mathbf{F}}'\right)^{-1}\hat{\mathbf{F}}'$, and $\hat{\mathbf{V}}$ is a diagonal matrix with the m largest eigenvalues of the matrix $\frac{1}{NT}\sum_{i=1}^{N}\left(\mathbf{y}_i - \mathbf{X}_i\hat{\mathbf{b}}_{PC}\right)\left(\mathbf{y}_i - \mathbf{X}_i\hat{\mathbf{b}}_{PC}\right)'$ arranged in decreasing order. The demeaning operator is applied to all variables before entering in the iterative procedure, to get rid of the fixed effects. The variance estimator of $\hat{\mathbf{b}}_{PC}$ is

$$\widehat{Var}\left(\hat{\mathbf{b}}_{PC}\right) = \frac{1}{NT}\mathbf{D}_0^{-1}\mathbf{D}_Z\mathbf{D}_0^{-1},$$

where

$$\begin{aligned}\mathbf{D}_0 &= \frac{1}{NT}\sum_{i=1}^{N}\mathbf{Z}_i\mathbf{Z}_i', \\ \mathbf{D}_Z &= \frac{1}{N}\sum_{i=1}^{N}\left(\hat{\sigma}_i^2\frac{1}{T}\sum_{t=1}^{T}\mathbf{z}_{it}\mathbf{z}_{it}'\right),\end{aligned}$$

with $\hat{\sigma}_i^2 = \frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_{it}^2$, $\mathbf{Z}_i = \mathbf{M}_{\hat{F}}\mathbf{X}_i - \frac{1}{N}\sum_{k=1}^{N}\left[\hat{\boldsymbol{\gamma}}_i'\left(\hat{\mathbf{L}}\hat{\mathbf{L}}'/N\right)^{-1}\hat{\boldsymbol{\gamma}}_k'\right]\mathbf{M}_{\hat{F}}\mathbf{X}_k$, and $\hat{\mathbf{L}} = (\hat{\boldsymbol{\gamma}}_1, ..., \hat{\boldsymbol{\gamma}}_N)'$ is the matrix of estimated factor loadings. When $T/N \to \rho > 0$, $\hat{\mathbf{b}}_{PC}$ is biased and, following Bai (2009), we estimate the bias as

$$bias = -\frac{1}{N}\mathbf{D}_0^{-1}\frac{1}{N}\sum_{i=1}^{N}\frac{\left(\mathbf{X}_i - \hat{\mathbf{V}}_i\right)'\hat{\mathbf{F}}}{T}\left(\frac{\hat{\mathbf{L}}\hat{\mathbf{L}}'}{N}\right)^{-1}\hat{\boldsymbol{\gamma}}_i\hat{\sigma}_i^2,$$

where $\hat{\mathbf{V}}_i = \frac{1}{N}\sum_{j=1}^{N}\hat{\boldsymbol{\gamma}}_i'\left(\hat{\mathbf{L}}\hat{\mathbf{L}}'/N\right)^{-1}\hat{\boldsymbol{\gamma}}_j'\mathbf{X}_j$.

6.1 Results

Results on the estimation of the slope parameters for the Experiments A and B are summarized in Tables 1-11. In what follows, we focus on the estimation of β_1; results for β_2 are very similar and are not reported. Notice that the power of the various tests is computed under the alternative $H_1 : \beta_1 = 0.95$.

Results reported in Tables 1 and 2 show that, as expected, the fixed effects estimator performs very poorly, is substantially biased, and is subject to large size distortions for all pairs of N and T, and for all values of m_2. Tables 3-6 show the results for the CCE estimators. The bias and RMSE of CCEP and the CCEMG estimators fall steadily with the sample size and tests of the null hypothesis based on them are correctly sized, regardless of whether the factors, $\{g_{\ell t}, \ell = 1, 2, ..., m_2\}$, are weak or semi-weak, and the choice of m_2. Further, we notice that the power of the tests based on CCE estimators is not affected by m_2, the number of weak (or semi-weak) factors. This is also confirmed by Figure 1, which shows that the power curves of tests based on the CCEP estimator do not change much with m_2.[11] The Monte Carlo results clearly show that augmenting the regression with cross

[11] Similar curves were obtained for CCEMG estimatos, which are not reported due to space considerations.

section averages seems to work well not only in the case of a few strong common factors, but also in the presence of an arbitrary, possibly infinite, number of (semi-) weak factors.

Tables 7-10 report the findings for the MGPC and PPC. First notice that these estimators, since they estimate the unobserved common factors by principal components analysis, only work in the case where the factors, $\{g_{\ell t}\}$, represent a set of weak factors, or when $m_2 = 0$ (i.e., in Experiment A). In fact, in the case of a semi-weak factor structure the covariance matrix of the idiosyncratic error would not have bounded column norm, a condition required by principal components analysis for consistent estimation of the factors and their loadings. However, as shown in Tables 7-8, even for Experiment A, these estimators show some distortions for small values of N (i.e., when $N = 20, 30$). One possible reason for this result is that the principal components approach requires estimating the number of (strong) factors via a selection criterion, which in turn introduces an additional source of uncertainty into the analysis. Therefore, not surprisingly tests based on MGPC and PPC estimators are severely oversized when a semi-weak (semi-strong) factor structure is considered.

Finally, Table 11 gives the results for the Bai (2009) PC iterative estimator. The bias and RMSE of the Bai estimates are comparable to CCE type estimators, but tests based on them are grossly over-sized, even when $m_2 = 0$. The problem seems to lie with the variance of the Bai estimators, an issue that clearly needs further investigation. In his Monte Carlo experiments, Bai does not provide size and power estimates of tests based on his estimator.

7 Concluding remarks

Cross section dependence is a rapidly growing field of study in panel data analysis. In this paper we have introduced the notions of weak and strong cross section dependence, and have shown that these are more general and more widely applicable than other characterizations of cross section dependence provided in the existing econometric literature. We have also investigated how our notions of CWD and CSD relate to the properties of common factor models that are widely used for modelling of contemporaneous correlation in regression models. Finally, we have provided further extensions of the CCE procedure advanced in Pesaran (2006) that allow for a large number of weak or semi-weak factors. Under this framework, we have shown that the CCE method still yields consistent estimates of the slope coefficients and the asymptotic normal theory continues to be applicable.

Table 1: MC results for FE estimator.

Experiment A: $m_1 = 3$ strong factors and m_2 weak factors.

		Bias (x100)					RMSE (x100)					Size (x100)					Power (x100)				
m_2	N/T	20	30	50	100	200	20	30	50	100	200	20	30	50	100	200	20	30	50	100	200
0	20	20.43	20.17	20.21	20.12	20.30	22.58	21.82	21.54	21.16	21.22	85.9	93.5	97.1	99.3	100.0	94.6	98.3	99.6	100.0	100.0
0	30	19.28	19.39	19.29	19.36	19.38	20.87	20.60	20.28	20.13	20.06	94.2	97.6	99.2	99.8	100.0	98.5	99.9	99.9	100.0	100.0
0	50	19.71	20.06	19.92	19.96	19.79	20.99	20.97	20.58	20.49	20.20	97.2	99.6	100.0	100.0	100.0	99.7	100.0	100.0	100.0	100.0
0	100	19.29	20.09	20.08	20.00	19.95	20.25	20.73	20.56	20.32	20.19	99.5	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0	200	19.88	19.80	19.75	19.81	19.81	20.67	20.34	20.12	20.03	19.97	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
4	20	20.24	20.19	20.29	19.97	19.96	22.35	21.92	21.65	20.98	20.85	85.2	92.6	97.5	99.4	99.9	94.7	98.1	99.6	100.0	100.0
6	30	20.21	19.62	19.98	19.85	19.83	21.65	20.89	20.94	20.62	20.46	95.5	97.2	99.5	100.0	100.0	99.1	99.8	100.0	100.0	100.0
10	50	17.35	17.57	17.32	17.47	17.30	18.83	18.63	18.12	18.05	17.78	94.7	98.3	99.6	99.9	100.0	98.7	99.8	100.0	100.0	100.0
20	100	21.32	21.10	21.49	21.38	21.43	22.18	21.77	21.95	21.68	21.63	99.8	99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
40	200	19.17	19.07	19.29	19.65	19.53	19.99	19.66	19.68	19.90	19.69	99.7	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
12	20	21.88	21.99	22.12	21.84	21.93	23.81	23.64	23.38	22.88	22.77	90.7	95.5	98.3	99.5	100.0	97.2	99.2	99.8	100.0	100.0
18	30	16.88	16.89	16.98	16.82	16.91	18.62	18.26	18.06	17.65	17.63	89.3	94.8	98.2	99.6	99.8	97.6	99.2	99.7	99.9	100.0
30	50	20.38	20.41	20.14	20.55	20.57	21.55	21.31	20.81	21.05	20.93	99.0	99.7	100.0	100.0	100.0	99.9	100.0	100.0	100.0	100.0
60	100	21.06	20.93	20.90	21.01	20.99	21.94	21.57	21.34	21.32	21.21	99.7	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
120	200	19.72	20.32	19.98	20.02	20.00	20.49	20.82	20.34	20.25	20.15	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
20	20	22.84	22.53	22.67	22.66	22.56	24.82	24.00	23.84	23.63	23.37	91.1	96.8	99.0	99.6	100.0	97.2	99.3	99.8	100.0	100.0
30	30	17.31	16.97	17.20	17.27	17.24	18.99	18.33	18.25	18.09	17.95	89.8	95.4	98.6	99.6	100.0	97.6	99.2	99.9	100.0	100.0
50	50	21.66	21.70	21.58	21.59	21.48	22.79	22.60	22.25	22.08	21.87	98.8	99.8	100.0	100.0	100.0	99.8	100.0	100.0	100.0	100.0
100	100	19.96	20.25	20.12	20.30	20.27	20.88	20.93	20.60	20.61	20.52	99.7	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
200	200	20.85	20.93	20.99	21.10	21.06	21.58	21.47	21.36	21.33	21.21	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Table 2: MC results for the FE estimator.

Experiment B: $m_1 = 3$ strong factors and m_2 semi-weak factors.

		Bias (x100)					RMSE (x100)					Size (x100)					Power (x100)				
m_2	N/T	20	30	50	100	200	20	30	50	100	200	20	30	50	100	200	20	30	50	100	200
4	20	22.79	22.71	23.10	23.04	23.39	24.63	24.13	24.25	23.95	24.14	91.3	96.8	99.2	99.9	100.0	97.4	99.3	99.9	100.0	100.0
6	30	20.00	20.07	20.01	19.99	20.27	21.63	21.33	20.98	20.70	20.89	93.5	97.5	99.4	100.0	100.0	98.2	99.6	100.0	100.0	100.0
10	50	16.87	17.09	16.96	17.23	16.93	18.36	18.21	17.86	17.80	17.42	93.8	97.5	99.2	100.0	100.0	98.4	99.7	100.0	100.0	100.0
20	100	18.73	18.63	18.79	18.78	18.77	19.81	19.40	19.34	19.14	19.03	98.8	100.0	100.0	100.0	100.0	99.9	100.0	100.0	100.0	100.0
40	200	19.51	19.87	19.94	19.87	19.96	20.35	20.47	20.32	20.12	20.13	99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
12	20	20.91	20.42	20.51	20.53	20.70	23.04	22.16	21.82	21.58	21.54	87.2	93.5	97.3	99.5	99.8	95.6	98.0	99.4	100.0	100.0
18	30	20.13	20.28	20.26	20.34	20.19	21.97	21.64	21.31	21.12	20.84	91.6	96.6	98.9	99.9	99.9	97.4	99.6	99.8	100.0	100.0
30	50	20.75	21.00	21.06	20.91	21.04	22.05	22.02	21.81	21.43	21.44	98.2	99.6	100.0	100.0	100.0	99.7	100.0	100.0	100.0	100.0
60	100	20.85	20.44	20.62	20.71	20.68	21.97	21.23	21.20	21.08	20.94	99.5	99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
120	200	20.28	20.60	20.88	20.75	20.75	21.26	21.28	21.34	21.02	20.92	99.7	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
20	20	20.83	20.86	20.77	20.86	21.02	23.35	22.81	22.21	21.95	21.89	83.5	91.7	96.8	99.2	99.7	92.5	97.5	99.6	99.9	100.0
30	30	20.85	21.66	21.59	21.56	21.59	22.83	23.12	22.71	22.37	22.23	91.4	96.7	99.2	99.9	100.0	97.9	99.3	99.9	100.0	100.0
50	50	21.14	21.36	21.63	21.27	21.33	22.53	22.33	22.39	21.78	21.71	97.7	99.6	99.9	100.0	100.0	99.7	100.0	100.0	100.0	100.0
100	100	19.99	20.09	20.12	20.18	20.13	21.29	21.00	20.77	20.58	20.41	98.6	99.8	100.0	100.0	100.0	99.7	100.0	100.0	100.0	100.0
200	200	19.26	19.23	19.35	19.31	19.23	20.38	20.06	19.88	19.62	19.42	99.4	99.7	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Table 3: MC results for CCEMG estimator.

Experiment A: $m_1 = 3$ strong factors and m_2 weak factors.

		Bias (x100)					RMSE (x100)					Size (x100)					Power (x100)				
m_2	N/T	20	30	50	100	200	20	30	50	100	200	20	30	50	100	200	20	30	50	100	200
0	20	0.04	-0.13	-0.03	0.08	0.01	8.91	7.13	6.13	5.44	5.19	6.80	6.70	7.10	7.10	7.80	11.70	13.05	15.20	19.00	19.70
0	30	0.37	0.11	0.00	0.04	-0.15	7.46	5.88	5.01	4.48	4.13	7.15	5.85	7.30	7.50	6.25	14.45	16.05	19.25	22.70	23.70
0	50	0.04	0.16	0.03	0.10	-0.09	5.86	4.68	3.93	3.42	3.23	5.95	6.45	5.70	6.30	5.15	16.35	20.25	26.80	32.35	34.30
0	100	-0.05	0.06	0.13	0.05	-0.05	4.06	3.34	2.73	2.49	2.34	5.35	5.85	6.05	6.20	5.65	24.60	35.30	46.50	54.65	58.20
0	200	-0.01	-0.04	0.00	-0.06	0.01	3.02	2.32	2.00	1.75	1.65	6.35	4.60	5.35	5.25	5.60	41.30	55.95	71.30	80.70	87.05
4	20	0.14	0.13	-0.16	-0.08	-0.11	8.65	7.13	6.15	5.50	5.03	6.15	7.30	6.90	6.85	7.45	10.40	13.65	15.85	18.20	18.75
6	30	0.20	-0.16	0.04	-0.10	-0.09	7.16	5.87	5.12	4.44	4.28	6.00	6.25	7.65	6.60	8.10	13.35	14.10	20.00	22.35	25.60
10	50	0.08	-0.05	-0.07	-0.06	-0.06	5.82	4.67	3.96	3.47	3.33	6.05	5.95	5.80	6.00	6.25	15.15	19.95	25.90	32.40	34.80
20	100	-0.05	-0.07	0.12	-0.05	0.00	4.05	3.29	2.79	2.48	2.29	5.00	4.65	5.45	6.30	4.80	22.40	32.25	44.45	53.15	59.35
40	200	0.01	-0.12	-0.05	0.04	0.04	2.83	2.35	1.95	1.72	1.63	4.50	5.35	5.10	5.30	4.65	40.50	54.35	71.25	83.70	87.30
12	20	0.04	0.15	-0.02	-0.07	0.06	9.27	7.61	6.20	5.51	5.05	7.75	8.60	7.20	8.15	7.90	11.50	14.35	15.65	19.05	21.15
18	30	-0.14	-0.01	-0.08	-0.20	0.03	7.03	6.02	4.99	4.51	4.20	5.35	5.45	5.95	6.85	6.90	11.35	16.25	18.35	21.75	25.65
30	50	0.17	-0.17	-0.03	-0.04	-0.11	5.69	4.56	3.90	3.43	3.16	5.45	6.05	5.60	5.70	5.00	16.35	19.50	25.30	31.50	34.50
60	100	0.06	0.08	-0.03	-0.03	0.01	3.97	3.30	2.74	2.52	2.24	4.90	5.15	5.05	6.60	5.25	23.75	34.15	44.85	53.95	58.70
120	200	-0.07	-0.01	-0.07	-0.04	0.04	2.90	2.33	2.01	1.71	1.63	5.45	5.15	5.65	4.50	5.40	40.00	56.80	71.15	82.00	86.65
20	20	0.22	-0.21	0.03	-0.09	-0.11	8.94	7.21	6.12	5.49	5.15	7.50	6.90	7.35	7.80	7.55	11.10	12.45	16.10	18.20	20.25
30	30	0.10	0.09	0.05	0.10	-0.03	7.17	5.89	5.03	4.34	4.20	5.70	6.40	6.55	6.45	7.00	12.65	16.30	19.95	22.10	24.85
50	50	0.20	0.00	-0.03	-0.01	-0.05	5.77	4.70	3.97	3.50	3.24	6.40	5.50	5.70	5.65	5.30	17.20	20.85	26.10	33.15	36.25
100	100	0.08	0.11	-0.03	-0.01	-0.06	4.13	3.28	2.82	2.44	2.34	5.95	5.20	5.60	5.55	5.90	24.90	34.20	45.00	54.20	59.10
200	200	0.12	0.06	-0.05	0.01	0.00	2.93	2.37	2.01	1.72	1.63	5.10	5.70	5.65	5.05	5.20	42.15	57.10	71.65	82.30	86.05

Table 4: MC results for CCEMG estimator.

Experiment B: $m_1 = 3$ strong factors and m_2 semi-weak factors.

		Bias (x100)					RMSE (x100)					Size (x100)					Power (x100)				
m_2	N/T	20	30	50	100	200	20	30	50	100	200	20	30	50	100	200	20	30	50	100	200
4	20	-0.10	-0.33	-0.07	-0.01	0.11	8.73	6.99	6.02	5.45	5.02	6.30	6.95	6.10	7.95	7.55	11.75	12.75	15.55	19.40	20.15
6	30	-0.18	-0.06	-0.13	-0.15	-0.08	7.16	5.77	5.03	4.52	4.12	6.45	6.90	6.35	7.35	6.45	12.20	15.90	19.15	22.80	23.90
10	50	0.09	0.09	0.04	0.10	0.00	5.62	4.72	3.93	3.48	3.23	4.60	6.10	5.65	6.30	5.35	15.15	21.60	27.75	32.60	35.70
20	100	-0.19	-0.06	-0.08	-0.01	-0.06	3.93	3.33	2.83	2.43	2.38	5.00	6.00	6.00	5.20	6.20	22.90	33.85	43.05	55.55	58.70
40	200	-0.04	0.07	-0.02	0.00	-0.05	2.83	2.36	1.89	1.74	1.66	5.15	5.45	4.30	4.80	6.20	42.15	59.50	72.10	82.55	86.40
12	20	-0.13	-0.25	-0.30	-0.11	-0.02	8.59	7.19	6.28	5.59	5.01	6.55	6.95	8.25	8.45	6.95	11.40	12.80	15.85	18.50	20.65
18	30	0.04	0.02	0.02	0.05	-0.11	7.31	5.94	5.10	4.47	4.26	7.05	5.85	6.15	6.65	7.15	12.05	15.25	19.05	23.50	24.90
30	50	0.10	0.05	-0.04	0.04	0.06	5.82	4.58	3.88	3.42	3.21	6.75	5.55	5.30	5.65	6.20	16.20	22.05	27.05	31.70	35.25
60	100	-0.11	0.05	-0.04	-0.01	0.05	4.11	3.35	2.78	2.42	2.26	5.40	6.05	5.55	5.10	4.45	24.15	34.55	43.70	54.45	60.40
120	200	-0.07	-0.04	0.01	-0.03	0.00	2.80	2.36	1.93	1.73	1.60	4.90	6.00	4.85	4.70	5.00	40.35	56.70	74.15	81.70	86.80
20	20	-0.12	-0.12	-0.20	-0.12	0.02	8.89	7.38	6.16	5.38	5.06	6.65	7.35	7.30	7.70	6.75	10.15	12.95	14.45	17.10	20.70
30	30	-0.41	0.20	0.03	0.09	-0.01	7.49	6.11	5.13	4.51	4.33	6.30	6.40	6.45	6.45	7.00	10.75	15.50	19.15	23.50	25.05
50	50	0.02	0.10	-0.01	0.06	0.03	5.47	4.49	3.81	3.45	3.16	5.70	5.85	5.35	6.35	5.25	15.95	21.15	26.15	33.00	36.80
100	100	0.17	-0.01	-0.09	-0.07	-0.02	4.05	3.27	2.83	2.46	2.26	5.85	5.30	5.25	5.05	4.35	26.20	34.05	42.30	52.25	58.20
200	200	0.03	-0.07	-0.05	-0.05	0.01	2.81	2.29	1.95	1.77	1.62	4.75	5.35	5.35	6.15	4.40	44.40	56.95	71.75	82.00	86.70

Table 5: MC results for CCEP estimator.

Experiment A: $m_1 = 3$ strong factors and m_2 weak factors.

		Bias (x100)					RMSE (x100)					Size (x100)					Power (x100)				
m_2	N/T	20	30	50	100	200	20	30	50	100	200	20	30	50	100	200	20	30	50	100	200
0	20	0.13	-0.11	0.01	0.04	0.03	8.43	7.12	6.27	5.66	5.39	7.50	7.60	7.90	7.25	7.80	13.00	13.60	16.45	18.20	18.85
0	30	0.24	0.13	-0.01	0.02	-0.15	6.87	5.87	5.22	4.69	4.29	6.45	6.55	7.50	6.85	6.85	14.85	16.15	19.50	22.40	23.30
0	50	0.08	0.08	0.10	0.05	-0.08	5.39	4.59	3.98	3.52	3.30	6.40	6.30	5.75	5.55	5.35	17.60	19.95	26.50	30.75	32.10
0	100	-0.15	0.12	0.13	0.05	-0.04	3.80	3.35	2.78	2.57	2.40	6.00	6.00	4.90	5.45	5.05	25.85	36.70	44.70	51.55	56.15
0	200	0.05	-0.03	0.02	-0.05	0.00	2.82	2.26	2.03	1.80	1.70	5.80	4.60	5.15	5.45	5.60	46.15	57.80	70.20	79.00	84.80
4	20	0.00	0.06	-0.13	-0.13	-0.16	8.13	7.19	6.32	5.67	5.23	6.60	6.65	7.40	7.95	7.50	12.55	14.25	16.00	16.90	17.45
6	30	0.23	-0.19	0.10	-0.13	-0.05	6.68	5.80	5.19	4.63	4.38	6.45	5.95	7.75	7.25	7.30	14.05	14.85	20.80	21.30	24.05
10	50	0.08	-0.02	-0.09	-0.06	-0.06	5.44	4.57	4.01	3.61	3.43	6.90	5.85	5.05	6.40	5.90	17.15	21.70	24.80	30.70	32.70
20	100	-0.11	-0.10	0.15	-0.02	0.00	3.76	3.29	2.82	2.55	2.34	5.00	5.75	5.50	5.45	5.00	26.30	33.10	43.20	51.30	56.80
40	200	-0.03	-0.12	-0.07	0.04	0.03	2.65	2.34	2.00	1.77	1.70	4.60	5.10	5.20	5.30	5.05	46.00	57.05	69.15	81.30	84.75
12	20	0.10	0.13	0.05	-0.10	0.04	8.74	7.49	6.37	5.69	5.22	8.65	8.45	7.50	7.65	7.40	13.75	13.55	16.15	18.35	20.05
18	30	-0.27	-0.02	-0.07	-0.23	0.00	6.71	5.88	5.12	4.65	4.34	6.30	6.70	6.55	6.95	6.75	12.15	16.25	18.75	20.45	24.25
30	50	0.19	-0.20	0.00	-0.05	-0.09	5.27	4.48	3.98	3.56	3.24	5.65	5.65	5.70	6.00	5.10	17.90	19.55	24.60	29.80	32.00
60	100	0.06	0.09	-0.04	-0.03	0.02	3.77	3.23	2.82	2.58	2.33	5.30	5.50	5.15	5.85	4.95	26.60	34.20	42.25	50.95	55.25
120	200	-0.05	0.00	-0.07	-0.03	0.03	2.72	2.29	2.06	1.76	1.68	5.90	5.40	5.70	4.00	5.30	45.15	57.15	69.05	79.15	84.70
20	20	0.22	-0.10	0.00	-0.09	-0.11	8.38	7.13	6.31	5.70	5.32	7.25	6.75	7.65	8.15	7.70	11.30	12.55	16.25	18.30	19.90
30	30	0.24	0.03	0.03	0.10	0.00	6.73	5.83	5.16	4.54	4.36	6.40	6.95	6.90	6.50	7.65	14.25	16.50	19.20	21.15	24.50
50	50	0.26	0.05	-0.04	0.02	-0.02	5.27	4.68	4.07	3.58	3.34	5.65	5.85	5.95	5.60	6.15	17.70	21.15	25.45	32.20	34.65
100	100	-0.04	0.11	0.00	0.01	-0.06	3.82	3.24	2.87	2.53	2.43	5.70	5.55	5.65	5.45	6.50	27.40	36.10	43.10	52.15	56.00
200	200	0.15	0.06	-0.03	0.02	0.00	2.70	2.33	2.04	1.79	1.68	5.75	5.10	5.60	5.45	4.85	49.05	59.20	69.20	80.15	83.75

Table 6: MC results for CCEP estimator.

Experiment B: $m_1 = 3$ strong factors and m_2 semi-weak factors.

		Bias (x100)					**RMSE (x100)**					**Size (x100)**					**Power (x100)**				
m_2	N/T	20	30	50	100	200	20	30	50	100	200	20	30	50	100	200	20	30	50	100	200
4	20	0.06	-0.27	-0.08	0.05	0.15	8.30	7.10	6.20	5.67	5.26	8.60	7.45	6.85	7.50	7.15	12.30	13.85	15.40	18.85	19.80
6	30	-0.20	-0.06	-0.18	-0.19	-0.08	6.68	5.72	5.16	4.66	4.35	6.20	6.60	7.30	7.15	7.00	13.65	15.55	18.70	21.05	22.60
10	50	0.05	0.09	0.05	0.14	-0.03	5.31	4.66	4.04	3.63	3.34	5.55	6.55	5.85	6.10	5.50	17.35	22.10	26.40	31.85	33.80
20	100	-0.12	-0.07	-0.07	-0.01	-0.06	3.77	3.23	2.91	2.50	2.46	5.40	5.05	6.10	4.95	5.65	27.25	33.50	42.55	51.45	55.70
40	200	0.04	0.07	0.01	-0.01	-0.04	2.68	2.32	1.91	1.80	1.72	5.30	5.75	4.10	5.00	5.55	47.90	61.25	71.25	79.95	83.10
12	20	0.08	-0.26	-0.34	-0.10	-0.01	8.19	7.10	6.52	5.83	5.28	6.55	7.25	8.70	8.95	7.10	12.85	13.90	16.50	19.20	20.85
18	30	-0.11	-0.11	0.01	0.05	-0.09	6.96	5.94	5.30	4.65	4.39	7.30	6.35	6.55	6.65	7.50	13.10	15.45	18.15	23.45	23.30
30	50	0.08	0.07	-0.02	0.07	0.08	5.26	4.49	3.95	3.58	3.38	5.95	6.05	5.95	6.00	6.40	16.55	20.80	25.95	30.15	34.20
60	100	-0.07	0.02	-0.05	-0.01	0.03	3.77	3.26	2.78	2.56	2.37	5.90	5.95	5.05	5.60	4.80	26.35	34.70	41.10	50.65	56.60
120	200	-0.10	-0.06	-0.02	-0.02	0.00	2.61	2.33	1.96	1.78	1.65	4.70	5.65	4.45	4.55	4.65	45.35	58.90	70.90	79.10	84.25
20	20	-0.12	-0.15	-0.18	-0.12	-0.04	8.37	7.36	6.41	5.71	5.44	7.00	7.20	7.60	7.85	7.35	11.15	13.80	14.95	17.40	18.70
30	30	-0.27	0.21	0.05	0.09	-0.03	7.00	6.17	5.23	4.74	4.48	6.30	7.10	6.35	6.70	7.00	11.75	16.95	19.60	22.95	23.65
50	50	-0.03	0.19	0.10	0.04	0.01	5.02	4.51	3.99	3.65	3.33	5.45	6.45	5.60	7.30	5.20	17.45	22.70	27.50	30.95	33.95
100	100	0.09	-0.03	-0.07	-0.04	-0.02	3.81	3.25	2.91	2.57	2.39	6.40	5.50	5.95	5.00	5.50	27.65	34.90	41.25	50.10	54.10
200	200	-0.01	-0.03	-0.03	-0.03	0.01	2.64	2.32	2.01	1.83	1.70	5.55	5.65	4.95	6.10	4.95	47.45	59.60	70.75	78.80	83.55

Table 7: MC results for MGPC estimator.

Experiment A: $m_1 = 3$ strong factors and m_2 weak factors.

		Bias (x100)					**RMSE (x100)**					**Size (x100)**					**Power (x100)**				
m_2	N/T	20	30	50	100	200	20	30	50	100	200	20	30	50	100	200	20	30	50	100	200
0	20	-15.93	-10.98	-8.33	-5.97	-4.55	21.34	14.54	11.18	8.28	6.71	22.00	24.30	26.10	22.45	17.75	12.90	12.05	12.10	8.10	6.85
0	30	-9.09	-6.08	-4.67	-3.43	-2.85	14.61	9.54	7.19	5.59	4.83	14.40	15.45	15.75	13.30	12.40	8.00	7.05	7.00	8.70	8.95
0	50	-4.63	-3.01	-2.48	-1.86	-1.59	9.54	6.20	4.76	3.77	3.44	10.55	9.25	9.75	9.30	8.95	6.15	7.25	10.05	16.75	20.25
0	100	-2.16	-1.42	-1.14	-0.89	-0.79	5.87	4.07	3.05	2.54	2.30	6.65	7.45	7.55	7.25	6.90	8.55	18.20	28.90	43.40	50.50
0	200	-1.04	-0.66	-0.66	-0.56	-0.36	3.99	2.69	2.12	1.75	1.56	5.95	5.10	6.30	7.00	6.10	19.25	38.10	57.15	75.80	87.90
4	20	-15.33	-11.03	-8.35	-6.07	-4.63	21.03	14.71	10.97	8.28	6.75	21.60	24.50	24.30	21.80	17.05	13.40	12.45	10.85	7.65	6.60
6	30	-8.99	-5.98	-4.58	-3.51	-2.82	14.08	9.49	7.18	5.61	4.89	13.60	15.35	15.60	14.50	12.35	6.50	6.95	6.95	7.05	10.25
10	50	-4.64	-3.21	-2.49	-1.86	-1.49	9.55	6.19	4.73	3.80	3.45	10.15	8.80	9.20	9.35	8.80	6.45	7.40	11.00	16.65	22.30
20	100	-1.99	-1.58	-1.02	-1.01	-0.75	5.76	4.12	2.98	2.59	2.26	6.15	7.30	6.80	8.75	6.20	9.10	16.55	29.55	41.80	50.75
40	200	-0.98	-0.81	-0.65	-0.44	-0.35	4.02	2.83	2.08	1.69	1.55	5.85	6.95	6.50	6.05	6.00	20.00	37.15	59.30	79.95	86.75
12	20	-15.62	-11.38	-8.52	-6.22	-4.69	21.53	15.17	11.27	8.45	6.76	22.75	25.65	27.20	24.20	17.70	14.15	12.55	11.40	7.85	6.10
18	30	-9.44	-5.91	-4.59	-3.68	-2.73	14.62	9.42	7.18	5.77	4.86	15.35	14.55	15.05	14.70	12.55	7.90	6.70	6.90	8.20	10.80
30	50	-4.83	-3.41	-2.72	-2.07	-1.72	9.63	6.39	4.93	3.92	3.43	9.30	10.55	11.55	10.45	9.15	5.85	7.10	10.40	15.80	18.70
60	100	-2.21	-1.41	-1.25	-0.99	-0.78	6.07	3.97	3.07	2.58	2.28	7.40	6.35	7.35	8.10	7.25	9.15	18.30	26.10	41.80	50.90
120	200	-1.03	-0.68	-0.69	-0.54	-0.34	4.08	2.71	2.15	1.72	1.54	6.50	5.45	6.90	6.20	5.20	19.85	37.65	58.70	77.35	87.00
20	20	-15.69	-11.09	-8.36	-6.14	-4.86	21.50	14.71	10.99	8.36	6.93	23.30	24.50	25.95	22.85	18.15	13.45	12.65	10.20	8.45	7.15
30	30	-9.19	-6.19	-4.78	-3.56	-2.89	14.51	9.67	7.33	5.63	4.91	15.30	16.30	16.75	14.55	12.00	7.55	7.50	6.80	7.00	9.90
50	50	-4.72	-3.31	-2.65	-2.03	-1.65	9.28	6.31	4.84	3.92	3.47	8.45	10.60	10.30	10.00	9.75	5.10	6.80	10.15	16.45	20.60
100	100	-1.89	-1.49	-1.32	-0.99	-0.82	5.99	4.00	3.17	2.53	2.35	8.00	6.70	8.40	7.25	7.40	10.00	16.15	27.50	41.40	50.65
200	200	-0.96	-0.69	-0.63	-0.50	-0.38	4.13	2.78	2.14	1.71	1.57	6.65	6.50	7.50	5.90	6.05	21.70	39.70	59.95	78.30	86.05

Table 8: MC results for MGPC estimator.

Experiment B: $m_1 = 3$ strong factors and m_2 semi-weak factors.

		Bias (x100)					**RMSE (x100)**					**Size (x100)**					**Power (x100)**				
m_2	N/T	20	30	50	100	200	20	30	50	100	200	20	30	50	100	200	20	30	50	100	200
4	20	-16.39	-11.94	-8.72	-6.38	-4.98	22.01	15.35	11.34	8.48	6.93	22.65	26.25	28.10	24.25	19.00	15.00	14.15	11.80	7.80	6.30
6	30	-10.27	-7.17	-5.43	-4.28	-3.47	15.50	10.46	7.79	6.19	5.24	17.55	17.90	18.30	18.95	14.20	9.75	7.75	7.10	6.75	7.20
10	50	-5.35	-4.11	-3.33	-2.61	-2.20	9.85	6.86	5.26	4.24	3.78	10.75	12.35	13.70	13.05	12.80	5.80	6.45	7.70	12.10	15.50
20	100	-3.17	-2.43	-2.06	-1.70	-1.51	6.48	4.56	3.61	2.88	2.69	8.35	10.20	12.05	11.65	12.25	6.90	12.05	19.20	29.80	38.35
40	200	-1.99	-1.54	-1.41	-1.24	-1.09	4.34	3.15	2.43	2.10	1.90	7.75	9.95	10.25	12.15	11.60	12.40	26.10	43.30	61.45	72.30
12	20	-18.36	-13.84	-11.09	-8.17	-6.89	23.38	17.17	13.49	10.08	8.43	26.65	33.75	37.60	32.75	29.05	16.55	19.25	17.75	11.85	7.30
18	30	-11.76	-8.69	-7.16	-5.79	-5.06	16.66	11.59	9.18	7.38	6.49	19.85	22.25	27.05	28.00	27.30	10.30	9.65	8.70	7.35	6.45
30	50	-7.46	-5.67	-4.97	-4.21	-3.74	11.33	8.15	6.58	5.38	4.83	15.25	19.00	22.85	23.75	23.05	6.80	7.00	5.90	6.25	6.75
60	100	-4.82	-4.02	-3.59	-3.12	-2.77	7.61	5.63	4.64	3.90	3.52	14.85	18.10	23.15	26.30	24.70	6.20	6.30	8.05	11.70	18.15
120	200	-3.61	-3.17	-2.93	-2.65	-2.41	5.43	4.28	3.64	3.17	2.84	14.90	21.80	29.45	34.90	34.65	7.00	11.65	18.20	29.70	38.85
20	20	-20.66	-15.49	-12.81	-10.18	-8.27	25.82	18.67	15.02	11.82	9.67	31.80	37.35	44.45	44.05	39.20	21.35	20.90	23.05	17.85	11.85
30	30	-13.53	-9.92	-8.45	-7.04	-6.25	18.07	12.76	10.30	8.42	7.48	23.75	27.40	34.00	35.55	34.95	13.25	11.40	10.55	8.95	6.85
50	50	-8.71	-7.20	-6.46	-5.58	-5.13	12.34	9.30	7.81	6.62	6.03	17.95	25.15	32.65	37.95	39.15	7.75	7.55	7.30	6.60	5.45
100	100	-5.88	-5.39	-5.02	-4.54	-4.12	8.54	6.75	5.92	5.19	4.67	17.40	27.60	39.80	45.75	47.15	6.85	5.65	5.85	6.25	6.35
200	200	-4.81	-4.57	-4.26	-3.82	-3.63	6.57	5.48	4.83	4.23	3.95	23.20	38.00	51.20	59.40	64.30	6.80	7.45	8.20	11.70	14.65

Table 9: MC results for PPC estimator.

Experiment A: $m_1 = 3$ strong factors and m_2 weak factors.

		Bias (x100)					RMSE (x100)					Size (x100)					Power (x100)				
m_2	N/T	20	30	50	100	200	20	30	50	100	200	20	30	50	100	200	20	30	50	100	200
0	20	-14.29	-10.00	-7.59	-5.62	-4.46	18.34	13.39	10.39	8.00	6.66	24.30	24.30	23.30	21.30	17.05	13.30	12.75	10.70	6.85	6.70
0	30	-8.16	-5.54	-4.32	-3.32	-2.90	12.53	8.68	6.86	5.57	4.95	16.60	14.80	14.25	14.05	13.40	8.75	6.35	7.15	8.75	9.00
0	50	-4.25	-2.77	-2.30	-1.89	-1.62	8.10	5.74	4.59	3.81	3.49	11.00	10.10	10.50	9.05	8.15	7.20	8.85	9.95	15.75	19.15
0	100	-1.93	-1.28	-1.08	-0.90	-0.81	4.88	3.72	2.97	2.56	2.32	6.80	7.15	7.10	7.25	6.75	12.55	20.55	30.70	40.70	50.25
0	200	-0.82	-0.63	-0.60	-0.55	-0.39	3.32	2.50	2.07	1.76	1.59	7.05	5.35	5.85	6.40	6.35	28.55	44.45	59.60	74.55	86.45
4	20	-14.10	-10.09	-7.63	-5.71	-4.59	18.44	13.35	10.26	7.98	6.75	23.90	23.10	23.30	19.50	17.05	13.80	12.30	9.80	6.80	6.75
6	30	-8.15	-5.57	-4.26	-3.44	-2.84	12.13	8.76	6.91	5.61	4.93	16.25	16.05	14.75	14.20	12.40	6.75	7.70	7.65	7.10	10.00
10	50	-4.16	-2.92	-2.34	-1.85	-1.55	8.07	5.72	4.56	3.83	3.51	10.60	9.55	8.85	8.80	9.30	7.20	8.00	10.90	16.55	22.10
20	100	-1.86	-1.44	-0.93	-0.98	-0.78	4.82	3.78	2.90	2.60	2.28	6.75	7.15	5.90	7.85	5.95	11.15	19.55	31.90	40.60	50.20
40	200	-0.83	-0.70	-0.60	-0.44	-0.37	3.23	2.55	2.05	1.70	1.58	6.10	6.75	6.00	6.05	5.85	27.30	42.70	60.75	78.80	85.55
12	20	-14.05	-10.13	-7.63	-5.79	-4.52	18.63	13.57	10.44	8.12	6.65	24.40	24.15	24.70	22.55	16.60	13.90	12.00	10.85	7.65	6.35
18	30	-8.60	-5.52	-4.30	-3.60	-2.77	12.48	8.68	6.95	5.76	4.89	16.25	14.95	15.15	14.95	11.80	7.50	7.20	7.10	8.05	9.70
30	50	-4.29	-3.12	-2.50	-2.04	-1.76	7.98	5.84	4.73	3.92	3.47	10.40	9.65	10.05	10.35	9.05	5.85	8.15	11.05	14.95	18.10
60	100	-1.93	-1.25	-1.17	-0.99	-0.82	4.97	3.68	3.03	2.59	2.32	7.30	6.75	6.95	7.95	6.90	12.00	19.65	28.10	40.75	49.10
120	200	-0.90	-0.61	-0.64	-0.53	-0.36	3.25	2.44	2.11	1.74	1.56	6.00	4.85	7.45	5.90	5.80	26.45	43.20	61.10	76.85	85.75
20	20	-13.80	-10.15	-7.64	-5.81	-4.71	18.14	13.46	10.27	8.14	6.87	22.90	25.45	23.50	21.55	18.25	13.40	12.10	8.95	7.30	6.60
30	30	-8.54	-5.59	-4.48	-3.47	-2.89	12.44	8.88	7.03	5.61	4.96	16.80	16.25	15.65	14.15	11.85	8.30	7.25	7.00	7.10	10.25
50	50	-4.09	-2.90	-2.46	-2.00	-1.68	7.70	5.83	4.69	3.91	3.50	9.05	9.55	9.95	9.70	9.20	5.50	9.30	11.05	15.80	20.25
100	100	-1.68	-1.33	-1.22	-0.97	-0.86	4.84	3.69	3.14	2.55	2.38	7.70	7.00	7.80	6.75	8.10	12.30	19.70	28.30	39.50	48.90
200	200	-0.80	-0.59	-0.58	-0.49	-0.40	3.24	2.52	2.08	1.73	1.59	5.90	6.45	6.55	5.90	5.70	27.60	45.15	60.80	77.30	83.95

Table 10: MC results for PPC estimator.

Experiment B: $m_1 = 3$ strong factors and m_2 semi-weak factors.

		Bias (x100)					RMSE (x100)					Size (x100)					Power (x100)				
m_2	N/T	20	30	50	100	200	20	30	50	100	200	20	30	50	100	200	20	30	50	100	200
4	20	-14.67	-10.83	-8.13	-6.12	-4.93	18.93	13.99	10.77	8.31	6.95	24.85	26.45	25.45	23.50	18.50	15.00	12.55	11.25	7.30	6.25
6	30	-9.44	-6.54	-5.16	-4.20	-3.51	13.23	9.55	7.51	6.13	5.31	19.25	18.30	18.20	18.10	14.10	8.90	7.50	7.00	6.00	7.40
10	50	-4.85	-3.62	-3.08	-2.55	-2.24	8.30	6.25	5.03	4.26	3.85	11.85	12.20	11.70	13.10	13.20	6.00	6.80	7.75	12.20	15.60
20	100	-2.57	-2.19	-1.92	-1.67	-1.51	5.30	4.09	3.51	2.89	2.71	9.30	9.60	11.55	10.25	11.60	9.90	13.85	21.65	29.00	37.25
40	200	-1.52	-1.32	-1.27	-1.22	-1.08	3.50	2.82	2.34	2.10	1.91	7.65	8.15	9.25	12.05	11.30	20.00	33.15	47.55	60.90	72.50
12	20	-16.80	-12.67	-10.18	-7.81	-6.75	20.80	15.59	12.52	9.74	8.31	30.25	33.30	34.60	30.40	28.05	18.85	16.70	14.90	10.45	6.80
18	30	-10.55	-7.89	-6.70	-5.59	-5.04	14.24	10.64	8.79	7.23	6.49	22.20	23.05	25.75	26.15	26.65	11.15	9.00	8.70	6.55	6.15
30	50	-6.39	-5.06	-4.60	-4.07	-3.77	9.55	7.26	6.22	5.28	4.87	17.20	17.60	21.65	21.85	23.45	6.85	5.90	6.10	6.25	6.75
60	100	-3.86	-3.43	-3.30	-3.01	-2.75	6.06	4.96	4.33	3.85	3.52	13.55	16.70	20.20	25.65	23.90	5.70	7.90	8.35	12.20	17.90
120	200	-2.79	-2.69	-2.61	-2.56	-2.39	4.28	3.74	3.36	3.09	2.84	14.35	19.55	26.00	32.05	32.95	12.30	16.30	23.25	29.95	38.50
20	20	-18.49	-14.19	-11.79	-9.70	-8.16	22.40	17.04	13.88	11.36	9.58	33.15	37.10	40.85	42.40	37.65	20.50	20.35	19.50	16.15	11.50
30	30	-12.15	-8.99	-7.79	-6.82	-6.22	15.46	11.51	9.65	8.23	7.49	26.25	27.25	30.60	32.80	34.70	12.85	10.35	9.35	8.60	7.45
50	50	-7.35	-6.24	-5.95	-5.45	-5.16	10.13	8.25	7.33	6.55	6.07	19.50	23.95	29.65	35.15	38.85	6.60	7.10	7.10	6.85	5.45
100	100	-4.74	-4.55	-4.50	-4.40	-4.13	6.81	5.88	5.42	5.07	4.69	17.85	25.20	33.65	43.85	46.65	6.35	5.75	5.75	6.30	6.80
200	200	-3.61	-3.73	-3.79	-3.68	-3.61	5.00	4.64	4.37	4.11	3.95	21.20	30.60	44.65	55.90	63.05	8.05	9.75	10.75	13.40	14.95

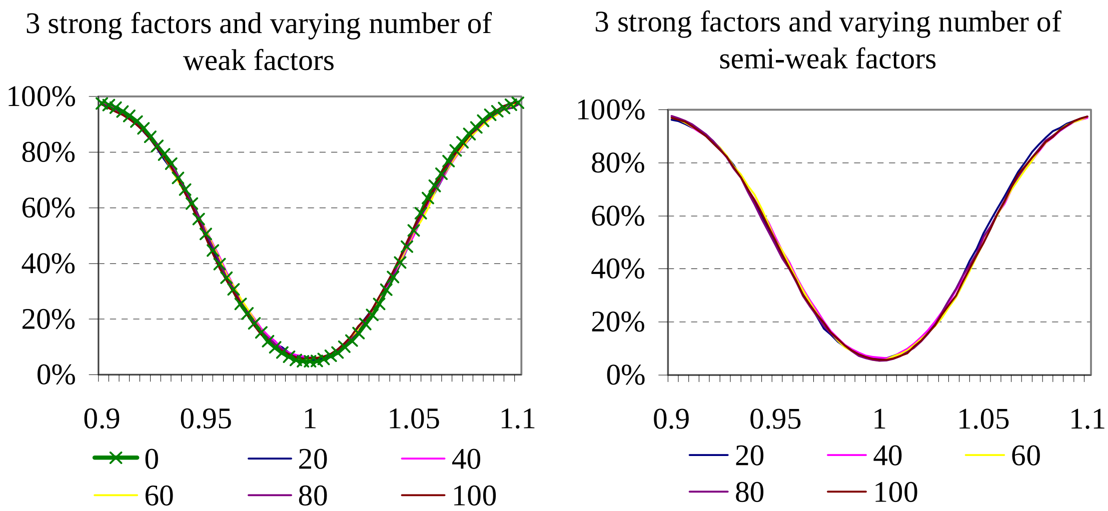

Figure 1: Power curves for the CCEP t-tests in experiments with $N = 100$, $T = 100$, 3 strong factors, and a varying number m_2 of weak factors (left chart) and semi-weak factors (right chart).

Table 11: MC results for Bai estimator.[12]

Experiment A and B: : $m_1 = 3$ strong factors and m_2 weak or semi-weak factors.

		Bias (x100)		**RMSE (x100)**		**Size (x100)**		**Power (x100)**	
m_2	N/T	20	100	20	100	20	100	20	100
		Weak factor structure $\{\boldsymbol{\lambda}_i'\mathbf{g}_t\}$							
0	20	0.47	-0.30	9.78	5.72	37.90	48.00	45.60	61.40
0	100	-0.01	0.02	3.57	2.50	21.50	47.20	58.70	91.10
4	20	0.62	-0.15	9.80	5.83	40.10	50.50	48.30	63.20
20	100	0.07	-0.09	3.48	2.47	21.40	44.90	56.20	91.50
20	20	0.30	0.09	9.91	6.07	37.90	52.40	46.50	64.20
100	100	0.10	0.03	3.47	2.42	21.10	45.30	59.80	91.90
		Semi-weak factor structure $\{\boldsymbol{\lambda}_i'\mathbf{g}_t\}$							
4	20	0.45	-0.23	9.40	6.08	35.50	52.10	42.70	65.10
20	100	-0.09	-0.17	3.70	2.60	23.60	46.80	58.30	88.70
20	20	1.28	-0.28	10.47	6.27	41.70	52.40	49.40	60.50
100	100	0.02	0.03	3.50	2.46	20.90	44.50	56.20	90.20

[12] Based on $R = 1000$ replications.

References

Anderson, B., M. Deistler, A. Filler, and C. Zinner (2009). Generalized linear dynamic factor models - an approach via singular autoregressions. Mimeo, Vienna University of Technology.

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

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

Bernstein, D. S. (2005). *Matrix mathematics: theory, facts, and formulas with application to linear systems theory*. Princeton University Press.

Chamberlain, G. (1983). Funds, factors, and diversification in arbitrage pricing models. *Econometrica 51*, 1281–1304.

Chamberlain, G. and M. Rothschild (1983). Arbitrage, factor structure and mean-variance analysis in large asset market. *Econometrica 51*, 1305–1324.

Chudik, A. and M. H. Pesaran (2009). Infinite dimensional vars and factor models. ECB working paper n. 998.

Connor, G. and R. A. Korajzcyk (1993). A test for the number of factors in an approximate factor structure. *The Journal of Finance 48*, 1263–1291.

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

Forni, M. and L. Reichlin (1998). Let's get real: a factor analytical approach to disaggregated business cycle dynamics. *Review of Economic Studies 65*, 453–473.

Horn, R. A. and C. A. Johnson (1985). *Matrix Analysis*. Cambridge University Press.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology 24*, 417–441 and 498–520.

Ingersoll, J. (1984). Some results in the theory of arbitrage pricing. *Journal of Finance 39*, 1021–1039.

Kapetanios, G. and M. Marcellino (2008). Factor-gmm estimation with large sets of possibly weak instruments. Mimeo, Queen Mary College, University of London.

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

Kapetanios, G., M. H. Pesaran, and T. Yagamata (2009). Panels with nonstationary multifactor error structures. CESifo Working Paper No. 1788, revised June 2009.

Kelejian, H. H. and I. Prucha (1999). A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review 40*, 509–533.

Lee, L. F. (2004). Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica 72*, 1899–1925.

Onatski, A. (2006). Asymptotic distribution of the principal components estimator of large factor models when factors are relatively weak. Mimeo, Columbia University.

Paul, D. (2007). Asymptotics of the leading sample eigenvalues for a spiked covariance model. *Statistica Sinica 17*, 1617–1642.

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

Pesaran, M. H. and E. Tosetti (2009). Large panels with common factors and spatial correlation. Mimeo, University of Cambridge.

Robinson, P. M. (2003). Long memory time series. In P. M. Robinson (Ed.), *Time Series with Long Memory*, pp. 4–32. Oxford University Press.

Ross, S. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory 13*, 341–360.

Stock, J. H. and M. W. Watson (1998). Diffusion indexes. NBER Working Paper No. 6702.

Stock, J. H. and M. W. Watson (2002). Has the business cycle changed and why? NBER Working Paper No. 9127.

Stone, R. N. (1947). On the interdependence of blocks of transactions. *Supplement to Journal of the Royal Statistical Society 9*, 1–45.

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

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.

Appendix

Let $\mathbf{Q} = \mathbf{G}\overline{\mathbf{P}}$, with

$$\overline{\mathbf{P}} = \begin{pmatrix} \mathbf{I}_n & \overline{\mathbf{B}} \\ \mathbf{0} & \overline{\mathbf{C}} \end{pmatrix},$$

and note that $\overline{\mathbf{H}} = \mathbf{Q} + \overline{\mathbf{U}}^*$, $\overline{\mathbf{U}}^* = \left(\mathbf{0}, \overline{\mathbf{U}}\right)$, and $\mathbf{X}_i = \mathbf{G}\mathbf{\Pi}_i + \mathbf{V}_i$.For any random variable x, $\|x\|_{L_1} = E|x|$ denotes L_1 norm of x. For any $k \times 1$ vector of random variables $\mathbf{x}_k = (x_1, ..., x_k)'$, $\|\mathbf{x}_k\|_{L_1} = \sum_{i=1}^{k} E|x_i|$. We use $\overset{L_1}{\to}$ to denote convergence in L_1 norm. We now provide some lemmas useful for proving Theorem 2.

Lemma 1 *Consider the panel data model (37) and (39) and suppose that Assumptions 8-12 hold, and m_1 does not vary with N. Then as $m_2, T, N \overset{j}{\to} \infty$, such that $N\sum_{\ell}^{m_2} \overline{\lambda}_\ell^2 < K < \infty$, we have:*

$$\sqrt{N}\frac{\overline{\boldsymbol{\eta}}'\mathbf{V}_i}{T} \overset{L_1}{\to} \mathbf{0}, \tag{49}$$

$$\sqrt{N}\frac{\overline{\boldsymbol{\eta}}'\mathbf{e}_i}{T} \overset{L_1}{\to} 0, \tag{50}$$

$$\sqrt{N}\frac{\overline{\boldsymbol{\eta}}'\mathbf{Q}}{T} \overset{L_1}{\to} \mathbf{0}, \tag{51}$$

$$\sqrt{N}\frac{\boldsymbol{\eta}_i'\overline{\mathbf{V}}}{T} \overset{L_1}{\to} \mathbf{0}, \tag{52}$$

$$\sqrt{N}\frac{\boldsymbol{\eta}_i'\overline{\mathbf{e}}}{T} \overset{L_1}{\to} 0, \tag{53}$$

$$N\frac{\overline{\boldsymbol{\eta}}'\overline{\mathbf{V}}}{T} \overset{L_1}{\to} \mathbf{0}, \tag{54}$$

$$N\frac{\overline{\boldsymbol{\eta}}'\overline{\mathbf{e}}}{T} \overset{L_1}{\to} 0, \tag{55}$$

and

$$N\frac{\overline{\boldsymbol{\eta}}'\overline{\boldsymbol{\eta}}}{T} - N\sum_{\ell=1}^{m_2} \overline{\lambda}_\ell^2 \overset{L_1}{\to} 0. \tag{56}$$

If in addition $N\sum_{\ell=1}^{m_2} \overline{\lambda}_\ell \lambda_{i\ell} < K < \infty$,

$$N\frac{\overline{\boldsymbol{\eta}}'\boldsymbol{\eta}_i}{T} - N\sum_{\ell=1}^{m_2} \overline{\lambda}_\ell \lambda_{i\ell} \overset{L_1}{\to} 0. \tag{57}$$

Proof. Let $T_N = T(N)$ and $m_{2,N} = m_2(N)$ be any non-decreasing integer-valued functions of N such that $\lim_{N\to\infty} T_N = \infty$, and $N\sum_{\ell}^{m_2} \overline{\lambda}_\ell^2 < K < \infty$.

(a) Consider now the following two-dimensional vector array $\left\{\{\boldsymbol{\kappa}_{Nt}, \mathcal{F}_t\}_{t=-\infty}^{\infty}\right\}_{N=1}^{\infty}$, defined by

$$\boldsymbol{\kappa}_{Nt} = \frac{\sqrt{N}}{T}\overline{\eta}_t \mathbf{v}_{it} = \frac{\sqrt{N}}{T_N}\sum_{\ell=1}^{m_{2,N}} \overline{\lambda}_\ell g_{\ell t}\mathbf{v}_{it}$$

where $\overline{\lambda}_\ell = \frac{1}{N}\sum_{j=1}^{N} \lambda_{j\ell}$ and $\{\mathcal{F}_t\}$ denotes an increasing sequence of σ-fields ($\mathcal{F}_{t-1} \subset \mathcal{F}_t$) such that $\mathcal{F}_t$ includes all information available at time t and $\boldsymbol{\kappa}_{Nt}$ is measurable with respect to $\mathcal{F}_t$ for any $N \in \mathbb{N}$. Set $c_{Nt} = \frac{1}{T_N}$ for all $t \in \mathbb{Z}$ and $N \in \mathbb{N}$. We have

$$E\left(\frac{\boldsymbol{\kappa}_{Nt}\boldsymbol{\kappa}_{Nt}'}{c_{Nt}^2}\right) = N \cdot \mathbf{\Sigma}_i \sum_{\ell=1}^{m_{2,N}} \overline{\lambda}_\ell^2,$$

where $E(\mathbf{v}_{it}\mathbf{v}_{it}') = \mathbf{\Sigma}_i$ and $E\left(g_{\ell t}^2\right) = 1$. It follows that

$$\left\|E\left(\frac{\boldsymbol{\kappa}_{Nt}\boldsymbol{\kappa}_{Nt}'}{c_{Nt}^2}\right)\right\| \leq \|\mathbf{\Sigma}_i\| N \sum_{\ell=1}^{m_{2,N}} \overline{\lambda}_\ell^2 < K < \infty. \tag{58}$$

Consider now

$$\left\| E\left\{ E\left(\frac{\boldsymbol{\kappa}_{Nt}}{c_{Nt}} \mid \mathcal{F}_{t-s} \right) E\left(\frac{\boldsymbol{\kappa}_{Nt}}{c_{Nt}} \mid \mathcal{F}_{t-s} \right)' \right\} \right\| \equiv \varsigma_s .$$

Equation (58) implies that $\varsigma_0 < K < \infty$ and by covariance stationarity of $\mathbf{v}_{it}$ and $g_{\ell t}$, we have $\varsigma_s \to 0$ as $s \to \infty$. By Liapunov's inequality, $E\left|E\left(\kappa_{Nt} \mid \mathcal{F}_{t-n}\right)\right| \leq \sqrt{E\left\{\left[E\left(\kappa_{Nt} \mid \mathcal{F}_{t-n}\right)\right]^2\right\}}$ (Davidson, 1994, Theorem 9.23) and the two-dimensional vector array $\left\{\{\boldsymbol{\kappa}_{Nt}, \mathcal{F}_t\}_{t=-\infty}^{\infty}\right\}_{N=1}^{\infty}$ is L_1-mixingale with respect to the constant array $\{c_{Nt}\}$. Equation (58) established that $\{\boldsymbol{\kappa}_{Nt}/c_{Nt}\}$ is uniformly bounded in L_2 norm, which implies uniform integrability.[13] Finally, note that the constant array $\{c_{Nt}\}$ satisfy the following conditions

$$\lim_{N\to\infty} \sum_{t=1}^{T_N} c_{Nt} = \lim_{N\to\infty} \sum_{t=1}^{T_N} \frac{1}{T_N} = 1 < \infty,$$

and

$$\lim_{N\to\infty} \sum_{t=1}^{T_N} c_{Nt}^2 = \lim_{N\to\infty} \sum_{t=1}^{T_N} \frac{1}{T_N^2} = 0.$$

It follows that array $\left\{\{\boldsymbol{\kappa}_{Nt}, \mathcal{F}_t\}_{t=-\infty}^{\infty}\right\}_{N=1}^{\infty}$ satisfies conditions of a mixingale weak law (Davidson, 1994, Theorem 19.11)., which implies $\sum_{t=1}^{T_N} \kappa_{Nt} \xrightarrow{L_1} 0$, that is

$$\sqrt{N}\frac{\overline{\boldsymbol{\eta}}'\mathbf{v}_i}{T} \xrightarrow{L_1} 0,$$

as $m_2, T, N \xrightarrow{j} \infty$, such that $N\sum_{\ell}^{m_2} \overline{\lambda}_{\ell}^2 < K < \infty$. This completes the proof of result (49). Results (50)-(51) can be proved in the same way.[14] Remaining results are proved below using the similar logical arguments.

(b) Next we establish result (52). Let

$$\boldsymbol{\kappa}_{Nt} = \frac{\sqrt{N}}{T_N}\eta_{it}\overline{\mathbf{v}}_t, \tag{59}$$

and as before consider the two-dimensional vector array $\left\{\{\boldsymbol{\kappa}_{Nt}, \mathcal{F}_t\}_{t=-\infty}^{\infty}\right\}_{N=1}^{\infty}$ defined by (59) and the same constant array c_{Nt}, namely $c_{Nt} = \frac{1}{T_N}$ for all $t \in \mathbb{Z}$ and $N \in \mathbb{N}$. We have

$$\left\| E\left(\frac{\boldsymbol{\kappa}_{Nt}\boldsymbol{\kappa}_{Nt}'}{c_{Nt}^2} \right) \right\| \leq N\left(\sum_{\ell=1}^{m_{2,N}} \lambda_{i\ell}^2 \right) \cdot \frac{1}{N^2}\sum_{i=1}^{N} \|\boldsymbol{\Sigma}_i\| < K < \infty.$$

Using similar arguments as before, $\left\{\{\boldsymbol{\kappa}_{Nt}, \mathcal{F}_t\}_{t=-\infty}^{\infty}\right\}_{N=1}^{\infty}$ is L_1-mixingale with respect to constant array $\{c_{Nt}\}$, and a mixingale weak law (Davidson, 1994, Theorem 19.11) imply $\sum_{t=1}^{T_N} \kappa_{Nt} \xrightarrow{L_1} 0$, that is

$$\sqrt{N}\frac{\boldsymbol{\eta}_i'\overline{\mathbf{v}}}{T} \xrightarrow{L_1} 0,$$

as $m_2, T, N \xrightarrow{j} \infty$, such that $N\sum_{\ell}^{m_2} \overline{\lambda}_{\ell}^2 < K < \infty$, which concludes the proof of result (52). Proof of result (53) is identical to the proof of result (52), but this time we set $\kappa_{Nt} = \frac{\sqrt{N}}{T_N}\eta_{it}\overline{e}_t$.

(c) Next we establish results (54) and (55) in a similar way. Define

$$\boldsymbol{\kappa}_{Nt} = \frac{N}{T_N}\overline{\eta}_t\overline{\mathbf{v}}_t.$$

As before, set $c_{Nt} = \frac{1}{T_N}$ for all $t \in \mathbb{Z}$ and $N \in \mathbb{N}$. Examining variance of $\boldsymbol{\kappa}_{Nt}\boldsymbol{\kappa}_{Nt}'/c_{Nt}^2$ yields

$$\left\| E\left(\frac{\boldsymbol{\kappa}_{Nt}\boldsymbol{\kappa}_{Nt}'}{c_{Nt}^2} \right) \right\| \leq \left(N\sum_{\ell=1}^{m_{2,N}} \overline{\lambda}_{\ell}^2 \right) \frac{1}{N}\sum_{i=1}^{N} \|\boldsymbol{\Sigma}_i\| < K < \infty.$$

[13] Sufficient condition for uniform integrability is $L_{1+\varepsilon}$ uniform boundedness for any $\varepsilon > 0$.

[14] Define $\kappa_{Nt} = \frac{\sqrt{N}}{T}\overline{\eta}_t e_{it}$ to prove result (50) and $\boldsymbol{\kappa}_{Nt} = \frac{\sqrt{N}}{T}\overline{\eta}_t\mathbf{f}_t$ to prove result (51)

Using the same arguments as before, $\left\{\{\boldsymbol{\kappa}_{Nt},\mathcal{F}_t\}_{t=-\infty}^{\infty}\right\}_{N=1}^{\infty}$ is L_1-mixingale with respect to constant array $\{c_{Nt}\}$, and a mixingale weak law (Davidson, 1994, Theorem 19.11) establishes result (54). Result (55) easily follows by noting that $Var\left(\overline{e}_t\right)$ and $\|Var\left(\overline{\mathbf{v}}_t\right)\|$ are both of order $O\left(N^{-1}\right)$.

(d) Next we prove equation (56). Set

$$\kappa_{Nt}=\frac{1}{T_N}N\overline{\eta}_t^2-\frac{1}{T_N}E\left(N\overline{\eta}_t^2\right),$$

and $c_{Nt}=\frac{1}{T_N}$ for all $t\in\mathbb{Z}$ and $N\in\mathbb{N}$. Note that

$$E\left(N\overline{\eta}_t^2\right)=N\sum_{\ell=1}^{m_{2,N}}\overline{\lambda}_\ell^2<K<\infty.$$

Furthermore,

$$\begin{aligned}E\left[\left(\frac{\kappa_{Nt}}{c_{Nt}}\right)^2\right] &= N^2\sum_{\ell=1}^{m_{2,N}}\sum_{s=1}^{m_{2,N}}\sum_{h=1}^{m_{2,N}}\sum_{r=1}^{m_{2,N}}\overline{\lambda}_\ell\overline{\lambda}_s\overline{\lambda}_h\overline{\lambda}_r E\left(g_{\ell t}g_{st}g_{ht}g_{rt}\right)-\left[E\left(N\overline{\eta}_t^2\right)\right]^2\\ &\leq N^2\sum_{\ell=1}^{m_{2,N}}\overline{\lambda}_\ell^4E\left(g_{\ell t}^4\right)+4N^2\sum_{\ell=1}^{m_{2,N}}\overline{\lambda}_\ell^2\sum_{s=1}^{m_{2,N}}\overline{\lambda}_s^2<K<\infty.\end{aligned}$$

Using the same arguments as before, $\left\{\{\boldsymbol{\kappa}_{Nt},\mathcal{F}_t\}_{t=-\infty}^{\infty}\right\}_{N=1}^{\infty}$ is L_1-mixingale with respect to constant array $\{c_{Nt}\}$, and a mixingale weak law (Davidson, 1994, Theorem 19.11) imply$\sum_{t=1}^{T_N}\kappa_{Nt}\overset{L_1}{\to}0$, namely $N\frac{\overline{\boldsymbol{\eta}}'\overline{\boldsymbol{\eta}}}{T}-N\sum_{\ell}^{m_2}\overline{\lambda}_\ell^2\overset{L_1}{\to}0$, as $m_2,T,N\overset{j}{\to}\infty$, such that $N\sum_{\ell}^{m_2}\overline{\lambda}_\ell^2<K<\infty$. This completes the proof of result (56).

(e) To establish result (57), define

$$\kappa_{Nt}=\frac{1}{T_N}N\overline{\eta}_t\eta_{it}-\frac{1}{T_N}E\left(N\overline{\eta}_t\eta_{it}\right),$$

and set again $c_{Nt}=\frac{1}{T_N}$ for all $t\in\mathbb{Z}$ and $N\in\mathbb{N}$. We have

$$E\left(N\overline{\eta}_t\eta_{it}\right)=N\sum_{\ell=1}^{m_{2,N}}\overline{\lambda}_\ell\lambda_{i\ell}<K<\infty,$$

and

$$\begin{aligned}E\left[\left(\frac{\kappa_{Nt}}{c_{Nt}}\right)^2\right] &= N^2\sum_{\ell=1}^{m_{2,N}}\sum_{s=1}^{m_{2,N}}\sum_{h=1}^{m_{2,N}}\sum_{r=1}^{m_{2,N}}\overline{\lambda}_\ell\lambda_{is}\overline{\lambda}_h\lambda_{ir}E\left(g_{\ell t}g_{st}g_{ht}g_{rt}\right)-\left[E\left(N\overline{\eta}_t\eta_{it}\right)\right]^2\\ &\leq N^2\sum_{\ell=1}^{m_{2,N}}\overline{\lambda}_\ell^2\lambda_{i\ell}^2E\left(g_{\ell t}^4\right)+3N^2\left(\sum_{\ell=1}^{m_{2,N}}\overline{\lambda}_\ell\lambda_{i\ell}\right)^2+N^2\sum_{\ell=1}^{m_{2,N}}\overline{\lambda}_\ell^2\lambda_{i\ell}^2<K<\infty.\end{aligned}$$

Using the same arguments as before, $\left\{\{\boldsymbol{\kappa}_{Nt},\mathcal{F}_t\}_{t=-\infty}^{\infty}\right\}_{N=1}^{\infty}$ is L_1-mixingale with respect to constant array $\{c_{Nt}\}$, and a mixingale weak law (Davidson, 1994, Theorem 19.11) imply result (57).

■

The following lemma collects several results presented in Pesaran (2006), Kapetanios Pesaran and Yamagata (2009), and Pesaran and Tosetti (2009).

Lemma 2 *Consider the panel data model (37) and (39) and suppose that Assumptions 8-12 hold, and m_1 does not vary with N. Then as $T,N\overset{j}{\to}\infty$ (at any rate) we have:*

$$\sqrt{N}\frac{\mathbf{Q}'\mathbf{V}_i}{T}\overset{L_1}{\to}\mathbf{0},\ \sqrt{N}\frac{\mathbf{Q}'\mathbf{e}_i}{T}\overset{L_1}{\to}\mathbf{0},\tag{60}$$

$$\sqrt{N}\frac{\mathbf{Q}'\overline{\mathbf{e}}}{T}\overset{L_1}{\to}\mathbf{0},\ \sqrt{N}\frac{\mathbf{Q}'\overline{\mathbf{V}}}{T}\overset{L_1}{\to}\mathbf{0},\tag{61}$$

$$\sqrt{N}\frac{\mathbf{V}_i'\overline{\mathbf{V}}}{T} \xrightarrow{L_1} \mathbf{0},\ \sqrt{N}\frac{\mathbf{V}_i'\overline{\mathbf{e}}}{T} \xrightarrow{L_1} \mathbf{0}, \tag{62}$$

$$\sqrt{N}\frac{\overline{\mathbf{e}}'\overline{\mathbf{e}}}{T} \xrightarrow{L_1} 0,\ \sqrt{N}\frac{\overline{\mathbf{V}}'\overline{\mathbf{V}}}{T} \xrightarrow{L_1} \mathbf{0},\ \sqrt{N}\frac{\overline{\mathbf{e}}'\overline{\mathbf{V}}}{T} \xrightarrow{L_1} \mathbf{0}, \tag{63}$$

and

$$\sqrt{N}\frac{\mathbf{e}_i'\overline{\mathbf{e}}}{T} \xrightarrow{L_1} 0,\ \sqrt{N}\frac{\mathbf{e}_i'\overline{\mathbf{V}}}{T} \xrightarrow{L_1} 0. \tag{64}$$

Proof. Lemma 2 follows directly from Pesaran (2006), Kapetanios Pesaran and Yamagata (2009), and Pesaran and Tosetti (2009). These results can also be established in the same way as Lemma 1 by using a mixingale weak law. ■

Lemma 3 *Consider the panel data model (37) and (39) and suppose that Assumptions 8-12 hold, and m_1 does not vary with N. Then as $m_2, T, N \xrightarrow{j} \infty$, such that $N\sum_{\ell}^{m_2} \overline{\lambda}_\ell \lambda_i < K < \infty$, we have:*

$$\sqrt{N}\frac{\mathbf{X}_i'\overline{\mathbf{M}}\boldsymbol{\eta}_i}{T} - \sqrt{N}\frac{\mathbf{X}_i'\mathbf{M}_g\boldsymbol{\eta}_i}{T} \xrightarrow{L_1} 0 \tag{65}$$

$$\sqrt{N}\frac{\mathbf{X}_i'\overline{\mathbf{M}}\mathbf{F}}{T} \xrightarrow{L_1} 0 \tag{66}$$

$$\sqrt{N}\frac{\mathbf{X}_i'\overline{\mathbf{M}}\mathbf{X}_i}{T} - \sqrt{N}\frac{\mathbf{X}_i'\mathbf{M}_g\mathbf{X}_i}{T} \xrightarrow{L_1} 0 \tag{67}$$

$$\sqrt{N}\frac{\mathbf{X}_i'\overline{\mathbf{M}}\mathbf{e}_i}{T} - \sqrt{N}\frac{\mathbf{X}_i'\mathbf{M}_g\mathbf{e}_i}{T} \xrightarrow{L_1} 0. \tag{68}$$

Proof. Throughout this proof we consider asymptotics $m_2, T, N \xrightarrow{j} \infty$, such that $N\sum_{\ell}^{m_2} \overline{\lambda}_\ell \lambda_i < K < \infty$. We start by establishing result (65). Consider

$$\begin{aligned}
\frac{\sqrt{N}}{T}\left\|\mathbf{X}_i'\overline{\mathbf{M}}\boldsymbol{\eta}_i - \mathbf{X}_i'\mathbf{M}_g\boldsymbol{\eta}_i\right\| &= \frac{\sqrt{N}}{T}\left\|\mathbf{X}_i'\overline{\mathbf{H}}\left(\overline{\mathbf{H}}'\overline{\mathbf{H}}\right)^{-1}\overline{\mathbf{H}}'\boldsymbol{\eta}_i - \mathbf{X}_i'\mathbf{Q}\left(\mathbf{Q}'\mathbf{Q}\right)^{-1}\mathbf{Q}'\boldsymbol{\eta}_i\right\| \\
&\leq \left\|\frac{\sqrt{N}\mathbf{X}_i'\left(\overline{\mathbf{H}} - \mathbf{Q}\right)}{T}\left(\frac{\overline{\mathbf{H}}'\overline{\mathbf{H}}}{T}\right)^{-1}\frac{\overline{\mathbf{H}}'\boldsymbol{\eta}_i}{T}\right\| + \\
&\quad + \left\|\frac{\mathbf{X}_i'\mathbf{Q}}{T}\left(\frac{\mathbf{Q}'\mathbf{Q}}{T}\right)^{-1}\frac{\sqrt{N}\left(\mathbf{Q}' - \overline{\mathbf{H}}'\right)\boldsymbol{\eta}_i}{T}\right\| + \\
&\quad + \left\|\frac{\mathbf{X}_i'\mathbf{Q}}{T}\frac{\sqrt{N}}{T}\left(\left(\overline{\mathbf{H}}'\overline{\mathbf{H}}\right)^{-1} - \left(\mathbf{Q}'\mathbf{Q}\right)^{-1}\right)\frac{\overline{\mathbf{H}}'\boldsymbol{\eta}_i}{T}\right\|.
\end{aligned} \tag{69}$$

We examine each of the three terms below. We have

$$\frac{\sqrt{N}\mathbf{X}_i'\left(\overline{\mathbf{H}} - \mathbf{Q}\right)}{T} = \frac{\sqrt{N}\left(\mathbf{G}\mathbf{\Pi}_i + \mathbf{V}_i\right)'\overline{\mathbf{U}}^*}{T}.$$

Equation (49) of Lemma 1 and equation (62) of Lemma 2 establish

$$\sqrt{N}\frac{\mathbf{V}_i'\overline{\mathbf{U}}}{T} \xrightarrow{L_1} 0. \tag{70}$$

In addition, equation (51) of Lemma 1 and equation (61) of Lemma 2 establish

$$\sqrt{N}\frac{\mathbf{G}'\overline{\mathbf{U}}}{T} \xrightarrow{L_1} 0. \tag{71}$$

Equations (70) and (71) imply

$$\frac{\sqrt{N}\mathbf{X}_i'\left(\overline{\mathbf{H}} - \mathbf{Q}\right)}{T} \xrightarrow{L_1} 0, \tag{72}$$

and noting that $\left(\frac{\overline{\mathbf{H}}'\overline{\mathbf{H}}}{T}\right)^{-1} = O_p(1)$, and $\frac{\overline{\mathbf{H}}'\boldsymbol{\eta}_i}{T} = O_p(1)$, establish

$$\left\| \frac{\sqrt{N}\mathbf{X}_i'\left(\overline{\mathbf{H}} - \mathbf{Q}\right)}{T}\left(\frac{\overline{\mathbf{H}}'\overline{\mathbf{H}}}{T}\right)^{-1}\frac{\overline{\mathbf{H}}'\boldsymbol{\eta}_i}{T}\right\|_{L_1} \to 0.$$

Now we focus on the second term of (69). Equations (52),(53) and (57) of Lemma 1 imply

$$\frac{\sqrt{N}\overline{\mathbf{U}}'\boldsymbol{\eta}_i}{T} \overset{L_1}{\to} 0.$$

It follows that $\frac{\sqrt{N}\left(\mathbf{Q}'-\overline{\mathbf{H}}'\right)\boldsymbol{\eta}_i}{T} \overset{L_1}{\to} 0$ and since $\frac{\mathbf{X}_i'\mathbf{Q}}{T} = O_p(1)$ and $\left(\frac{\mathbf{Q}'\mathbf{Q}}{T}\right)^{-1} = O_p(1)$, we have

$$\left\| \frac{\mathbf{X}_i'\mathbf{Q}}{T}\left(\frac{\mathbf{Q}'\mathbf{Q}}{T}\right)^{-1}\frac{\sqrt{N}\left(\mathbf{Q}' - \overline{\mathbf{H}}'\right)\boldsymbol{\eta}_i}{T}\right\|_{L_1} \to 0.$$

In order to establish the last term of (69), we write

$$\frac{\sqrt{N}}{T}\left(\left(\overline{\mathbf{H}}'\overline{\mathbf{H}}\right)^{-1} - \left(\mathbf{Q}'\mathbf{Q}\right)^{-1}\right) = \left(\frac{\overline{\mathbf{H}}'\overline{\mathbf{H}}}{T}\right)^{-1}\frac{\sqrt{N}\left(\mathbf{Q}\mathbf{Q} - \overline{\mathbf{H}}'\overline{\mathbf{H}}\right)}{T}\left(\frac{\mathbf{Q}'\mathbf{Q}}{T}\right)^{-1},$$

where (note that $\overline{\mathbf{H}} = \mathbf{Q} + \overline{\mathbf{U}}^*$)

$$\frac{\sqrt{N}\left(\mathbf{Q}\mathbf{Q} - \overline{\mathbf{H}}'\overline{\mathbf{H}}\right)}{T} = -\frac{\sqrt{N}\overline{\mathbf{U}}^{*\prime}\overline{\mathbf{U}}^*}{T} - \frac{\sqrt{N}\mathbf{Q}'\overline{\mathbf{U}}^*}{T} - \frac{\sqrt{N}\overline{\mathbf{U}}^{*\prime}\mathbf{Q}}{T}.$$

Equations (54),(55) and (56) of Lemma 1 and equation (63) of Lemma 2 imply

$$\frac{\sqrt{N}\overline{\mathbf{U}}^{*\prime}\overline{\mathbf{U}}^*}{T} \overset{L_1}{\to} 0.$$

Similarly, equation (51) of Lemma 1 and equation (61) of Lemma 2 imply

$$\frac{\sqrt{N}\mathbf{Q}'\overline{\mathbf{U}}^*}{T} \overset{L_1}{\to} 0, \text{ as well as } \frac{\sqrt{N}\overline{\mathbf{U}}^{*\prime}\mathbf{Q}}{T} \overset{L_1}{\to} 0. \tag{73}$$

Noting that $\frac{\mathbf{X}_i'\mathbf{Q}}{T} = O_p(1)$, $\frac{\overline{\mathbf{H}}'\boldsymbol{\eta}_i}{T} = O_p(1)$, $\left(\frac{\overline{\mathbf{H}}'\overline{\mathbf{H}}}{T}\right)^{-1} = O_p(1)$ and $\left(\frac{\mathbf{Q}'\mathbf{Q}}{T}\right)^{-1} = O_p(1)$, it follows that

$$\frac{\sqrt{N}}{T}\left(\left(\overline{\mathbf{H}}'\overline{\mathbf{H}}\right)^{-1} - \left(\mathbf{Q}'\mathbf{Q}\right)^{-1}\right) \overset{L_1}{\to} 0, \tag{74}$$

and therefore

$$\left\| \frac{\mathbf{X}_i'\mathbf{Q}}{T}\frac{\sqrt{N}}{T}\left(\left(\overline{\mathbf{H}}'\overline{\mathbf{H}}\right)^{-1} - \left(\mathbf{Q}'\mathbf{Q}\right)^{-1}\right)\frac{\overline{\mathbf{H}}'\boldsymbol{\eta}_i}{T}\right\|_{L_1} \to 0.$$

This completes the proof of result (65).

In order to establish result (66), note that $\mathbf{M}_g\mathbf{F} = \mathbf{0}$ and therefore (66) is equivalent with the following statement,

$$\sqrt{N}\frac{\mathbf{X}_i'\overline{\mathbf{M}}\mathbf{F}}{T} - \sqrt{N}\frac{\mathbf{X}_i'\mathbf{M}_g\mathbf{F}}{T} \overset{L_1}{\to} 0.$$

Using similar steps as in deriving equation (69), we have:

$$\begin{aligned}
\frac{\sqrt{N}}{T}\left\|\mathbf{X}_i'\overline{\mathbf{M}}\mathbf{F}-\mathbf{X}_i'\mathbf{M}_g\mathbf{F}\right\|_{L_1} \quad \leq \quad & \left\|\frac{\sqrt{N}\mathbf{X}_i'\left(\overline{\mathbf{H}}-\mathbf{Q}\right)}{T}\left(\frac{\overline{\mathbf{H}}'\overline{\mathbf{H}}}{T}\right)^{-1}\frac{\overline{\mathbf{H}}'\mathbf{F}}{T}\right\|_{L_1}+ \\
& +\left\|\frac{\mathbf{X}_i'\mathbf{Q}}{T}\left(\frac{\mathbf{Q}'\mathbf{Q}}{T}\right)^{-1}\frac{\sqrt{N}\left(\mathbf{Q}'-\overline{\mathbf{H}}'\right)\mathbf{F}}{T}\right\|_{L_1}+ \\
& +\left\|\frac{\mathbf{X}_i'\mathbf{Q}}{T}\frac{\sqrt{N}}{T}\left(\left(\overline{\mathbf{H}}'\overline{\mathbf{H}}\right)^{-1}-\left(\mathbf{Q}'\mathbf{Q}\right)^{-1}\right)\frac{\overline{\mathbf{H}}'\mathbf{F}}{T}\right\|_{L_1}. \qquad (75)
\end{aligned}$$

Since $\frac{\overline{\mathbf{H}}'\mathbf{F}}{T}=O_p(1)$, convergence of the first and the last term of (75) to zero directly follows from earlier results, in particular equations (72) and (74). Furthermore, equation (73) implies

$$\frac{\sqrt{N}\left(\mathbf{Q}'-\overline{\mathbf{H}}'\right)\mathbf{F}}{T}=\frac{\sqrt{N}\overline{\mathbf{U}}^{*\prime}\mathbf{F}}{T}\xrightarrow{L_1}0,$$

and it follows that also the second term of (75) converges to zero. This establishes that $\frac{\sqrt{N}}{T}\left\|\mathbf{X}_i'\overline{\mathbf{M}}\mathbf{F}-\mathbf{X}_i'\mathbf{M}_g\mathbf{F}\right\|_{L_1}\to 0$, which completes the proof of result (66).

Result (67) is established next in a similar fashion. Consider

$$\begin{aligned}
\frac{\sqrt{N}}{T}\left\|\mathbf{X}_i'\overline{\mathbf{M}}\mathbf{X}_i-\mathbf{X}_i'\mathbf{M}_g\mathbf{X}_i\right\|_{L_1} \quad \leq \quad & \left\|\frac{\sqrt{N}\mathbf{X}_i'\left(\overline{\mathbf{H}}-\mathbf{Q}\right)}{T}\right\|_{L_1}\left\|\left(\frac{\overline{\mathbf{H}}'\overline{\mathbf{H}}}{T}\right)^{-1}\right\|_{L_1}\left\|\frac{\overline{\mathbf{H}}'\mathbf{X}_i}{T}\right\|_{L_1}+ \\
& +\left\|\frac{\mathbf{X}_i'\mathbf{Q}}{T}\right\|_{L_1}\left\|\left(\frac{\mathbf{Q}'\mathbf{Q}}{T}\right)^{-1}\right\|_{L_1}\left\|\frac{\sqrt{N}\left(\mathbf{Q}'-\overline{\mathbf{H}}'\right)\mathbf{X}_i}{T}\right\|_{L_1}+ \\
& +\left\|\frac{\mathbf{X}_i'\mathbf{Q}}{T}\right\|_{L_1}\left\|\frac{\sqrt{N}}{T}\left(\left(\overline{\mathbf{H}}'\overline{\mathbf{H}}\right)^{-1}-\left(\mathbf{Q}'\mathbf{Q}\right)^{-1}\right)\right\|_{L_1}\left\|\frac{\overline{\mathbf{H}}'\mathbf{X}_i}{T}\right\|_{L_1}. \qquad (76)
\end{aligned}$$

Using equations (72) and (74), and noting that the remaining elements are bounded, we have

$$\frac{\sqrt{N}}{T}\left\|\mathbf{X}_i'\overline{\mathbf{M}}\mathbf{X}_i-\mathbf{X}_i'\mathbf{M}_g\mathbf{X}_i\right\|_{L_1}\to 0,$$

which completes the proof of result (67).

Finally, consider

$$\begin{aligned}
\frac{\sqrt{N}}{T}\left\|\mathbf{X}_i'\overline{\mathbf{M}}\mathbf{e}_i-\mathbf{X}_i'\mathbf{M}_g\mathbf{e}_i\right\|_{L_1} \quad \leq \quad & \left\|\frac{\sqrt{N}\mathbf{X}_i'\left(\overline{\mathbf{H}}-\mathbf{Q}\right)}{T}\right\|_{L_1}\left\|\left(\frac{\overline{\mathbf{H}}'\overline{\mathbf{H}}}{T}\right)^{-1}\right\|_{L_1}\left\|\frac{\overline{\mathbf{H}}'\mathbf{e}_i}{T}\right\|_{L_1}+ \\
& +\left\|\frac{\mathbf{X}_i'\mathbf{Q}}{T}\right\|_{L_1}\left\|\left(\frac{\mathbf{Q}'\mathbf{Q}}{T}\right)^{-1}\right\|_{L_1}\left\|\frac{\sqrt{N}\left(\mathbf{Q}'-\overline{\mathbf{H}}'\right)\mathbf{e}_i}{T}\right\|_{L_1}+ \\
& +\left\|\frac{\mathbf{X}_i'\mathbf{Q}}{T}\right\|_{L_1}\left\|\frac{\sqrt{N}}{T}\left(\left(\overline{\mathbf{H}}'\overline{\mathbf{H}}\right)^{-1}-\left(\mathbf{Q}'\mathbf{Q}\right)^{-1}\right)\right\|_{L_1}\left\|\frac{\overline{\mathbf{H}}'\mathbf{e}_i}{T}\right\|_{L_1}. \qquad (77)
\end{aligned}$$

Equation (50) of Lemma 1 and equation (64) of Lemma 2 imply

$$\left\|\frac{\sqrt{N}\left(\mathbf{Q}'-\overline{\mathbf{H}}'\right)\mathbf{e}_i}{T}\right\|_{L_1}=\left\|\frac{\sqrt{N}\overline{\mathbf{U}}^{*\prime}\mathbf{e}_i}{T}\right\|_{L_1}\to 0. \qquad (78)$$

Equations (72), (74) and (78) imply $\frac{\sqrt{N}}{T}\left\|\mathbf{X}_i'\overline{\mathbf{M}}\mathbf{e}_i-\mathbf{X}_i'\mathbf{M}_g\mathbf{e}_i\right\|_{L_1}\rightarrow 0$, which completes the proof of result (68). ■

Proof of Theorem 2. We prove Theorem 2 in two parts. First, we establish result (46) for the CCE pooled estimation and in the second part we establish result (45) for the CCE mean group estimation.

Let $\eta_{it}=\sum_{\ell=1}^{m_2}\lambda_{i\ell}g_{\ell t}$ and consider

$$\sqrt{N}\left(\widehat{\boldsymbol{\beta}}_{CCEP}-\boldsymbol{\beta}\right)=\left(\frac{1}{N}\sum_{i=1}^{N}\frac{\mathbf{X}_i'\overline{\mathbf{M}}\mathbf{X}_i}{T}\right)^{-1}\sum_{i=1}^{N}\frac{1}{\sqrt{N}}\frac{\mathbf{X}_i'\overline{\mathbf{M}}\left(\mathbf{X}_i\boldsymbol{v}_i+\mathbf{F}\gamma_i+\boldsymbol{\eta}_i+\mathbf{e}_i\right)}{T}, \tag{79}$$

We focus first on the new term $\sum_{i=1}^{N}\frac{1}{T\sqrt{N}}\mathbf{X}_i'\overline{\mathbf{M}}\boldsymbol{\eta}_i$, which is introduced by possibly infinite factor structure $\{g_{\ell t}\}_{\ell=1}^{m_2}$ and which is not present in previous studies by Pesaran and Tosetti (2009), Kapetanios Pesaran and Yamagata (2009), and Pesaran (2006). Equation (65) of Lemma 3 implies

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{\mathbf{X}_i'\overline{\mathbf{M}}\boldsymbol{\eta}_i}{T}-\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{\mathbf{V}_i'\mathbf{M}_g\boldsymbol{\eta}_i}{T}\overset{L_1}{\rightarrow}0, \tag{80}$$

as $m_2,T,N\overset{j}{\rightarrow}\infty$, such that $N\sum_{\ell}^{m_2}\overline{\lambda}_\ell\lambda_i<K<\infty$. Let $\widetilde{\mathbf{V}}_i=\mathbf{M}_g\mathbf{V}_i$ and denote t-th row of matrix $\widetilde{\mathbf{V}}_i$ as $\widetilde{\mathbf{v}}_{it}'$. Using this notation we write,

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{\mathbf{V}_i'\mathbf{M}_g\boldsymbol{\eta}_i}{T}=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{\widetilde{\mathbf{V}}_i\boldsymbol{\eta}_i}{T}=\sum_{t-1}^{T}\left(\frac{1}{T\sqrt{N}}\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{\mathbf{v}}_{it}\eta_{it}\right).$$

Let $T_N=T(N)$ and $m_{2,N}=m_2(N)$ be any non-decreasing integer-valued functions of N such that $\lim_{N\rightarrow\infty}T_N=\infty$ and such that Assumption 9.b holds, namely $\lim_{N\rightarrow\infty}\sum_{\ell=1}^{m_{2,N}}\lambda_{i\ell}^2<K$. Consider now the following two-dimensional vector array $\left\{\{\boldsymbol{\kappa}_{Nt},\mathcal{F}_t\}_{t=-\infty}^{\infty}\right\}_{N=1}^{\infty}$, defined by

$$\boldsymbol{\kappa}_{Nt}=\frac{1}{T_N\sqrt{N}}\sum_{i=1}^{N}\widetilde{\mathbf{v}}_{it}\eta_{it}=\frac{1}{T_N\sqrt{N}}\sum_{i=1}^{N}\widetilde{\mathbf{v}}_{it}\sum_{\ell=1}^{m_{2,N}}\lambda_{i\ell}g_{\ell t}, \tag{81}$$

and $\{\mathcal{F}_t\}$ denotes an increasing sequence of σ-fields ($\mathcal{F}_{t-1}\subset\mathcal{F}_t$) such that $\mathcal{F}_t$ includes all information available at time t and $\boldsymbol{\kappa}_{Nt}$ is measurable with respect to $\mathcal{F}_t$ for any $N\in\mathbb{N}$. Let $\left\{\{c_{Nt}\}_{t=-\infty}^{\infty}\right\}_{N=1}^{\infty}$ be two-dimensional array of constants and set $c_{Nt}=\frac{1}{T_N}$ for all $t\in\mathbb{Z}$ and $N\in\mathbb{N}$. Note that

$$\begin{aligned}E\left(\frac{\boldsymbol{\kappa}_{Nt}\boldsymbol{\kappa}_{Nt}'}{c_{Nt}^2}\right)&=\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}E\left(\widetilde{\mathbf{v}}_{it}\widetilde{\mathbf{v}}_{jt}'\eta_{it}\eta_{jt}\right)\\&=\frac{1}{N}\sum_{i=1}^{N}E\left(\widetilde{\mathbf{v}}_{it}\widetilde{\mathbf{v}}_{it}'\right)E\left(\eta_{it}^2\right),\end{aligned}$$

where the second equality follow from independence of $\mathbf{v}_{it}$ and $\mathbf{v}_{jt}$ for any $i\neq j$. $E\left(\widetilde{\mathbf{v}}_{it}\widetilde{\mathbf{v}}_{it}'\right)=\boldsymbol{\Sigma}_i$ and by Assumption 10 there exists a constant $K_1<\infty$, which does not depend on i nor on N and such that $\|\boldsymbol{\Sigma}_i\|<K_1$. Further, using independence of factors $g_{\ell t}$ and $g_{\ell' t}$ for any $\ell\neq\ell'$ and noting that $E\left(g_{\ell t}^2\right)=1$, we have

$$E\left(\eta_{it}^2\right)=\sum_{\ell=1}^{m_{2,N}}\lambda_{i\ell}^2<K_2<\infty,$$

where the existence of uniform upper bound K_2, which does not depend on i,N is assumed in Assumption 9. It follows that

$$\left\|E\left(\frac{\boldsymbol{\kappa}_{Nt}\boldsymbol{\kappa}_{Nt}'}{c_{Nt}^2}\right)\right\|<K<\infty, \tag{82}$$

where the constant $K = K_1 K_2$ and it does not depend on N. Consider now

$$\left\| E\left\{ E\left(\frac{\boldsymbol{\kappa}_{Nt}}{c_{Nt}} \mid \mathcal{F}_{t-s} \right) E\left(\frac{\boldsymbol{\kappa}_{Nt}}{c_{Nt}} \mid \mathcal{F}_{t-s} \right)' \right\} \right\| = \varsigma_s .$$

Equation (82) implies that $\varsigma_0 < K < \infty$ and by stationarity of $\mathbf{v}_{it}$ and $g_{\ell t}$, we have $\varsigma_s \to 0$ as $s \to \infty$. By Liapunov's inequality, $E\left|E\left(\kappa_{Nt} \mid \mathcal{F}_{t-n}\right)\right| \leq \sqrt{E\left\{\left[E\left(\kappa_{Nt} \mid \mathcal{F}_{t-n}\right)\right]^2\right\}}$ (Davidson, 1994, Theorem 9.23) and the two-dimensional array $\left\{\{\kappa_{Nt}, \mathcal{F}_t\}_{t=-\infty}^{\infty}\right\}_{N=1}^{\infty}$ is L_1-mixingale with respect to the constant array $\{c_{Nt}\}$. Equation (82) established that $\{\kappa_{Nt}/c_{Nt}\}$ is uniformly bounded in L_2 norm, which implies uniform integrability.[15] Finally, note that the constant array $\{c_{Nt}\}$ satisfies the following conditions

$$\lim_{N\to\infty} \sum_{t=1}^{T_N} c_{Nt} = \lim_{N\to\infty} \sum_{t=1}^{T_N} \frac{1}{T_N} = 1 < \infty,$$

$$\lim_{N\to\infty} \sum_{t=1}^{T_N} c_{Nt}^2 = \lim_{N\to\infty} \sum_{t=1}^{T_N} \frac{1}{T_N^2} = 0.$$

It follows that array $\left\{\{\kappa_{Nt}, \mathcal{F}_t\}_{t=-\infty}^{\infty}\right\}_{N=1}^{\infty}$ satisfies conditions of a mixingale weak law (Davidson, 1994, Theorem 19.11)., which establish $\sum_{t=1}^{T_N} \kappa_{Nt} \xrightarrow{L_1} 0$, that is:

$$\frac{1}{T\sqrt{N}} \sum_{i=1}^{N} \sum_{t=1}^{T} \widetilde{\mathbf{v}}_{it} \eta_{it} \xrightarrow{L_1} 0, \tag{83}$$

as $m_2, N, T \xrightarrow{j} \infty$ (at any rate) or m_2 is fixed and $N, T \xrightarrow{j} \infty$. Equations (80) and (83) imply

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{\mathbf{X}_i' \overline{\mathbf{M}} \boldsymbol{\eta}_i}{T} \xrightarrow{L_1} 0,$$

as $m_2, T, N \xrightarrow{j} \infty$, such that $N \sum_{\ell}^{m_2} \overline{\lambda}_\ell \lambda_i < K < \infty$. Convergence results for the remaining terms on the right side of equation (79) can be established in the same way as in Pesaran and Tosetti (2009) or Pesaran (2006). In particular, results (66)-(68) of Lemma 3 imply

$$\left(\frac{1}{N} \sum_{i=1}^{N} \frac{\mathbf{X}_i' \overline{\mathbf{M}} \mathbf{X}_i}{T} \right)^{-1} \sum_{i=1}^{N} \frac{1}{\sqrt{N}} \frac{\mathbf{X}_i' \overline{\mathbf{M}} \mathbf{X}_i \boldsymbol{\upsilon}_i}{T} - \left(\frac{1}{N} \sum_{i=1}^{N} \frac{\mathbf{X}_i' \mathbf{M}_g \mathbf{X}_i}{T} \right)^{-1} \sum_{i=1}^{N} \frac{1}{\sqrt{N}} \frac{\mathbf{X}_i' \mathbf{M}_g \mathbf{X}_i \boldsymbol{\upsilon}_i}{T} \xrightarrow{L_1} 0,$$

and

$$\left(\frac{1}{N} \sum_{i=1}^{N} \frac{\mathbf{X}_i' \overline{\mathbf{M}} \mathbf{X}_i}{T} \right)^{-1} \sum_{i=1}^{N} \frac{1}{\sqrt{N}} \frac{\mathbf{X}_i' \overline{\mathbf{M}} \left(\mathbf{F} \gamma_i + \mathbf{e}_i \right)}{T} \xrightarrow{L_1} 0,$$

as $m_2, T, N \xrightarrow{j} \infty$, such that $N \sum_{\ell}^{m_2} \overline{\lambda}_\ell \lambda_i < K < \infty$. It follows that

$$\sqrt{N} \left(\widehat{\boldsymbol{\beta}}_{CCEP} - \boldsymbol{\beta} \right) \xrightarrow{D} N\left(0, \boldsymbol{\Sigma}_{CCEP}\right),$$

as $m_2, T, N \xrightarrow{j} \infty$, such that $N \sum_{\ell}^{m_2} \overline{\lambda}_\ell \lambda_i < K < \infty$. This completes the proof of result (46).

Next we establish result (45) for the CCE mean group estimation. Let again $\eta_{it} = \sum_{\ell=1}^{m_2} \lambda_{i\ell} g_{\ell t}$ and consider

$$\begin{aligned} \sqrt{N} \left(\widehat{\boldsymbol{\beta}}_{CCEMG} - \boldsymbol{\beta} \right) &= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \boldsymbol{\upsilon}_i + \frac{1}{N} \sum_{i=1}^{N} \widehat{\boldsymbol{\Psi}}_{iT}^{-1} \left(\frac{\sqrt{N} \mathbf{X}_i' \overline{\mathbf{M}} \mathbf{F}}{T} \right) \gamma_i + \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \widehat{\boldsymbol{\Psi}}_{iT}^{-1} \left(\frac{\mathbf{X}_i' \overline{\mathbf{M}} \boldsymbol{\eta}_i}{T} \right) + \\ &\quad + \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \widehat{\boldsymbol{\Psi}}_{iT}^{-1} \left(\frac{\mathbf{X}_i' \overline{\mathbf{M}} \mathbf{e}_i}{T} \right), \end{aligned} \tag{84}$$

[15] Sufficient condition for uniform integrability is $L_{1+\varepsilon}$ uniform boundedness for any $\varepsilon > 0$.

where $\widehat{\mathbf{\Psi}}_{iT} = T^{-1}\mathbf{X}_i'\overline{\mathbf{M}}\mathbf{X}_i$. Compared to Pesaran (2006), and Pesaran and Tosetti (2009), equation (84) has the extra term $\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\widehat{\mathbf{\Psi}}_{iT}^{-1}\left(\frac{\mathbf{X}_i'\overline{\mathbf{M}}\boldsymbol{\eta}_i}{T}\right)$, not encountered pepreviously. We focus on this new term first. Lemma 3 implies

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\widehat{\mathbf{\Psi}}_{iT}^{-1}\left(\frac{\mathbf{X}_i'\overline{\mathbf{M}}\boldsymbol{\eta}_i}{T}\right) - \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\frac{\mathbf{V}_i'\mathbf{M}_g\mathbf{V}_i}{T}\right)^{-1}\left(\frac{\mathbf{V}_i'\mathbf{M}_g\boldsymbol{\eta}_i}{T}\right) \xrightarrow{L_1} 0, \tag{85}$$

as $m_2, T, N \xrightarrow{j} \infty$, such that $N\sum_{\ell}^{m_2}\overline{\lambda}_\ell\lambda_i < K < \infty$. Let $\mathbf{W}_{i\cdot}' = \left(\frac{\mathbf{V}_i'\mathbf{M}_g\mathbf{V}_i}{T}\right)^{-1}\mathbf{V}_i'\mathbf{M}_g$ and denote the tth row of matrix $\mathbf{W}_{i\cdot}$ as $\mathbf{w}_{it}$. Using this notation we write:

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\frac{\mathbf{V}_i'\mathbf{M}_g\mathbf{V}_i}{T}\right)^{-1}\left(\frac{\mathbf{V}_i'\mathbf{M}_g\boldsymbol{\eta}_i}{T}\right) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mathbf{W}_{i\cdot}'\boldsymbol{\eta}_i = \frac{1}{T\sqrt{N}}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{w}_{it}\eta_{it}$$

Using the same method as in the first part of the proof, we define two-dimensional vector array $\left\{\{\boldsymbol{\kappa}_{Nt},\mathcal{F}_t\}_{t=-\infty}^{\infty}\right\}_{N=1}^{\infty}$, as

$$\boldsymbol{\kappa}_{Nt} = \frac{1}{T_N\sqrt{N}}\sum_{i=1}^{N}\mathbf{w}_{it}\sum_{\ell=1}^{m_{2,N}}\lambda_{i\ell}g_{\ell t}, \tag{86}$$

which is identical to (81) except that $\mathbf{w}_{it}$ is used instead of $\widetilde{\mathbf{v}}_{it}$. Following the same steps as in the first part of this proof, we have that $\left\{\{\boldsymbol{\kappa}_{Nt},\mathcal{F}_t\}_{t=-\infty}^{\infty}\right\}_{N=1}^{\infty}$ is L_1-mixingale with respect to constant array $\{c_{Nt}\}$, and a mixingale weak law (Davidson, 1994, Theorem 19.11) establishes $\sum_{t=1}^{T_N}\kappa_{Nt}\xrightarrow{L_1}0$, that is:

$$\frac{1}{T\sqrt{N}}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{w}_{it}\eta_{it}\xrightarrow{L_1}0, \tag{87}$$

as $m_2, T, N \xrightarrow{j} \infty$, such that $N\sum_{\ell}^{m_2}\overline{\lambda}_\ell\lambda_i < K < \infty$. It follows from equations (85) and (87) that

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\widehat{\mathbf{\Psi}}_{iT}^{-1}\left(\frac{\mathbf{X}_i'\overline{\mathbf{M}}\boldsymbol{\eta}_i}{T}\right)\xrightarrow{L_1}0.,$$

as $m_2, T, N \xrightarrow{j} \infty$, such that $N\sum_{\ell}^{m_2}\overline{\lambda}_\ell\lambda_i < K < \infty$. Convergence results for remaining terms on the right side of equation (84) can be established in the same way as in Pesaran and Tosetti (2009) or Pesaran (2006). In particular, we have

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\boldsymbol{\upsilon}_i\xrightarrow{D}N\left(0,\mathbf{\Omega}_\upsilon\right), \text{ as } N\to\infty,$$

and Lemma 3 implies

$$\frac{1}{N}\sum_{i=1}^{N}\widehat{\mathbf{\Psi}}_{iT}^{-1}\left(\frac{\sqrt{N}\mathbf{X}_i'\overline{\mathbf{M}}\mathbf{F}}{T}\right)\gamma_i\xrightarrow{L_1}0,$$

and

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\widehat{\mathbf{\Psi}}_{iT}^{-1}\left(\frac{\mathbf{X}_i'\overline{\mathbf{M}}\mathbf{e}_i}{T}\right)\xrightarrow{L_1}0,$$

as $m_2, T, N \xrightarrow{j} \infty$, such that $N\sum_{\ell}^{m_2}\overline{\lambda}_\ell\lambda_i < K < \infty$. This completes the proof of result (45). ■