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 and CIMF

April 19, 2010

Abstract

This paper introduces the concepts of time-specific weak and strong cross section dependence, and investigates how these notions are related to the concepts of weak, strong and semi-strong common factors, frequently used for modelling residual cross section correlations in panel data models. It then focuses on the problems of estimating slope coefficients in large panels, where cross section units are subject to possibly a large number of unobserved common factors. It is established that the Common Correlated Effects (CCE) estimator introduced by Pesaran (2006) remains asymptotically normal under certain conditions on factors loadings of an infinite factor error structure, including cases where methods relying on principal components fail. 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, Common Correlated Effects (CCE) Estimator

JEL Classification: C10, C31, C33.

*We are grateful to Takashi Yagamata for helpful comments. 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 Manfred Deistler, Pierre Perron (the Editor), and two anonymous referees. The views expressed in the paper are those of the authors and do not necessarily reflect those of the European Central Bank.

†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

The problem of error cross section dependence in panel regressions has attracted considerable attention over the past decade. It is increasingly recognized that conditioning on variables specific to the cross section units alone need not deliver cross section error independence, and neglecting such dependencies can lead to biased estimates and spurious inference. How best to account for cross correlation of errors in panels depends on the nature of the cross dependence, and the size of the time series dimension (T) of the panel relative to its cross section dimension (N). When N is small relative to T, and the errors are uncorrelated with the regressors cross section dependence can be modelled using the Seemingly Unrelated Regression Equations (SURE) approach of Zellner (1962). But when N is large relative to T, the SURE procedure is not feasible. 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)), as a data shrinkage procedure where correlations across many units or variables are modelled by means of a small number of latent factors.

In this paper we show that factor models can be employed more generally to characterize other forms of dependence such as dependence across space or social networks. Initially we introduce the concepts of weak and strong cross section dependence defined at a point in time and with respect to a given information set. These concepts generalize the notions of weak (or idiosyncratic) and strong cross section dependence advanced in the literature. Forni and Lippi (2001), building on Forni and Reichlin (1998), consider a double index process over both dimensions (time and space) simultaneously, and define it as idiosyncratic (or weakly dependent) if the weighted average of the process, computed over both dimensions, converges to zero in quadratic mean for all sets of weights satisfying certain granularity conditions. The double index process is said to be strongly dependent (again over both dimensions) if the weighted averages do not tend to zero.[1] These concepts, that are applicable to dynamic factor models, provide a generalization of the notions of weak and strong dependence developed by Chamberlain (1983) and Chamberlain and Rothschild (1983) for the analysis of static factor models.

Our notions of weak and strong cross section dependence are more widely applicable and does not require the double index process to be stationarity over time, and allow a finer distinction between strong and semi-strong cross section dependence. 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 need not be stationary, and concepts of weak and strong dependence that are more generally applicable are needed. We also

[1] For further developments and discussions see Anderson et al. (2009).

investigate how weak and strong cross section dependence are related to the notions of weak, strong and semi-strong common factors, which may be used to represent very general forms of cross section dependence.

We then turn our attention to the second main concern of this paper, namely the estimation of slope coefficients in the context of panel data models with general cross section error dependence. Building on the first part of the paper, we show that general linear error dependence in panels can be modelled in terms of a factor model with a fixed number of strong factors and a large number of non-strong factors. We allow the number of non-strong factors to rise with N, and establish that the Common Correlated Effects (CCE) estimator introduced by Pesaran (2006) remains consistent and 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 (including the recent iterative Principle Component (PC) procedure proposed by Bai (2009)) and the CCE estimators under alternative assumptions on the nature of unobserved common effects. In particular, we examine and compare the performance of these estimators when the errors are subject to a finite number of unobserved strong factors and an infinite number of weak and/or semi-strong unobserved common factors. As predicted by the theory the CCE estimator performs well and show very little size distortions, which is in contrast with the iterated PC approach of Bai (2009) which exhibit significant size distortions. The latter is partly due to the fact that in the presence of weak or semi-strong factors the PC estimates of factors need not be consistent. This problem does not affect the CCE estimator since it does not aim at consistent estimation of the factors but deals with error cross section dependence generally by using cross section averages to mop up such effects. As shown in Pesaran (2006), the CCE estimator continues to be valid even if the number of factors is larger than the number of cross section averages. The present paper goes one step further and shows that this property holds even if the number of weak factors tend to infinity with N. Note that for variances of the observables to be bounded the number of strong factors must be fixed and can not vary with N.

The plan of the remainder of the paper is as follows. Section 2 introduces the concepts of strong and weak cross section dependence. Section 3 discusses the notions of weak, semi-strong and strong common factors. Section 4 introduces the CCE estimators in the context of panels with an infinite number of common factors. Section 5 describes the Monte Carlo design and discusses the results. Finally, Section 6 provides some concluding remarks. The mathematical details are relegated to appendices.

Notations: $|\lambda_1(\mathbf{A})| \geq |\lambda_2(\mathbf{A})| \geq ... \geq |\lambda_n(\mathbf{A})|$ 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 the Moore-Penrose generalized inverse of $\mathbf{A}$. The column norm of $\mathbf{A} \in \mathbb{M}^{n\times n}$ 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 $\|\mathbf{A}\| = [\lambda_1(\mathbf{AA}')]^{1/2}$, and $\|\mathbf{A}\|_2 = [Tr\,(\mathbf{AA}')]^{1/2}$.

K is used for a fixed positive constant that does not depend on N. Joint convergence of N and T will be denoted by $(N,T) \overset{j}{\to} \infty$. For any random variable x, $\|x\|_{L_p} = (E|x|^p)^{1/p}$, for $p > 1$, denotes L_p norm of x. For any $k \times 1$ vector of random variables $\mathbf{x}_k = (x_1, x_2, ..., x_k)'$, $\|\mathbf{x}_k\|_{L_p} = \left(\sum_{i=1}^k E|x_i|^p\right)^{1/p}$. We use $\overset{L_p}{\to}$ to denote convergence in L_p norm.

2 Cross section dependence in large panels

Consider the double index process $\{z_{it}, i \in \mathbb{N}, t \in \mathbb{Z}\}$, where z_{it} is defined on a suitable probability space, the index t refers to an ordered set such as time, and i refers to units of an unordered population. Our primary focus is on characterizing the correlation structure of the double index process $\{z_{it}\}$ over the cross sectional dimension at a given point in time, t. To this end, we make the following assumptions:

Assumption 1 *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 the conditional mean, $E(\mathbf{z}_{Nt} |\mathcal{I}_{t-1}) = \mathbf{0}$, and the conditional variance, $Var(\mathbf{z}_{Nt} |\mathcal{I}_{t-1}) = \mathbf{\Sigma}_{Nt}$, where $\mathbf{\Sigma}_{Nt}$ is an $N \times N$ symmetric, nonnegative definite matrix. The (i,j)-th element of $\mathbf{\Sigma}_{Nt}$, denoted by $\sigma_{N,ijt}$ is bounded such that $0 < \sigma_{N,iit} \leq K$, for $i = 1, 2, ..., N$, where K is a finite constant independent of N.*

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

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

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

Zero conditional mean in Assumption 1 can be relaxed to $E(\mathbf{z}_{Nt} |\mathcal{I}_{t-1}) = \boldsymbol{\mu}_{N,t-1}$, with $\boldsymbol{\mu}_{N,t-1}$ being a pre-determined function of the elements of $\mathcal{I}_{t-1}$. Assumption 2, known in finance as the granularity condition, ensures that the weights $\{w_{N,it}\}$ 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, $\mathcal{I}_{t-1}$, the weights, $\mathbf{w}_{Nt}$, are distributed independently of $\mathbf{z}_{Nt}$. To simplify the notations in the rest of the paper we suppress the explicit dependence of $\mathbf{z}_{Nt}$, $\mathbf{w}_{Nt}$ and other vectors and matrices and their elements on N.

In the following, we describe our notions of weak and strong cross sectionally dependent processes, and then introduce the related concepts of weak, strong, and semi-strong factors.

2.1 Weak and strong cross section dependence

Consider 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 1 and 2. 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 1 *(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 the information set $\mathcal{I}_{t-1}$, if for any sequence of weight vectors $\{\mathbf{w}_t\}$ satisfying the granularity conditions (1)-(2) we have*

$$\lim_{N \rightarrow \infty} Var(\mathbf{w}_t' \mathbf{z}_t \,|\mathcal{I}_{t-1}) = 0. \tag{3}$$

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

$$Var(\mathbf{w}_t' \mathbf{z}_t \,|\mathcal{I}_{t-1}) \geq K > 0. \tag{4}$$

The concepts of weak and strong cross section dependence proposed here are defined conditional on a given information set, $\mathcal{I}_{t-1}$, which allows us 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 the application under consideration. For dynamic (possibly nonstationary) models the information set could contain all lagged realizations of the process $\{z_{it}\}$, that is $\mathcal{I}_{t-1} = \{\mathbf{z}_{t-1}, \mathbf{z}_{t-2}, ...\}$, or only the starting values of the process. For stationary panels, unconditional variances of cross section averages could be considered. Conditioning information set could also contain contemporaneous realizations, which might be useful in applications where a particular unit has a dominant influence on the rest of the units in the system.

Remark 1 *Anderson et al. (2009) propose definitions of weak and strong cross section dependence for covariance stationary processes, with spectral density $\mathbf{F}_z(\omega)$ (see also Forni and Lippi (2001)). According to their definition, $\{z_{it}\}$ is weakly dependent if the the largest eigenvalue of the spectral density matrix, $\lambda_1^z(\omega)$, is uniformly bounded in ω and N. $\{z_{it}\}$ is strongly dependent if the first $m \geq 1$ $(m < K)$ eigenvalues $(\lambda_1^z(\omega), ..., \lambda_m^z(\omega))$ diverge to infinity as $N \rightarrow \infty$, for all frequencies. 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 note that the notion of weak dependence does not necessarily involve an ordering of the observations or the specification of a distance metric across the observations.*

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

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 for any N sufficiently large (and as $N \to \infty$), $N^{-1}\lambda_1(\mathbf{\Sigma}_t) \geq K > 0$.

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

$$Var(\mathbf{w}_t'\mathbf{z}_t\,|\mathcal{I}_{t-1}) = \mathbf{w}_t'\mathbf{\Sigma}_t\mathbf{w}_t \leq \left(\mathbf{w}_t'\mathbf{w}_t\right)\lambda_1(\mathbf{\Sigma}_t), \tag{5}$$

and under the granularity conditions (1)-(2) 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 (5), it follows that $\lambda_1(\mathbf{\Sigma}_t)$ tends to infinity at least at the rate N. Hence, under CSD $N^{-1}\lambda_1(\mathbf{\Sigma}_t) \geq K > 0$ for any N sufficiently large. Note that $\lambda_1(\mathbf{\Sigma}_t) \leq \sum_{i=1}^{N} \sigma_{ii,t}$ where, under Assumption 1, $\sigma_{ii,t}$ are finite, $\lambda_1(\mathbf{\Sigma}_t)$ cannot diverge to infinity at a rate faster than N. To prove the reverse relation, first note that, from the Rayleigh-Ritz theorem[2],

$$\lambda_1(\mathbf{\Sigma}_t) = \max_{\mathbf{v}_t'\mathbf{v}_t=1} \mathbf{v}_t'\mathbf{\Sigma}_t\mathbf{v}_t = \mathbf{v}_t^{*\prime}\mathbf{\Sigma}_t\mathbf{v}_t^{*}. \tag{6}$$

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

$$\lambda_1(\mathbf{\Sigma}_t) = N \cdot Var(\mathbf{w}_t^{*\prime}\mathbf{z}_t\,|\mathcal{I}_{t-1}). \tag{7}$$

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

Since $\lambda_1(\mathbf{\Sigma}_t) \leq \|\mathbf{\Sigma}_t\|_1$, [3] it follows from (5) that both the spectral radius and the column norm of the covariance matrix of a CSD process are unbounded in N. A similar condition also arises in the case of time series processes with long memory or strong temporal dependence where the autocorrelation coefficients are not absolutely summable. (Robinson (2003)).

Remark 3 *The definition of idiosyncratic process by Forni and Lippi (2001) differs from our definition of CWD in terms of the weights used to construct the weighted averages. While Forni and*

[2] See Horn and Johnson (1985), p.176.
[3] See Horn and Johnson (1985), pp. 297-298.

Lippi assume $\lim_{N\to\infty}\|\mathbf{w}\| = 0$*, our granularity conditions (1)-(2) 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 important implications for the cross sectional properties of the processes. In particular, under* $\lim_{N\to\infty}\|\mathbf{w}_t\| = 0$*, 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 (1)-(2), it is clear that if* $\lambda_1(\mathbf{\Sigma}_t) = O(N^{1-\epsilon})$ *for any* $\epsilon > 0$*, then, using (5),*

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

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.*

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 needed. In Section 4 we consider estimation of slope coefficients in panels with an infinite factor structure, where eigenvalues of the error covariance matrix are allowed to increase at a rate slower than N.

3 Common factor models

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

$$z_{it} = \gamma_{i1}f_{1t} + \gamma_{i2}f_{2t} + ... + \gamma_{iN}f_{Nt} + \varepsilon_{it}, \quad i = 1, 2, ..., N, \tag{8}$$

or in matrix notations

$$\mathbf{z}_t = \mathbf{\Gamma}\mathbf{f}_t + \boldsymbol{\varepsilon}_t, \tag{9}$$

where $\mathbf{f}_t = (f_{1t}, f_{2t}, ..., f_{Nt})'$, $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \varepsilon_{2t}, ..., \varepsilon_{Nt})'$, and the common factors, $f_{\ell t}$, and the idiosyncratic errors, ε_{it}, satisfy the following assumptions:

Assumption 3 *The* $N \times 1$ *vector* $\mathbf{f}_t$ *is a zero mean 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 4 $Var(\varepsilon_{it} |\mathcal{I}_{t-1}) = \sigma_i^2 < K < \infty$*,* ε_{it} *and* ε_{jt} *are independently distributed for all* $i \neq j$ *and for all* t*. Specifically,* $\max_i(\sigma_i^2) = \sigma_{\max}^2 < K < \infty$*.*

The process z_{it} in (8) has conditional variance

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

For the conditional variance of z_{it} to be bounded in N, as required by Assumption 1, we must have

$$\sum_{\ell=1}^{N} \gamma_{i\ell}^2 \leq K < \infty, \text{ for } i = 1, 2, ..., N. \tag{10}$$

In what follows we also consider the slightly stronger absolute summability condition

$$\sum_{\ell=1}^{N} |\gamma_{i\ell}| \leq K < \infty, \text{ for } i = 1, 2, ..., N. \tag{11}$$

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

$$\lim_{N\to\infty} N^{-1} \sum_{i=1}^{N} |\gamma_{i\ell}| = K > 0. \tag{12}$$

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

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

The literature on large factor models has focussed on the case where the factors are strong. The case of weak factors is recently considered by Onatski (2009). It is also possible to consider semi-strong or semi-weak factors. In general, let α be a positive constant in the range $0 \leq a \leq 1$ and consider the condition

$$\lim_{N\to\infty} N^{-\alpha} \sum_{i=1}^{N} |\gamma_{i\ell}| = K < \infty. \tag{14}$$

The strong and weak factors correspond to the two values of $\alpha = 1$ and $\alpha = 0$, respectively. For any other values of $\alpha \in (0, 1)$ the factor $f_{\ell t}$ can be said to be semi-strong or semi-weak. It will prove useful to associate the semi-weak factors with values of $0 < \alpha < 1/2$, and the semi-strong factors with values of $1/2 \leq \alpha < 1$. In Section 4 we provide some practical examples where such semi-strong factors may exist.

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

Theorem 1 *Consider the factor model (9) and suppose that Assumptions 1-4, and the absolute summability condition (11) hold, and there exists a positive constant α in the range $0 \leq a \leq 1$, such that condition (14) hold for any $\ell = 1, 2, .., N$. Then the following statements hold:*

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

(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. Using (9), the covariance of $\mathbf{z}_t$ is given by

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

where $\mathbf{\Lambda}_\varepsilon$ is a diagonal matrix with elements σ_i^2. Since condition (14) holds for $\ell = 1, 2, ..., N$ then $\|\mathbf{\Gamma}\|_1 = O(N^\alpha)$, and noting that $\|\mathbf{\Gamma}'\|_1 = \|\mathbf{\Gamma}\|_\infty = O(1)$ *by* (11) then

$$\lambda_1(\mathbf{\Sigma}) \leq \left\|\mathbf{\Gamma}\mathbf{\Gamma}' + \mathbf{\Lambda}_\varepsilon\right\|_1 \leq \|\mathbf{\Gamma}\|_1 \left\|\mathbf{\Gamma}'\right\|_1 + \sigma_{\max}^2 = O(N^\alpha). \tag{15}$$

But using (5),

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

and when $\alpha < 1$, we have,

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

for any weights $\mathbf{w}$ satisfying condition (1). It follows that $\{z_{it}\}$ is CWD, which establishes result (i). Now suppose that $\{z_{it}\}$ is CSD. Then, noting that $\sigma_{\max}^2 < K < \infty$,

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

Given that, by assumption, $\|\mathbf{\Gamma}'\|_1$ is bounded in N, it follows that $\lim_{N\to\infty} N^{-1}\|\mathbf{\Gamma}\|_1 = K > 0$, and there exists at least one strong factor in (9). To prove the reverse, assume that there exists at least one strong factor in (9) (i.e., $\lim_{N\to\infty} N^{-1}\|\mathbf{\Gamma}\|_1 = K > 0$). Noting that[4]

$$\lambda_1^{1/2}(\mathbf{\Sigma}) \geq \lambda_1^{1/2}\left(\mathbf{\Gamma}\mathbf{\Gamma}'\right) \geq \frac{\|\mathbf{\Gamma}\|_1}{\sqrt{N}}. \tag{16}$$

it follows that $\lim_{N\to\infty} N^{-1}\lambda_1(\mathbf{\Sigma}) = K > 0$ and the process is CSD, which establishes result (ii). ■

Under (12)-(13), z_{it} can be decomposed as

$$z_{it} = z_{it}^s + z_{it}^w, \tag{17}$$

where

$$z_{it}^s = \sum_{\ell=1}^{m} \gamma_{i\ell} f_{\ell t}; \quad z_{it}^w = \sum_{\ell=m+1}^{N} \gamma_{i\ell} f_{\ell t} + \varepsilon_{it}, \tag{18}$$

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

and $\gamma_{i\ell}$ satisfy conditions (12) for $\ell = 1, ..., m$, and (13) for $\ell = m+1, ..., N$. In the light of Theorem 1, it follows that z_{it}^s is CSD and z_{it}^w is CWD. Also, notice that when $m = 0$, we have a model with no strong factors and potentially an infinite number of weak factors.

Remark 4 *Consider the following general spatial process*

$$\mathbf{z}_t = \mathbf{R}\mathbf{v}_t, \tag{19}$$

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 (2010) 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 (19). 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 across the units, 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 an infinite number of 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* $i, \ell = 1, ..., N$. *Under the bounded column and row norms of* $\mathbf{R}$, *the loadings in the above factor structure satisfy (13), and hence* z_{it} *will be a CWD process.*

Remark 5 *Consistent estimation of factor models with weak or semi-strong factors may be problematic. To see this, consider the following single factor model with known factor loadings*

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

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

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

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. *See also Onatski (2009).*

4 CCE estimation of panel data models with an infinite number of factors

In this section we focus on consistent estimation of slopes in panel regression models where the error terms have an infinite order factor structure. Let y_{it} be the observation on the i^{th} cross section unit at time t, for $i = 1, 2, ..., N$, and $t = 1, 2, ..., T$, and suppose that it is generated as

$$y_{it} = \boldsymbol{\alpha}_i' \mathbf{d}_t + \boldsymbol{\beta}_i' \mathbf{x}_{it} + e_{it}, \tag{20}$$

where $\mathbf{d}_t = (d_{1t}, d_{2t}, ..., d_{m_d t})'$ is a $m_d \times 1$ vector of *observed* common effects, and $\mathbf{x}_{it}$ is a $k \times 1$ vector of observed individual specific regressors. The parameters of interest are the means of individual slope coefficients, $\boldsymbol{\beta} = E(\boldsymbol{\beta}_i)$.[5] The error term, e_{it}, is given by the following general factor structure,

$$e_{it} = \sum_{\ell=1}^{m_f} \gamma_{i\ell} f_{\ell t} + \sum_{\ell=1}^{m_n} \lambda_{i\ell} n_{\ell t} + \varepsilon_{it}, \tag{21}$$

where we have distinguished between two types of unobserved common factors, $\mathbf{f}_t = \left(f_{1t}, f_{2t}, ..., f_{m_f t}\right)'$ and $\mathbf{n}_t = (n_{1t}, n_{2t}, ..., n_{m_n t})'$. The former are strong factors that are possibly correlated with the regressors $\mathbf{x}_{it}$, while the latter are the weak, semi-weak or semi-strong factors that are assumed to be uncorrelated with the regressors. The associated vectors of factor loadings will be denoted by $\boldsymbol{\gamma}_i = (\gamma_{i1}, \gamma_{i2}, ..., \gamma_{im_f})'$ and $\boldsymbol{\lambda}_i = (\lambda_{i1}, \lambda_{i2}, ..., \lambda_{im_n})'$, respectively. The cross section dependence of errors are modelled using the unobserved common factors, $\mathbf{f}_t$ and $\mathbf{n}_t$, and without loss of generality it is assumed that the idiosyncratic errors, ε_{it}, are cross sectionally uncorrelated (although they can be serially correlated).

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

$$\mathbf{x}_{it} = \mathbf{A}_i' \mathbf{d}_t + \boldsymbol{\Gamma}_i' \mathbf{f}_t + \mathbf{v}_{it}, \tag{22}$$

where $\mathbf{A}_i'$ and $\boldsymbol{\Gamma}_i'$ are $k \times m_d$ and $k \times m_f$ factor loading matrices, and $\mathbf{v}_{it}$ is the individual component of $\mathbf{x}_{it}$, assumed to be distributed independently of the innovations e_{it}.

Similar panel data models have been analyzed by Pesaran (2006), Kapetanios, Pesaran, and Yagamata (2010), and Pesaran and Tosetti (2010). Pesaran (2006) introduced CCE estimators in a panel model where m_f is fixed, $m_n = 0$, 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_f could be larger than $k + 1$. Kapetanios, Pesaran, and Yagamata (2010) extended the results of Pesaran (2006) by allowing unobserved common factors to follow unit root processes. In both papers, innovations $\{\varepsilon_{it}\}$ are assumed to be cross sectionally independent although possibly serially correlated. This assumption is relaxed by Pesaran and Tosetti (2010) who assume that $\{\varepsilon_{it}\}$ is a weakly dependent process with bounded row and column norms of its variance matrix, 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 somewhere in between. Our analysis is thus an extension of Pesaran (2006) to the case where there are an infinite number

[5]We assume that individual slope coefficients are drawn from a common distribution with mean $\boldsymbol{\beta}$.

of factors, a fixed number of which are strong and the rest are either weak, semi-weak or semi-strong factors.

The special case where both m_f and m_n are fixed has already been analyzed in the above cited papers. The case where $f_{1t}, f_{2t}, ..., f_{m_f t}$ are strong and $m_f = m_f(N) \to \infty$ as $N \to \infty$, is not that meaningful as it will lead to unbounded variances as $N \to \infty$. However, it would be possible to let the number of non-strong factors to rise with N, whilst keeping the number of strong factors fixed. We show below that the CCE type estimators continue to be consistent and asymptotically normal under these types of infinite-factor error structures. We use notations $m_n(N)$ to emphasize the dependence of the number of non-strong factors on N in the remainder of this paper.

Equations (20) and (22) can be written more compactly as

$$\mathbf{z}_{it} = \begin{pmatrix} y_{it} \\ \mathbf{x}_{it} \end{pmatrix} = \mathbf{B}_i'\mathbf{d}_t + \mathbf{C}_i'\mathbf{f}_t + \mathbf{u}_{it}, \tag{23}$$

where

$$\begin{aligned} \mathbf{B}_i &= \begin{pmatrix} \boldsymbol{\alpha}_i & \mathbf{A}_i \end{pmatrix} \mathbf{D}_i,\ \mathbf{C}_i = \begin{pmatrix} \boldsymbol{\gamma}_i & \boldsymbol{\Gamma}_i \end{pmatrix} \mathbf{D}_i, \\ \mathbf{D}_i &= \begin{pmatrix} 1 & \mathbf{0}_{1\times k} \\ \boldsymbol{\beta}_i & \mathbf{I}_k \end{pmatrix},\ \mathbf{u}_{it} = \begin{pmatrix} \boldsymbol{\lambda}_i'\mathbf{n}_t + \varepsilon_{it} + \boldsymbol{\beta}_i'\mathbf{v}_{it} \\ \mathbf{v}_{it} \end{pmatrix}. \end{aligned} \tag{24}$$

Stacking the T observations for each i we also have

$$\begin{aligned} \mathbf{y}_i &= \mathbf{D}\boldsymbol{\alpha}_i + \mathbf{X}_i\boldsymbol{\beta}_i + \mathbf{e}_i, \\ \mathbf{X}_i &= \mathbf{G}\boldsymbol{\Pi}_i + \mathbf{V}_i, \\ \mathbf{Z}_i &= \mathbf{D}\mathbf{B}_i + \mathbf{F}\mathbf{C}_i + \mathbf{U}_i, \end{aligned} \tag{25}$$

where $\mathbf{y}_i = (y_{i1}, y_{i2}, ..., y_{iT})'$, $\mathbf{D} = (\mathbf{d}_1, \mathbf{d}_2, ..., \mathbf{d}_T)'$, $\mathbf{X}_i = (\mathbf{x}_{i1}, \mathbf{x}_{i2}, ..., \mathbf{x}_{iT})'$, $\mathbf{G} = (\mathbf{D}, \mathbf{F})$, $\mathbf{F} = (\mathbf{f}_1, \mathbf{f}_2, ..., \mathbf{f}_T)'$, $\mathbf{V}_i = (\mathbf{v}_{i1}, \mathbf{v}_{i2}, ..., \mathbf{v}_{iT})'$, $\mathbf{Z}_i = (\mathbf{z}_{i1}, \mathbf{z}_{i2}, ..., \mathbf{z}_{iT})'$, $\mathbf{U}_i = (\mathbf{u}_{i1}, \mathbf{u}_{i2}, ..., \mathbf{u}_{iT})'$, and $\boldsymbol{\Pi}_i = (\mathbf{A}_i', \boldsymbol{\Gamma}_i')'$.

For the development of the CCE estimators we need the cross section averages of the individual specific variables $\mathbf{z}_{it} = (y_{it}, \mathbf{x}_{it}')'$, which we denote by $\overline{\mathbf{z}}_{wt} = \sum_{i=1}^N w_i\mathbf{z}_{it}$, where $\mathbf{w} = (w_1, w_1, ..., w_N)'$ is any vector of weights that satisfy the granularity conditions (1)-(2). Further, let $\overline{\mathbf{M}}_w = \mathbf{I}_T - \overline{\mathbf{H}}_w \left(\overline{\mathbf{H}}_w'\overline{\mathbf{H}}_w\right)^+ \overline{\mathbf{H}}_w'$, $\overline{\mathbf{H}}_w = \left(\mathbf{D}, \overline{\mathbf{Z}}_w\right)$, $\overline{\mathbf{Z}}_w = (\overline{\mathbf{z}}_{w1}, \overline{\mathbf{z}}_{w2}, ..., \overline{\mathbf{z}}_{wT})'$, $\overline{\mathbf{M}}_q = \mathbf{I}_T - \mathbf{Q}\left(\mathbf{Q}'\mathbf{Q}\right)^+\mathbf{Q}'$, $\mathbf{Q} = \mathbf{G}\overline{\mathbf{P}}_w$,

$$\underset{(m_d+m_f)\times(m_d+k+1)}{\overline{\mathbf{P}}_w} = \begin{pmatrix} \mathbf{I}_{m_d} & \underset{m_d\times(k+1)}{\overline{\mathbf{B}}_w} \\ \underset{m_f\times m_d}{\mathbf{0}} & \underset{m_f\times(k+1)}{\overline{\mathbf{C}}_w} \end{pmatrix}, \tag{26}$$

$$\overline{\mathbf{B}}_w = \sum_{i=1}^N w_i\mathbf{B}_i, \text{ and } \overline{\mathbf{C}}_w = \sum_{i=1}^N w_i\mathbf{C}_i. \tag{27}$$

Also, define the matrices associated with $\overline{\mathbf{M}}_q$ and $\overline{\mathbf{P}}_w$ as $\mathbf{M}_g = \mathbf{I}_T - \mathbf{G}\left(\mathbf{G}'\mathbf{G}\right)^{-1}\mathbf{G}'$, and

$$\mathbf{P} = \begin{pmatrix} \mathbf{I}_{m_d} & \mathbf{B} \\ \underset{m_f \times m_d}{\mathbf{0}} & \mathbf{C} \end{pmatrix}, \tag{28}$$

where $\mathbf{B} = E(\mathbf{B}_i)$, and $\mathbf{C} = E(\mathbf{C}_i)$. As we shall see below the asymptotic theory of the CCE type estimators depends on the rank of $\overline{\mathbf{C}}_w$ both for a finite N, and as $N \to \infty$.

We make the following assumptions on the unobserved common factors $\mathbf{f}_t$ and $\mathbf{n}_t$ and their loadings.

Assumption 5 *(Common factors) The $(m_d + m_f) \times 1$ vector $\mathbf{g}_t = (\mathbf{d}_t', \mathbf{f}_t')'$ is a covariance stationary process, with absolute summable autocovariances and finite second-order moments. In particular, $\|\mathbf{\Sigma}_g\| < K$ for some constant K, where $\mathbf{\Sigma}_g = E(\mathbf{g}_t\mathbf{g}_t')$ is a positive definite matrix. For each $\ell = 1, 2, ..., m_n(N)$, common factor $n_{\ell t}$ follows a covariance stationary process with absolute summable autocovariances, zero mean, unit variance, and finite fourth-order moment uniformly bounded in ℓ. $n_{\ell t}$ is independently distributed of $\mathbf{g}_t$ and of $n_{\ell' t}$ for all $\ell \neq \ell'$ and t.*

Assumption 6 *(Factor loadings)*

(a) Factor loadings $\boldsymbol{\gamma}_i$, and $\mathbf{\Gamma}_i$ are independently and identically distributed across i, and of the common factors $\mathbf{g}_t$, $\mathbf{n}_t$, for all i and t, with fixed mean $\boldsymbol{\gamma}$ and $\mathbf{\Gamma}$, and uniformly bounded second moments. In particular,

$$\boldsymbol{\gamma}_i = \boldsymbol{\gamma} + \boldsymbol{\eta}_{\gamma i},\ \boldsymbol{\eta}_{\gamma i} \sim IID\left(\mathbf{0}, \mathbf{\Omega}_\gamma\right), \text{ for } i = 1, 2, ..., N,$$

and

$$vec\left(\mathbf{\Gamma}_i\right) = vec\left(\mathbf{\Gamma}\right) + \boldsymbol{\eta}_{\Gamma i},\ \boldsymbol{\eta}_{\Gamma i} \sim IID\left(\mathbf{0}, \mathbf{\Omega}_\Gamma\right), \text{ for } i = 1, 2, ..., N,$$

where $\mathbf{\Omega}_\gamma$ and $\mathbf{\Omega}_\Gamma$ are $m_f \times m_f$ and $k\, m_f \times k\, m_f$ symmetric nonnegative definite matrices, $\|\boldsymbol{\gamma}\| < K$, $\|\mathbf{\Omega}_\gamma\| < K$, $\|\mathbf{\Gamma}\| < K$, and $\|\mathbf{\Omega}_\Gamma\| < K$ for some constant K.

(b) Factor loadings $\lambda_{i\ell}$, for $i = 1, 2, ..., N$, and $\ell = 1, 2, ..., m_n(N)$, are non-stochastic. For each $i = 1, 2, ..., N$, the factor loadings, $\lambda_{i\ell}$, satisfy the following absolute summability condition

$$\sum_{\ell=1}^{m_n(N)} |\lambda_{i\ell}| < K. \tag{29}$$

Remark 6 *The absolute summability condition (29) is sufficient for ensuring bounded variances of $\vartheta_{it} = \boldsymbol{\lambda}_i'\mathbf{n}_t = \sum_{\ell=1}^{m_n(N)} \lambda_{i\ell} n_{\ell t}$ for each $i = 1, 2, ..., N$, as $m_n(N) \to \infty$. This condition alone does not, however, rule out strong, semi strong, or semi weak factor structures. Additional requirements on the sum of absolute values of the loadings $\lambda_{i\ell}$ across i will be postulated in theorems below.*

The following assumptions are similar to Pesaran (2006).

Assumption 7 *The individual-specific errors ε_{it} and $\mathbf{v}_{it}$ are independently distributed across i, independently distributed of the common factors $\mathbf{g}_t$, $\mathbf{n}_t$ and of the factor loadings $\boldsymbol{\gamma}_j$, $\boldsymbol{\Gamma}_j$, for each i,j and each t. $\mathbf{v}_{it}$, for $i = 1, 2, ..., N$, follow linear stationary processes with absolute summable autocovariances, zero mean, and finite second-order moments uniformly bounded in i. For each i,*

$$E(\mathbf{v}_{it}\mathbf{v}_{it}') = \boldsymbol{\Sigma}_{v_i},$$

where $\boldsymbol{\Sigma}_{vi}$ is a positive definite matrix, such that $\sup_i \|\boldsymbol{\Sigma}_{vi}\| < K$, for some positive constant K. Errors ε_{it}, for $i = 1, 2, ..., N$, follow a linear stationary process with absolute summable autocovariances, zero mean, and finite second-order moments uniformly bounded in i.

Assumption 8 *Coefficient matrices $\mathbf{B}_i$ are independently and identically distributed across i, independently distributed of the common factors $\mathbf{g}_t$ and $\mathbf{n}_t$, of the factor loadings $\boldsymbol{\gamma}_j$ and $\boldsymbol{\Gamma}_j$, and of the errors ε_{jt} and $\mathbf{v}_{jt}$, for all i, j and t, with fixed mean $\mathbf{B}$, and uniformly bounded second moments.*

Assumption 9 *The slope coefficients follow the random coefficient model*

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

where $\|\boldsymbol{\beta}\| < K$, $\|\boldsymbol{\Omega}_\beta\| < K$, $\boldsymbol{\Omega}_\beta$ is a symmetric non-negative definite matrix, and the random deviations $\boldsymbol{\upsilon}_i$ are distributed independently of the common factors $\mathbf{g}_t$ and $\mathbf{n}_t$, of the factor loadings $\boldsymbol{\gamma}_j$ and $\boldsymbol{\Gamma}_j$, of the errors ε_{jt} and $\mathbf{v}_{jt}$, and of the coefficients in $\boldsymbol{\alpha}_j$ and $\mathbf{A}_j$ for all i, j and t.

Assumption 10

(a) The matrix $\lim_{N\to\infty}\sum_{i=1}^N w_i\boldsymbol{\Sigma}_{iq} = \boldsymbol{\Psi}^*$ exists and is nonsingular, and $\sup_i \left\|\boldsymbol{\Sigma}_{iq}^{-1}\right\| < K$, where $\boldsymbol{\Sigma}_{iq} = \boldsymbol{\Sigma}_{vi} + \boldsymbol{\Pi}_i^{*\prime}\boldsymbol{\Sigma}_g\boldsymbol{\Pi}_i^*$, and $\boldsymbol{\Pi}_i^* = \left[\mathbf{I} - \mathbf{P}\left(\mathbf{P}'\mathbf{P}\right)^+\mathbf{P}'\right]\boldsymbol{\Pi}_i$.

(b) Denote the t-th row of matrix $\widetilde{\mathbf{X}}_i = \overline{\mathbf{M}}_q\mathbf{X}_i$ by $\widetilde{\mathbf{x}}_{it}' = (\widetilde{x}_{i1t}, \widetilde{x}_{i2t}, ..., \widetilde{x}_{ikt})$. Individual elements of the vector $\widetilde{\mathbf{x}}_{it}'$ have uniformly bounded fourth moments, namely there exists a positive constant K such that $E\left(\widetilde{x}_{ist}^4\right) < K$ for any $t = 1, 2, ..., T$, $i = 1, 2, ..., N$ and $s = 1, 2, ..., k$. Furthermore, fourth moments of $f_{\ell t}$, for $\ell = 1, 2, ..., m_f$, are bounded.

(c) There exists T_0 such that for all $T \geq T_0$, $\left(\sum_{i=1}^N w_i\mathbf{X}_i'\overline{\mathbf{M}}_w\mathbf{X}_i/T\right)^{-1}$ exists.

(d) 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 $\left(\mathbf{X}_i'\overline{\mathbf{M}}_w\mathbf{X}_i/T\right)^{-1}$ and $\left(\mathbf{X}_i'\mathbf{M}_g\mathbf{X}_i/T\right)^{-1}$ exist for all i.

The CCE approach is motivated by the fact that, to estimate $\boldsymbol{\beta}$, one does not necessarily need to compute consistent estimates of the unobservable common factors. It is sufficient to account for their effects by including cross section averages of the observables in the regressions, since such cross section averages indirectly reflect the overall importance of the factors for the estimation of $\boldsymbol{\beta}$. Two types of CCE estimators are considered. The common correlated effects mean group estimator (CCEMG) which is given by

$$\widehat{\boldsymbol{\beta}}_{MG} = \frac{1}{N}\sum_{i=1}^{N}\widehat{\boldsymbol{\beta}}_i, \tag{30}$$

where $\widehat{\boldsymbol{\beta}}_i = \left(\mathbf{X}_i'\overline{\mathbf{M}}_w\mathbf{X}_i\right)^{-1}\mathbf{X}_i'\overline{\mathbf{M}}_w\mathbf{y}_i$, and the common correlated effects pooled (CCEP) estimator which is defined by

$$\widehat{\boldsymbol{\beta}}_P = \left(\sum_{i=1}^{N} w_i\mathbf{X}_i'\overline{\mathbf{M}}_w\mathbf{X}_i\right)^{-1}\sum_{i=1}^{N} w_i\mathbf{X}_i'\overline{\mathbf{M}}_w\mathbf{y}_i. \tag{31}$$

The following theorem establishes consistency of CCE estimators in case of panels with (possibly) an infinite number of factors.

Theorem 2 *(Consistency of CCE estimators) Consider the panel data model (20) and (22), and suppose that Assumptions 5-10 hold, and there exist constants* α *and* K *such that* $0 \leq \alpha < 1$,

$$\sum_{i=1}^{N}|\lambda_{i\ell}| < K\,N^{\alpha} \text{ for each } \ell = 1, 2, ..., m_n(N), \tag{32}$$

and

$$\lim_{N\to\infty}\frac{m_n(N)}{N^{2(1-\alpha)}} \to 0. \tag{33}$$

Then common correlated effects mean group and pooled estimators, defined by (30) and (31), respectively, are consistent, that is as $(N, T) \xrightarrow{j} \infty$ *we have*

$$\widehat{\boldsymbol{\beta}}_{MG} - \boldsymbol{\beta} \xrightarrow{p} \mathbf{0}, \tag{34}$$

and

$$\widehat{\boldsymbol{\beta}}_{P} - \boldsymbol{\beta} \xrightarrow{p} \mathbf{0}. \tag{35}$$

A proof is provided in the Appendix.

Assumptions of Theorem 2 rule out the case where $\vartheta_{it} = \boldsymbol{\lambda}_i'\mathbf{n}_t = \sum_{\ell=1}^{m_n(N)}\lambda_{i\ell}n_{\ell t}$ is a strong factor structure, but allow for the possibility of semi-strong ($1/2 \leq \alpha < 1$), semi-weak ($0 < \alpha < 1/2$), or weak factors ($\alpha = 0$) so long as the number of factors $m_n(N)$ is appropriately bounded. The sufficient bound for $m_n(N)$ is given by condition (33). Note that conditions (32)-(33) and $0 \leq \alpha < 1$ ensure that $Var\left(\overline{\vartheta}_{wt}\right) \to 0$, as $N \to \infty$, and therefore ϑ_{it} is CWD.

The following theorem establishes asymptotic distribution of CCE estimator in case of weak ($\alpha = 0$) and semi weak ($0 < \alpha < 1/2$) infinite factor structures.

Theorem 3 *(Distribution of CCE estimators) Consider the panel data model (20) and (22), and* *suppose that Assumptions 5-10 hold, and there exist constants α and K such that $0 \leq \alpha < 1/2$,*

$$\sum_{i=1}^{N} |\lambda_{i\ell}| < K\, N^{\alpha} \text{ for each } \ell = 1, 2, ..., m_n(N), \tag{36}$$

and

$$m_n(N) < K\, N^{1-2\alpha}. \tag{37}$$

Then, as $(N,T) \xrightarrow{j} \infty$,

$$\sqrt{N}\left(\widehat{\boldsymbol{\beta}}_{MG} - \boldsymbol{\beta}\right) \xrightarrow{d} N\left(\mathbf{0}, \boldsymbol{\Sigma}_{MG}\right), \tag{38}$$

where $\widehat{\boldsymbol{\beta}}_{MG}$ is given by (30), and $\boldsymbol{\Sigma}_{MG}$ is given by equation (B.25) in the Appendix. Furthermore,

$$\left(\sum_{i=1}^{N} w_i^2\right)^{-1/2} \left(\widehat{\boldsymbol{\beta}}_{P} - \boldsymbol{\beta}\right) \xrightarrow{d} N\left(\mathbf{0}, \boldsymbol{\Sigma}_{P}\right), \tag{39}$$

where $\widehat{\boldsymbol{\beta}}_{P}$ is given by (31), and $\boldsymbol{\Sigma}_{P}$ is given by equation (B.19) in the Appendix.

A proof is provided in the Appendix.

Remark 7 *Following Pesaran (2006), it is also possible to provide semi-parametric estimators of variances of $\widehat{\boldsymbol{\beta}}_{MG}$ and $\widehat{\boldsymbol{\beta}}_{P}$. Consistent estimators of $\boldsymbol{\Sigma}_{MG}$ and $\boldsymbol{\Sigma}_{P}$ are given by equations (58) and (69) of Pesaran (2006), respectively.*

Remark 8 *As it was mentioned earlier, CCE estimators are valid irrespective whether $\overline{\mathbf{C}}_w$ defined by (27) has full column rank, or is rank deficient, and therefore m_f, the number of factors in $\mathbf{f}_t$, could be larger than $k+1$. If assumption of full column rank of $\overline{\mathbf{C}}_w$ (for any $N \in \mathbb{N}$, as well as $N \to \infty$) is satisfied, then Assumption 6.a on factor loadings and Assumption 8 on coefficient matrices could be relaxed. In particular, it would be sufficient to assume that factor loadings $\boldsymbol{\gamma}_i$ and $\boldsymbol{\Gamma}_i$, and the coefficients $\boldsymbol{\alpha}_i$ and $\mathbf{A}_i$ are non-stochastic and uniformly bounded.*

Current factor literature assumes that eigenvalues of the spectral density matrix of the underlying double indexed processes either rise with N at the rate N or are bounded in N, while they are not allowed to rise at any rate slower than N. As the sources of cross section dependence are generally unknown (factors are latent and in general not identified), such assumptions seem to have been adopted for technical convenience rather than on grounds of their empirical validity. However, in several empirical applications it seems reasonable to consider cases where the eigenvalues of the spectral density rise at a rate slower than N. Semi-strong factors may exist if there is a cross section unit or an unobserved common factor that affects, rather then all units, only a subset of them expanding at rate slower than N. One can think of an unobserved common shock that hits only a subset of the population. For example a new law that affects only large firms. As the number of firms, N, increases, one reasonable assumption is that the number of large firms increases at a rate slower than

N. Similarly, the performance of medium-sized firms may have impact only on a subset of firms in the market. If we assume that the range of influence of this firm is proportional to its dimension, then as N increases, the subset of units that is affected by it expands at a rate slower than N.

We observe that practical difficulties encountered when estimating the number of factors in large data sets could be related to the presence of semi-strong factors, as existing techniques for determining the number of factors assume that there are no semi-weak (or semi-strong) factors and that all factors under consideration are either weak or strong.

5 Monte Carlo experiments

We consider the following data generating process

$$y_{it} = \alpha_i d_{1t} + \beta_{i1} x_{i1t} + \beta_{i2} x_{i2t} + u_{it}, \tag{40}$$

for $i = 1, 2, ..., N$ and $t = 1, 2, ..., T$. We assume heterogeneous slopes, and set $\beta_{ij} = \beta_j + \eta_{ij}$, with $\eta_{ij} \sim IIDN\,(1, 0.04)$, for $i = 1, 2, ..., N$ and $j = 1, 2$, varying across replications. The errors, u_{it}, are generated as

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

where $\varepsilon_{it} \sim N(0, \sigma_i^2)$, $\sigma_i^2 \sim IIDU\,(0.5, 1.5)$, 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.

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

The first three factors will be assumed to be strong, in the sense that the sum of the absolute values of their loadings is 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_n factors $n_{\ell t}$:

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

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

It is easily seen that for each ℓ, $\sum_{i=1}^{N} |\lambda_{i\ell}| = O(1)$ and for each i, $\sum_{\ell=1}^{m_n} \lambda_{i\ell}^2 = O(m_n/N^2)$.

Therefore, asymptotically as $N \to \infty$, the R_i^2 is only affected by the strong factors, even if $m_n \to \infty$.

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

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

In this case, for each ℓ, $\sum_{i=1}^{N} |\lambda_{i\ell}| = O(N^{1/2})$, and for each i, $\sum_{\ell=1}^{m_n} \lambda_{i\ell}^2 = O(m_n/N)$, and the signal-to-noise ratio of the regressions deteriorate as m_n is increased for any given N. In Section 5.1, we will investigate this issue further, to check if the effect of m_n 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:

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

where

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

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

The observed common effects are generated as

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

When generating v_{ijt} and the common factors $f_{\ell t}, n_{\ell t}$ 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 $m_n = 0, N/5, 3N/5, N$. For example, for $N = 100$, we consider $m_n = 0, 20, 60, 100$. We report bias, RMSE, size and power for six estimators: the FE estimator with standard variance, the

CCEMG and CCEP estimators given by (30) and (31), 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 (30) and (31) except that the cross section averages are replaced by estimated common factors using the Bai and Ng (2002) procedure to $\mathbf{z}_{it} = (y_{it}, \mathbf{x}'_{it})'$. 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 $\hat{m}_f$ 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 $\mathbf{D}_0 = (NT)^{-1}\sum_{i=1}^{N}\mathbf{Z}'_i\mathbf{Z}_i$, $\mathbf{D}_Z = N^{-1}\sum_{i=1}^{N}\hat{\sigma}_i^2\left(T^{-1}\sum_{t=1}^{T}\mathbf{z}_{it}\mathbf{z}'_{it}\right)$, with $\hat{\sigma}_i^2 = T^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_{it}^2$, $\mathbf{Z}_i = \mathbf{M}_{\hat{F}}\mathbf{X}_i - N^{-1}\sum_{k=1}^{N}\left[\hat{\boldsymbol{\gamma}}'_i\left(\hat{\mathbf{L}}'\hat{\mathbf{L}}/N\right)^{-1}\hat{\boldsymbol{\gamma}}_k\right]\mathbf{M}_{\hat{F}}\mathbf{X}_k$, 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 $\hat{\mathbf{V}}_i = N^{-1}\sum_{j=1}^{N}\hat{\boldsymbol{\gamma}}'_i\left(\hat{\mathbf{L}}'\hat{\mathbf{L}}/N\right)^{-1}\hat{\boldsymbol{\gamma}}_j\mathbf{X}_j$. The selection of the number of strong common factors (m_f) in the Kapetanios and Pesaran (2007) and in the Bai (2009) estimators has been based on Bai and Ng (2002) IC_{p1} criterium.

5.1 Results

Results on the estimation of the slope parameters for the Experiments A and B are summarized in Tables 1-5. 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 $H_1 : \beta_1 = 0.95$.

We do not report results for the FE estimator since they show that, as expected, this 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_n. Tables 1-2 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, $\{n_{\ell t},\ \ell = 1, 2, ..., m_n\}$, are weak or semi-strong, and the choice of m_n. Further, we notice that the power of the tests based on CCE estimators is not affected by m_n, the number of weak (or semi-strong) 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_n.[6] The Monte Carlo results clearly show that augmenting the regression with cross 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 3-4 report the findings for the MGPC and PPC. First notice that these estimators, since they estimate the unobserved common factors by principal components, only work in the case where the factors, $\{n_{\ell t}\}$, represent a set of weak factors, or when $m_n = 0$ (i.e., in Experiment A). In fact, in the case of a semi-strong factor structure the covariance matrix of the idiosyncratic error would not have bounded column norm, a condition required by the principal components analysis for consistent estimation of the factors and their loadings. However, as shown in Table 1, even for Experiment A, these estimators show some size 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-strong factor structure is considered.

Finally, Table 5 gives the results for the Bai (2009) PC iterative estimator. The bias and RMSE of the Bai estimators are comparable to CCE type estimators, but tests based on them are grossly over-sized, even when $m_n = 0$. The problem seems to lie with the variance of the Bai estimator, 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 proposed estimator.

6 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-strong factors. Under this framework, we have shown that the CCE method still yields consistent estimates of the mean of the slope coefficients and the asymptotic normal theory continues to be applicable.

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

Table 1: Results for CCEMG estimator.

		Bias (x100)					**RMSE (x100)**					**Size (x100)**					**Power (x100)**				
m_n	N/T	20	30	50	100	200	20	30	50	100	200	20	30	50	100	200	20	30	50	100	200
		Experiment A: $m_f = 3$ strong factors and m_n weak factors.																			
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
		Experiment B: $m_f = 3$ strong factors and m_n semi-weak factors.																			
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 2: Results for CCEP estimator.

		Bias (x100)					**RMSE (x100)**					**Size (x100)**					**Power (x100)**				
m_n	N/T	20	30	50	100	200	20	30	50	100	200	20	30	50	100	200	20	30	50	100	200
		Experiment A: $m_f = 3$ strong factors and m_n weak factors.																			
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
		Experiment B: $m_f = 3$ strong factors and m_n semi-strong factors.																			
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 3: Results for MGPC estimator.

		Bias (x100)					**RMSE (x100)**					**Size (x100)**					**Power (x100)**				
m_n	N/T	20	30	50	100	200	20	30	50	100	200	20	30	50	100	200	20	30	50	100	200
		Experiment A: $m_f = 3$ strong factors and m_n weak factors.																			
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
		Experiment B: $m_f = 3$ strong factors and m_n semi-strong factors.																			
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 4: Results for PPC estimator.

		Bias (x100)					**RMSE (x100)**					**Size (x100)**					**Power (x100)**				
m_n	N/T	20	30	50	100	200	20	30	50	100	200	20	30	50	100	200	20	30	50	100	200
		Experiment A: $m_f = 3$ strong factors and m_n weak factors.																			
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
		Experiment B: $m_f = 3$ strong factors and m_n semi-strong factors.																			
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

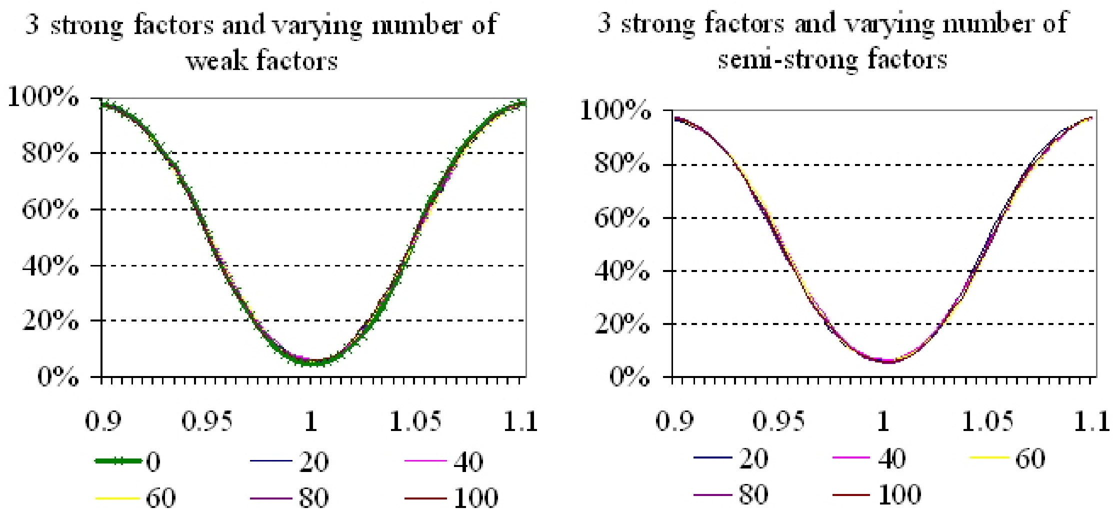

Figure 1: Power curves for the CCEP t-tests in experiments with $N = 100$, $T = 100$, 3 strong factors, and a varying number m_n of weak factors (left chart) and semi-strong factors (right chart).

Table 5: Results for Bai estimator. Experiment A and B: $m_f = 3$ strong factors and m_n weak or semi-strong factors.[7]

		Bias (x100)		**RMSE (x100)**		**Size (x100)**		**Power (x100)**	
m_n	N/T	20	100	20	100	20	100	20	100
		Weak factor structure $\{\boldsymbol{\lambda}_i'\mathbf{n}_t\}$							
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-strong factor structure $\{\boldsymbol{\lambda}_i'\mathbf{n}_t\}$							
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

[7] 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.

Andrews, D. W. K. (1987). Asymptotic results for generalized Wald tests. *Econometric Theory 3*, 348–358.

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.

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

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.

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.

Davidson, J. (1994). *Stochastic Limit Theory*. Oxford University Press.

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.

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 (2010). Panels with nonstationary multifactor error structures. CESifo Working Paper No. 1788, August 2006, revised April 2010.

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. (2009). Asymptotics of the principal components estimator of large factor models with weak factors. Economics Department, Columbia University.

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 (2010). Large panels with common factors and spatial correlations. CESifo Working Paper No. 2103, September 2007, revised January 2010.

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

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.

A Statements and proofs of lemmas

We state and prove a number of lemmas that we shall use in proofs of Theorems 2 and 3.

Lemma A.1 *Suppose Assumptions 5-9 hold and* $(N,T) \xrightarrow{j} \infty$. *Then,*

$$\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}_t\overline{\vartheta}_{wt} \xrightarrow{L_1} \mathbf{0}, \quad \frac{\sqrt{N}}{T}\sum_{t=1}^{T}\mathbf{g}_t\overline{\varepsilon}_{wt} \xrightarrow{L_1} \mathbf{0}, \tag{A.1}$$

$$\frac{\sqrt{N}}{T}\sum_{t=1}^{T}\mathbf{g}_t\overline{\mathbf{v}}'_{wt} \xrightarrow{L_1} \mathbf{0}, \quad \frac{1}{T}\sum_{t=1}^{T}\mathbf{v}_{it}\overline{\vartheta}_{wt} \xrightarrow{L_1} \mathbf{0} \text{ uniformly in } i, \tag{A.2}$$

$$\frac{\sqrt{N}}{T}\sum_{t=1}^{T}\mathbf{v}_{it}\overline{\varepsilon}_{wt} \xrightarrow{L_1} \mathbf{0} \text{ uniformly in } i, \quad \frac{\sqrt{N}}{T}\sum_{t=1}^{T}\mathbf{v}_{it}\overline{\mathbf{v}}'_{wt} \xrightarrow{L_1} \mathbf{0} \text{ uniformly in } i, \tag{A.3}$$

$$\frac{1}{T}\sum_{t=1}^{T}\vartheta_{it}\overline{\vartheta}_{wt} \xrightarrow{L_1} 0 \text{ uniformly in } i, \quad \frac{\sqrt{N}}{T}\sum_{t=1}^{T}\vartheta_{it}\overline{\varepsilon}_{wt} \xrightarrow{L_1} 0 \text{ uniformly in } i, \tag{A.4}$$

$$\frac{\sqrt{N}}{T}\sum_{t=1}^{T}\vartheta_{it}\overline{\mathbf{v}}_{wt} \xrightarrow{L_1} \mathbf{0} \text{ uniformly in } i, \quad \frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}\overline{\vartheta}_{wt} \xrightarrow{L_1} 0, \text{ uniformly in } i \tag{A.5}$$

$$\frac{\sqrt{N}}{T}\sum_{t=1}^{T}\varepsilon_{it}\bar{\varepsilon}_{wt}\overset{L_1}{\to}0 \text{ uniformly in } i, \text{ and } \frac{\sqrt{N}}{T}\sum_{t=1}^{T}\varepsilon_{it}\overline{\mathbf{v}}_{wt}\overset{L_1}{\to}\mathbf{0} \text{ uniformly in } i, \tag{A.6}$$

where $\mathbf{g}_t=(\mathbf{d}_t',\mathbf{f}_t')'$, $\overline{\vartheta}_{wt}=\sum_{i=1}^{N}w_i\vartheta_{it}$, $\vartheta_{it}=\sum_{\ell=1}^{m_n(N)}\lambda_{i\ell}n_{\ell t}$, $\bar{\varepsilon}_{wt}=\sum_{i=1}^{N}w_i\varepsilon_{it}$, *and* $\overline{\mathbf{v}}_{wt}=\sum_{\ell=1}^{N}w_i\mathbf{v}_{it}$. *If in addition there exist constants* α *and* K *such that* $0\leq\alpha<1$ *and conditions (32) and (33) hold, then*

$$\frac{1}{T}\sum_{t=1}^{T}\overline{\mathbf{u}}_{wt}\overline{\mathbf{u}}_{wt}'\overset{L_1}{\to}\mathbf{0}, \tag{A.7}$$

where $\overline{\mathbf{u}}_{wt}=\sum_{i=1}^{N}w_i\mathbf{u}_{it}$, *and* $\mathbf{u}_{it}$ *is defined by (24). If conditions (36) and (37) hold instead of conditions (32) and (33),* $0\leq\alpha<1/2$, *and the remaining assumptions are unchanged, then*

$$\frac{\sqrt{N}}{T}\sum_{t=1}^{T}\mathbf{g}_t\overline{\vartheta}_{wt}\overset{L_1}{\to}\mathbf{0},\ \frac{\sqrt{N}}{T}\sum_{t=1}^{T}\mathbf{v}_{it}\overline{\vartheta}_{wt}\overset{L_1}{\to}\mathbf{0} \text{ uniformly in } i, \tag{A.8}$$

$$\frac{\sqrt{N}}{T}\sum_{t=1}^{T}\vartheta_{it}\overline{\vartheta}_{wt}\overset{L_1}{\to}0 \text{ uniformly in } i, \text{ and } \frac{\sqrt{N}}{T}\sum_{t=1}^{T}\varepsilon_{it}\overline{\vartheta}_{wt}\overset{L_1}{\to}0 \text{ uniformly in } i. \tag{A.9}$$

Proof. We use L_1 mixingale weak law to establish results (A.1)-(A.9). Let $T_N=T(N)$ such that $T_N\to\infty$ as $N\to\infty$ and let $c_{Nt}=\frac{1}{T_N}$ for all $N\in\mathbb{N}$, and all $t\in\mathbb{Z}$. To establish the first part of (A.1) define

$$\boldsymbol{\kappa}_{Nt}=\frac{1}{T_N}\mathbf{g}_t\overline{\vartheta}_{wt}=\frac{1}{T_N}\mathbf{g}_t\sum_{\ell=1}^{m_n(N)}\overline{\lambda}_{w\ell}n_{\ell t},$$

where $\overline{\lambda}_{w\ell}=\sum_{i=1}^{N}w_i\lambda_{i\ell}$. We have

$$E\left(\frac{\boldsymbol{\kappa}_{Nt}\boldsymbol{\kappa}_{Nt}'}{c_{Nt}^2}\right)=\boldsymbol{\Sigma}_g\sum_{\ell=1}^{m_n(N)}\overline{\lambda}_{w\ell}^2,$$

where $\|\boldsymbol{\Sigma}_g\|<K$ by Assumption 5. Consider the term $\sum_{\ell=1}^{m_n(N)}\overline{\lambda}_{w\ell}^2$. Since absolute summability implies square summability, a sufficient condition for the existence of an upper bound for $\sum_{\ell=1}^{m_n(N)}\overline{\lambda}_{w\ell}^2$ is the existence of an upper bound for $\sum_{\ell=1}^{m_n(N)}\left|\overline{\lambda}_{w\ell}\right|$. But

$$\sum_{\ell=1}^{m_n(N)}\left|\overline{\lambda}_{w\ell}\right|=\sum_{\ell=1}^{m_n(N)}\left|\sum_{i=1}^{N}w_i\lambda_{i\ell}\right|\leq\sum_{i=1}^{N}|w_i|\left(\sum_{\ell=1}^{m_n(N)}|\lambda_{i\ell}|\right)<K,$$

where $\sum_{\ell=1}^{m_n(N)}|\lambda_{i\ell}|<K$ by condition (29) of Assumption 6, and $\sum_{i=1}^{N}|w_i|$ is bounded by (1)-(2). It follows that array $\{\boldsymbol{\kappa}_{Nt}/c_{Nt}\}$ is uniformly bounded in L_2-norm and therefore uniformly integrable.[8] Furthermore, $\mathbf{g}_t$ and $n_{\ell t}$, for $\ell=1,2,...,m_n(N)$, are covariance stationary processes with absolute summable autocovariances, and therefore $\|E(\mathbf{g}_t\mid\mathcal{I}_{t-s})\|_{L_1}\to0$ and $\|E(n_{\ell t}\mid\mathcal{I}_{t-s})\|_{L_1}\to0$, as $s\to\infty$, and array $\{\boldsymbol{\kappa}_{Nt}\}$ is uniformly integrable L_1-mixingale with respect to the constant array $\{c_{Nt}\}$. Since $\lim_{N\to\infty}\sum_{t=1}^{T_N}c_{Nt}=\lim_{N\to\infty}\sum_{t=1}^{T_N}T_N^{-1}=1<\infty$, and $\lim_{N\to\infty}\sum_{t=1}^{T_N}c_{Nt}^2=\lim_{N\to\infty}\sum_{t=1}^{T_N}T_N^{-2}=0$, a mixingale weak law (Davidson (1994), Theorem 19.11) can

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

be applied, and we have

$$\sum_{t=1}^{T} \boldsymbol{\kappa}_{Nt} = \frac{1}{T_N} \sum_{t=1}^{T} \mathbf{g}_t \overline{\vartheta}_{wt} \overset{L_1}{\rightarrow} \mathbf{0},$$

as required. Similarly, to establish the first part of (A.8) define

$$\boldsymbol{\kappa}_{Nt} = \frac{\sqrt{N}}{T_N} \mathbf{g}_t \overline{\vartheta}_{wt} = \frac{\sqrt{N}}{T_N} \mathbf{g}_t \sum_{\ell=1}^{m_n(N)} \overline{\lambda}_{w\ell} n_{\ell t},$$

where as before $\overline{\lambda}_{w\ell} = \sum_{i=1}^{N} w_i \lambda_{i\ell}$. Hence

$$E\left(\frac{\boldsymbol{\kappa}_{Nt}\boldsymbol{\kappa}_{Nt}'}{c_{Nt}^2}\right) = N \ \boldsymbol{\Sigma}_g \sum_{\ell=1}^{m_n(N)} \overline{\lambda}_{w\ell}^2,$$

and note that

$$\sum_{\ell=1}^{m_n(N)} \overline{\lambda}_{w\ell}^2 < K \ N^{2\alpha-2} \ m_n\left(N\right) = O\left(N^{-1}\right),$$

under conditions (36) and (37). Thus, same as before $\{\boldsymbol{\kappa}_{Nt}\}$ is uniformly integrable L_1-mixingale with respect to a constant array c_{Nt}, and applying a mixingale weak law yields

$$\sum_{t=1}^{T} \boldsymbol{\kappa}_{Nt} = \frac{\sqrt{N}}{T_N} \sum_{t=1}^{T} \mathbf{g}_t \overline{\vartheta}_{wt} \overset{L_1}{\rightarrow} \mathbf{0}.$$

Remaining results can also be established in a similarly way. For example, in order to establish the second part of (A.1) define $\boldsymbol{\kappa}_{Nt} = \frac{\sqrt{N}}{T_N} \mathbf{g}_t \overline{\varepsilon}_{wt}$, and note that

$$E\left(\frac{\boldsymbol{\kappa}_{Nt}\boldsymbol{\kappa}_{Nt}'}{c_{Nt}^2}\right) = N \ \boldsymbol{\Sigma}_g E\left(\overline{\varepsilon}_{wt}^2\right) = N \ \boldsymbol{\Sigma}_g \sum_{i=1}^{N} w_i^2 E\left(\varepsilon_{it}^2\right) < K.$$

Similarly, in order to establish the first part of (A.2), define $\boldsymbol{\kappa}_{Nt} = \frac{\sqrt{N}}{T_N} \mathbf{g}_t \overline{\mathbf{v}}_{wt}$, and note that

$$E\left(\frac{\boldsymbol{\kappa}_{Nt}\boldsymbol{\kappa}_{Nt}'}{c_{Nt}^2}\right) = N\boldsymbol{\Sigma}_g E\left(\overline{\mathbf{v}}_{wt}^2\right) = N\boldsymbol{\Sigma}_g \sum_{i=1}^{N} w_i^2 \left\|\boldsymbol{\Sigma}_{vi}\right\| < K.$$

The following four results, the second part of (A.3), the first part of (A.6), (A.7), and the first part of (A.9), deserve more attention. In the case of the second part of (A.3), we have

$$\frac{\sqrt{N}}{T_N} \sum_{t=1}^{T_N} \mathbf{v}_{it} \overline{\mathbf{v}}_{wt}' = \frac{\sqrt{N}}{T_N} \sum_{t=1}^{T_N} w_i \mathbf{v}_{it} \mathbf{v}_{it}' + \frac{\sqrt{N}}{T_N} \sum_{t=1}^{T_N} \mathbf{v}_{it} \sum_{j \neq i} w_j \mathbf{v}_{jt}. \tag{A.10}$$

Note that since $\mathbf{v}_{it}$ is ergodic in variance then $T_N^{-1} \sum_{t=1}^{T_N} \mathbf{v}_{it}\mathbf{v}_{it}' \overset{L_1}{\rightarrow} \boldsymbol{\Sigma}_{vi}$, and since $\sqrt{N} w_i \rightarrow 0$ as $N \rightarrow \infty$ and $\sup_i \left\|\boldsymbol{\Sigma}_{vi}\right\| < K$, it follows that $\frac{\sqrt{N}}{T_N} \sum_{t=1}^{T_N} w_i \mathbf{v}_{it}\mathbf{v}_{it}' \overset{L_1}{\rightarrow} \mathbf{0}$. To establish convergence of the second term on the right side of (A.10), define $\boldsymbol{\kappa}_{Nt}=\frac{\sqrt{N}}{T_N}\mathbf{v}_{it}\sum_{j\neq i}w_j\mathbf{v}_{jt}$, and note that

$$\left\|E\left(\frac{\boldsymbol{\kappa}_{Nt}\boldsymbol{\kappa}_{Nt}'}{c_{Nt}^2}\right)\right\|<N\left\|\boldsymbol{\Sigma}_{vi}\right\|\sum_{j\neq i}w_j^2\left\|\boldsymbol{\Sigma}_{vj}\right\|<K.$$

Using now the same arguments as in the proof of the first part of (A.1) we have $\sum_{t=1}^{T_N}\boldsymbol{\kappa}_{Nt}\xrightarrow{L_1}\mathbf{0}$, which completes the proof of the second part of (A.3). The first part of (A.6) can be established along the same lines followed to prove the second part of (A.3). To establish the first part of (A.9), consider

$$\kappa_{Nt}=\frac{\sqrt{N}}{T_N}\vartheta_{it}\overline{\vartheta}_{wt}=\frac{\sqrt{N}}{T_N}\vartheta_{it}\sum_{j=1}^{N}w_j\vartheta_{jt}.$$

We have

$$\left\|E\left(\frac{\kappa_{Nt}^2}{c_{Nt}^2}\right)\right\|=N\sum_{j=1}^{N}\sum_{k=1}^{N}w_jw_kE\left(\vartheta_{it}^2\vartheta_{jt}\vartheta_{kt}\right), \tag{A.11}$$

where

$$E\left(\vartheta_{it}^2\vartheta_{jt}\vartheta_{kt}\right)=\sum_{\ell_1=1}^{m_n(N)}\sum_{\ell_2=1}^{m_n(N)}\sum_{\ell_3=1}^{m_n(N)}\sum_{\ell_4=1}^{m_n(N)}\lambda_{i\ell_1}\lambda_{i\ell_2}\lambda_{j\ell_3}\lambda_{k\ell_4}E\left(n_{\ell_1t}n_{\ell_2t}n_{\ell_3t}n_{\ell_4t}\right),$$

in which $E\left(n_{\ell_1t}n_{\ell_2t}n_{\ell_3t}n_{\ell_4t}\right)$ is nonzero only in the following three cases: (i) $\ell_1=\ell_2=\ell_3=\ell_4$, (ii) $\ell_1=\ell_2$ and $\ell_3=\ell_4$, and (iii) $\ell_1=\ell_3$ and $\ell_2=\ell_4$. It follows that

$$E\left(\vartheta_{it}^2\vartheta_{jt}\vartheta_{kt}\right)=\sum_{\ell=1}^{m_n(N)}\lambda_{i\ell}^2\lambda_{j\ell}\lambda_{k\ell}E\left(n_{\ell t}^4\right)+\sum_{\ell_1=1}^{m_n(N)}\sum_{\ell_3\neq\ell_1}\lambda_{i\ell_1}^2\lambda_{j\ell_3}\lambda_{k\ell_3}+\sum_{\ell_1=1}^{m_n(N)}\sum_{\ell_2\neq\ell_1}\lambda_{i\ell_1}\lambda_{i\ell_2}\lambda_{j\ell_1}\lambda_{k\ell_2}, \tag{A.12}$$

where $E\left(n_{\ell t}^2\right)=1$, and $E\left(n_{\ell t}^4\right)<K$ by Assumption 5. Using conditions (36) and (37), and the absolute summability condition (29) of Assumption 6, we obtain

$$N\sum_{j=1}^{N}\sum_{k=1}^{N}w_jw_k\left(\sum_{\ell=1}^{m_n(N)}\lambda_{i\ell}^2\lambda_{j\ell}\lambda_{k\ell}E\left(n_{\ell t}^4\right)\right)=N\sum_{\ell=1}^{m_n(N)}\lambda_{i\ell}^2\overline{\lambda}_{w\ell}^2<K, \tag{A.13}$$

$$N\sum_{j=1}^{N}\sum_{k=1}^{N}w_jw_k\left(\sum_{\ell_1=1}^{m_n(N)}\sum_{\ell_3\neq\ell_1}\lambda_{i\ell_1}^2\lambda_{j\ell_3}\lambda_{k\ell_3}\right)=N\sum_{\ell_1=1}^{m_n(N)}\lambda_{i\ell_1}^2\sum_{\ell_3\neq\ell_1}\overline{\lambda}_{w\ell_3}^2<K$$

$$N\sum_{j=1}^{N}\sum_{k=1}^{N}w_jw_k\left(\sum_{\ell_1=1}^{m_n(N)}\sum_{\ell_2\neq\ell_1}\lambda_{i\ell_1}\lambda_{i\ell_2}\lambda_{j\ell_1}\lambda_{k\ell_2}\right)=N\sum_{\ell_1=1}^{m_n(N)}\lambda_{i\ell_1}\overline{\lambda}_{w\ell_1}\sum_{\ell_2\neq\ell_1}^{m_n(N)}\lambda_{i\ell_2}\overline{\lambda}_{w\ell_2}<K \tag{A.14}$$

Now substitute (A.12) in (A.11) for $E\left(\vartheta_{it}^2\vartheta_{jt}\vartheta_{kt}\right)$, and use (A.13)-(A.14) to obtain

$$\left\|E\left(\frac{\kappa_{Nt}^2}{c_{Nt}^2}\right)\right\|<K. \tag{A.15}$$

Using the same arguments as in the proof of the first part of (A.1), κ_{Nt} is uniformly integrable L_1-mixingale with respect to the constant array c_{Nt}, and applying a mixingale weak law yields $\sum_{t=1}^{T_N}\kappa_{Nt}\xrightarrow{L_1}0$, as required.

In order to establish (A.7) note that

$$\overline{\mathbf{u}}_{wt} = \begin{pmatrix} \overline{\vartheta}_{wt} + \overline{\varepsilon}_{wt} + \sum_{i=1}^{N} w_i \boldsymbol{\beta}_i' \mathbf{v}_{it} \\ \overline{\mathbf{v}}_{wt} \end{pmatrix}.$$

Convergence of $T^{-1}\sum_{t=1}^{T} \overline{\vartheta}_{wt}^2$ can also be established using a mixingale weak law. Let $\kappa_{Nt} = T_N^{-1}\overline{\vartheta}_{wt}^2$, and note that

$$\begin{aligned} E\left(\overline{\vartheta}_{wt}^2\right) &= E\left(\sum_{j=1}^{N} w_j \vartheta_{jt}\right)^2 = \sum_{i=1}^{N}\sum_{j=1}^{N} w_i w_j E\left(\vartheta_{it}\vartheta_{jt}\right) \\ &= \sum_{i=1}^{N}\sum_{j=1}^{N} w_i w_j \sum_{\ell=1}^{m_n(N)} \lambda_{i\ell}\lambda_{j\ell} \\ &\leq \sum_{\ell=1}^{m_n(N)} \left(\sum_{i=1}^{N} w_i \lambda_{i\ell} \cdot \sum_{j=1}^{N} w_j \lambda_{j\ell}\right) \\ &\leq K \cdot m_n(N) N^{2\alpha-2} \to 0, \end{aligned}$$

where $|w_i| < K/N$ under the granularity conditions (1)-(2), $\sum_{i=1}^{N} |\lambda_{i\ell}| < KN^{\alpha}$ by (32), and $m_n(N) N^{2\alpha-2} \to 0$ by (33). Similarly as in the proof of the first part of (A.9), it can be shown that $E\left(\kappa_{Nt}^2/c_{Nt}^2\right)$ is bounded and that $\sum_{t=1}^{T_N} \kappa_{Nt} \xrightarrow{L_1} 0$. The convergence of the remaining elements of (A.7) can be established using similar arguments as in Lemma 2 of Pesaran (2006), or by applying a mixingale weak law. ■

Lemma A.2 *Suppose Assumptions 5-9 hold and $(N,T) \xrightarrow{j} \infty$. Then,*

$$\frac{\mathbf{V}_i'\mathbf{V}_i}{T} \xrightarrow{p} \boldsymbol{\Sigma}_{vi} \text{ uniformly in } i,\ \frac{\mathbf{V}_i'\mathbf{G}}{T} \xrightarrow{p} \mathbf{0} \text{ uniformly in } i, \tag{A.16}$$

$$\frac{\mathbf{V}_i'\mathbf{Q}}{T} \xrightarrow{p} \mathbf{0} \text{ uniformly in } i,\ \frac{\mathbf{G}'\mathbf{G}}{T} \xrightarrow{p} \boldsymbol{\Sigma}_g, \tag{A.17}$$

$$\frac{\mathbf{Q}'\mathbf{G}}{T} = O_p(1),\ \frac{\mathbf{Q}'\mathbf{Q}}{T} = O_p(1) \tag{A.18}$$

$$\overline{\boldsymbol{\Pi}}_i^* - \boldsymbol{\Pi}_i^* \xrightarrow{p} \mathbf{0} \text{ uniformly in } i,\ \frac{\overline{\mathbf{H}}_w'\overline{\mathbf{H}}_w}{T} = O_p(1), \tag{A.19}$$

$$\frac{\overline{\mathbf{H}}_w'\boldsymbol{\vartheta}_i}{T} = o_p(1) \text{ uniformly in } i,\ \frac{\mathbf{X}_i'\mathbf{Q}}{T} = O_p(1) \text{ uniformly in } i, \tag{A.20}$$

$$\frac{\overline{\mathbf{H}}_w'\boldsymbol{\varepsilon}_i}{T} = o_p(1) \text{ uniformly in } i,\ \frac{\overline{\mathbf{H}}_w'\mathbf{X}_i}{T} = O_p(1) \text{ uniformly in } i \tag{A.21}$$

and

$$\frac{\overline{\mathbf{H}}_w'\mathbf{F}}{T} = O_p(1), \tag{A.22}$$

where $\boldsymbol{\Pi}_i^* = \left[\mathbf{I} - \mathbf{P}\left(\mathbf{P}'\mathbf{P}\right)^{+}\mathbf{P}'\right]\boldsymbol{\Pi}_i$, $\overline{\boldsymbol{\Pi}}_i^* = \left[\mathbf{I} - \overline{\mathbf{P}}_w\left(\overline{\mathbf{P}}_w'\overline{\mathbf{P}}_w\right)^{+}\overline{\mathbf{P}}_w'\right]\boldsymbol{\Pi}_i$, $\overline{\mathbf{P}}_w$ *is defined by (26),* $\mathbf{P} = E(\overline{\mathbf{P}}_w)$,

$\mathbf{G}=(\mathbf{D},\mathbf{F})$, $\mathbf{Q}=\mathbf{G}\overline{\mathbf{P}}_w$, $\overline{\mathbf{H}}_w=\left(\mathbf{D},\overline{\mathbf{Z}}_w\right)$, *and* $\boldsymbol{\vartheta}_i=(\vartheta_{i1},\vartheta_{i1},...,\vartheta_{iT})'$ *with* $\vartheta_{it}=\sum_{\ell=1}^{m_n(N)}\lambda_{i\ell}n_{\ell t}$.

Proof. The first part of (A.16) follows directly by observing that the covariance stationary process $\mathbf{v}_{it}$ is ergodic in variance. Since $\mathbf{g}_t=(\mathbf{d}_t',\mathbf{f}_t')$ is also a covariance stationary process with absolute summable autocovariances, it follows that

$$\frac{1}{T}\sum_{t=1}^{T}\mathbf{v}_{it}\mathbf{g}_t'\xrightarrow{p}E\left(\mathbf{v}_{it}\mathbf{g}_t'\right)=\mathbf{0},$$

where the convergence is uniform in i since the second moments of $\mathbf{v}_{it}$ are uniformly bounded in i. This establishes the second part of (A.16). The first part of (A.17) can be established using the same arguments. The second part of (A.17) can be established similarly to the first part of (A.16) by noting that $\boldsymbol{\Sigma}_g=E\left(\mathbf{g}_t\mathbf{g}_t'\right)$. In the same spirit,

$$\frac{1}{T}\sum_{t=1}^{T}\mathbf{q}_t\mathbf{g}_t'=\frac{1}{T}\sum_{t=1}^{T}\overline{\mathbf{P}}_w'\mathbf{g}_t\mathbf{g}_t'\xrightarrow{p}E\left(\overline{\mathbf{P}}_w'\mathbf{g}_t\mathbf{g}_t'\right),$$

as $(N,T)\xrightarrow{j}\infty$. But $E\left(\overline{\mathbf{P}}_w'\mathbf{g}_t\mathbf{g}_t'\right)=\mathbf{P}'\boldsymbol{\Sigma}_g$, and $\left\|\mathbf{P}'\boldsymbol{\Sigma}_g\right\|\leq\left\|\mathbf{P}\right\|\left\|\boldsymbol{\Sigma}_g\right\|<K$, where $\left\|\boldsymbol{\Sigma}_g\right\|<K$ by Assumption 5 and $\left\|\mathbf{P}\right\|<K$ by Assumptions 6, 8, and 9, which completes the proof of the first part of (A.18). Noting that $\mathbf{Q}=\mathbf{G}\overline{\mathbf{P}}_w$, and that $\overline{\mathbf{P}}_w\xrightarrow{p}\mathbf{P}$, the second part of (A.17) implies

$$\frac{\mathbf{Q}'\mathbf{Q}}{T}-\mathbf{P}'\boldsymbol{\Sigma}_g\mathbf{P}\xrightarrow{p}\mathbf{0},$$

as $(N,T)\xrightarrow{j}\infty$. But, same as before, $\left\|\mathbf{P}'\boldsymbol{\Sigma}_g\mathbf{P}\right\|\leq\left\|\mathbf{P}\right\|^2\left\|\boldsymbol{\Sigma}_g\right\|<K$ and it follows that $\mathbf{Q}'\mathbf{Q}/T=O_p(1)$, as required. To establish the first part of (A.19) note that $\overline{\mathbf{P}}_w-\mathbf{P}\xrightarrow{p}\mathbf{0}$ as $(N,T)\xrightarrow{j}\infty$, and

$$\lim_{N\rightarrow\infty}\Pr\left[rank\left(\overline{\mathbf{P}}_w'\overline{\mathbf{P}}_w\right)=rank\left(\mathbf{P}'\mathbf{P}\right)\right]=1.$$

It follows, using also Theorem 2 of Andrews (1987), that $\overline{\boldsymbol{\Pi}}_i^*-\boldsymbol{\Pi}_i^*\xrightarrow{p}\mathbf{0}$. The remaining results can be established in a similar way, as results (A.16)-(A.18), using ergodicity in mean and variance of covariance stationary series with absolute summable autocovariances and Lemma A.1. ■

Lemma A.3 *Suppose Assumptions 5-9 hold,* $(N,T)\xrightarrow{j}\infty$, *and there exist constants* α *and* K *such that* $0\leq\alpha<1$, *and conditions (32) and (33) hold. Then,*

$$\frac{\mathbf{X}_i'\left(\overline{\mathbf{H}}_w-\mathbf{Q}\right)}{T}\xrightarrow{p}\mathbf{0}\text{ uniformly in }i,\ \frac{\left(\mathbf{Q}'-\overline{\mathbf{H}}_w'\right)\boldsymbol{\vartheta}_i}{T}\xrightarrow{p}\mathbf{0}\text{ uniformly in }i,\tag{A.23}$$

$$\left[\left(\frac{\overline{\mathbf{H}}_w'\overline{\mathbf{H}}_w}{T}\right)^+-\left(\frac{\mathbf{Q}'\mathbf{Q}}{T}\right)^+\right]\xrightarrow{p}\mathbf{0},\tag{A.24}$$

$$\frac{\left(\mathbf{Q}'-\overline{\mathbf{H}}_w'\right)\boldsymbol{\varepsilon}_i}{T}\xrightarrow{p}\mathbf{0}\text{ uniformly in }i,\text{ and }\frac{\left(\mathbf{Q}'-\overline{\mathbf{H}}_w'\right)\mathbf{F}}{T}\xrightarrow{p}\mathbf{0}.\tag{A.25}$$

If conditions (36) and (37) hold instead of conditions (32) and (33), $0\leq\alpha<1/2$, *and the remaining assumptions* *are unchanged, then*

$$\frac{\sqrt{N}\mathbf{X}_i'\left(\overline{\mathbf{H}}_w-\mathbf{Q}\right)}{T}\xrightarrow{p}\mathbf{0}\text{ uniformly in } i,\ \frac{\sqrt{N}\left(\mathbf{Q}'-\overline{\mathbf{H}}_w'\right)\boldsymbol{\vartheta}_i}{T}\xrightarrow{p}\mathbf{0}\text{ uniformly in } i, \tag{A.26}$$

$$\sqrt{N}\left[\left(\frac{\overline{\mathbf{H}}_w'\overline{\mathbf{H}}_w}{T}\right)^+-\left(\frac{\mathbf{Q}'\mathbf{Q}}{T}\right)^+\right]\xrightarrow{p}\mathbf{0}, \tag{A.27}$$

$$\frac{\sqrt{N}\left(\mathbf{Q}'-\overline{\mathbf{H}}_w'\right)\boldsymbol{\varepsilon}_i}{T}\xrightarrow{p}\mathbf{0}\text{ uniformly in } i,\text{ and } \frac{\sqrt{N}\left(\mathbf{Q}'-\overline{\mathbf{H}}_w'\right)\mathbf{F}}{T}\xrightarrow{p}\mathbf{0}. \tag{A.28}$$

Proof. Using the notations in Section (4) we note that $\overline{\mathbf{H}}_w=\mathbf{Q}+\overline{\mathbf{U}}_w^*$, where $\overline{\mathbf{U}}^*=\left(\mathbf{0},\overline{\mathbf{U}}_w\right)$, $\overline{\mathbf{U}}_w=\sum_{i=1}^N w_i\mathbf{U}_i$. Also recall that $\mathbf{X}_i=\mathbf{G}\boldsymbol{\Pi}_i+\mathbf{V}_i$. Hence both parts of (A.23) and (A.25) directly follow from results (A.1)-(A.3) of Lemma A.1. However, because Moore-Penrose inverse is not a continuous function it is not sufficient that

$$\left(\frac{\overline{\mathbf{H}}_w'\overline{\mathbf{H}}_w}{T}\right)-\left(\frac{\mathbf{Q}'\mathbf{Q}}{T}\right)=o_p\left(1\right), \tag{A.29}$$

for (A.24) to hold. We establish (A.24) in a similarly way as Kapetanios, Pesaran, and Yagamata (2010). By Theorem 2 of Andrews (1987), (A.29) is sufficient for (A.24), if additionally, as $(N,T)\xrightarrow{j}\infty$,

$$\lim_{(N,T)\xrightarrow{j}\infty}\Pr\left[rank\left(\frac{\overline{\mathbf{H}}_w'\overline{\mathbf{H}}_w}{T}\right)=rank\left(\frac{\mathbf{Q}'\mathbf{Q}}{T}\right)\right]=1, \tag{A.30}$$

where $rank\left(\mathbf{A}\right)$ denotes rank of $\mathbf{A}$. But

$$\frac{\overline{\mathbf{H}}_w'\overline{\mathbf{H}}_w}{T}=\frac{\mathbf{Q}'\mathbf{Q}}{T}+\frac{\mathbf{Q}'\overline{\mathbf{U}}_w^*}{T}+\frac{\overline{\mathbf{U}}_w^{*\prime}\mathbf{Q}}{T}+\frac{\overline{\mathbf{U}}_w^{*\prime}\overline{\mathbf{U}}_w^*}{T},$$

where

$$\lim_{(N,T)\xrightarrow{j}\infty}\Pr\left(\left\|\frac{\mathbf{Q}'\overline{\mathbf{U}}_w^*}{T}+\frac{\overline{\mathbf{U}}_w^{*\prime}\mathbf{Q}}{T}+\frac{\overline{\mathbf{U}}_w^{*\prime}\overline{\mathbf{U}}_w^*}{T}\right\|>\epsilon\right)=0$$

for all $\epsilon>0$. Also

$$rank\left(\frac{\mathbf{Q}'\mathbf{Q}}{T}\right)=m_d+rank\left(\overline{\mathbf{C}}_w\right),$$

for all N and T, with $rank\left(\mathbf{Q}'\mathbf{Q}/T\right)\to m_d+rank\left(\mathbf{C}\right)\leq m_d+m_f$, as $(N,T)\xrightarrow{j}\infty$. Using these results, it is now easily seen that condition (A.30) in fact holds. Hence, the desired result (A.24) follows.

Results (A.26)-(A.28) can be established in a similar way as results (A.23)-(A.25). ■

Lemma A.4 *Suppose Assumptions 5-10 hold, and $(N,T)\xrightarrow{j}\infty$. Then,*

$$\frac{\mathbf{X}_i'\overline{\mathbf{M}}_q\boldsymbol{\varepsilon}_i}{T}\xrightarrow{p}\mathbf{0}\text{ uniformly in } i, \tag{A.31}$$

and

$$\frac{\mathbf{X}_i'\overline{\mathbf{M}}_q\boldsymbol{\vartheta}_i}{T}\xrightarrow{p}\mathbf{0}\text{ uniformly in } i, \tag{A.32}$$

where $\boldsymbol{\vartheta}_i = (\vartheta_{i1}, \vartheta_{i2}, ..., \vartheta_{iT})'$, *and* $\vartheta_{it} = \sum_{\ell=1}^{m_n(N)} \lambda_{i\ell} n_{\ell t}$.

Proof. Consider

$$\frac{\mathbf{X}_i'\overline{\mathbf{M}}_q\boldsymbol{\vartheta}_i}{T} = \frac{\widetilde{\mathbf{X}}_i'\boldsymbol{\vartheta}_i}{T} = \frac{1}{T}\sum_{t=1}^{T}\widetilde{\mathbf{x}}_{it}\vartheta_{it},$$

where $\widetilde{\mathbf{X}}_i = \overline{\mathbf{M}}_q\mathbf{X}_i$. Let $T_N = T(N)$ be any non-decreasing integer-valued functions of N such that $\lim_{N\to\infty} T_N = \infty$ and define

$$\boldsymbol{\kappa}_{Nt} = \frac{1}{T_N}\widetilde{\mathbf{x}}_{it}\vartheta_{it} = \frac{1}{T_N}\widetilde{\mathbf{x}}_{it}\sum_{\ell=1}^{m_n(N)}\lambda_{i\ell}n_{\ell t}. \tag{A.33}$$

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 $N \in \mathbb{N}$. We have

$$E\left(\frac{\boldsymbol{\kappa}_{Nt}\boldsymbol{\kappa}_{Nt}'}{c_{Nt}^2}\right) = E\left(\widetilde{\mathbf{x}}_{it}\widetilde{\mathbf{x}}_{it}'\vartheta_{it}^2\right) = E\left(\widetilde{\mathbf{x}}_{it}\widetilde{\mathbf{x}}_{it}'\right)E\left(\vartheta_{it}^2\right),$$

where the second equality follow from independence of $\widetilde{\mathbf{x}}_{it}$ and ϑ_{it}. By Assumption 10 there exists a constant $K < \infty$ such that $\sup_i \left\|E\left(\widetilde{\mathbf{x}}_{it}\widetilde{\mathbf{x}}_{it}'\right)\right\| < K$. Further, using independence of factors $n_{\ell t}$ and $n_{\ell' t}$ for any $\ell \neq \ell'$ and noting that $E\left(n_{\ell t}^2\right) = 1$, we have

$$E\left(\vartheta_{it}^2\right) = \sum_{\ell=1}^{m_n(N)}\lambda_{i\ell}^2 < K < \infty.$$

It follows that

$$\left\|E\left(\frac{\boldsymbol{\kappa}_{Nt}\boldsymbol{\kappa}_{Nt}'}{c_{Nt}^2}\right)\right\| < K < \infty. \tag{A.34}$$

(A.34) established that $\{\boldsymbol{\kappa}_{Nt}/c_{Nt}\}$ is uniformly bounded in L_2 norm, which implies uniform integrability. Using similar arguments as in proof of Lemma A.1, $\{\boldsymbol{\kappa}_{Nt}\}$ is L_1-mixingale with respect to the constant array $\{c_{Nt}\}$, and applying a mixingale weak law (Davidson (1994), Theorem 19.11) establishes $\sum_{t=1}^{T_N}\boldsymbol{\kappa}_{Nt} \overset{L_1}{\to} \mathbf{0}$, that is $T^{-1}\sum_{t=1}^{T}\widetilde{\mathbf{x}}_{it}\vartheta_{it} \overset{L_1}{\to} \mathbf{0}$, as $(N,T) \overset{j}{\to} \infty$. This completes the proof of (A.32).

Result (A.31) can be established in a similar way, but this time we need to define $\boldsymbol{\kappa}_{Nt} = T_N^{-1}\widetilde{\mathbf{x}}_{it}\varepsilon_{it}$ and noting that $\sup_i E\left(\varepsilon_{it}^2\right) < K$ by Assumption 7. ■

Lemma A.5 *Suppose Assumptions 5-9 hold and* $(N,T) \overset{j}{\to} \infty$. *Then*

$$\frac{\mathbf{X}_i'\overline{\mathbf{M}}_q\mathbf{X}_i}{T} \overset{p}{\to} \mathbf{\Sigma}_{iq} \text{ uniformly in } i, \tag{A.35}$$

and

$$\frac{\mathbf{X}_i'\overline{\mathbf{M}}_q\mathbf{F}}{T} \overset{p}{\to} \mathbf{Q}_{if} \text{ uniformly in } i, \tag{A.36}$$

where $\mathbf{\Sigma}_{iq}$ *is positive definite and given by*

$$\mathbf{\Sigma}_{iq} = \mathbf{\Sigma}_{vi} + \mathbf{\Pi}_i^{*\prime}\mathbf{\Sigma}_g\mathbf{\Pi}_i^*, \tag{A.37}$$

and

$$\mathbf{Q}_{if} = \mathbf{\Pi}_i^{*\prime}\mathbf{\Sigma}_g\mathbf{S}_f^*, \tag{A.38}$$

in which

$$\mathbf{\Pi}_i^* = \left[\mathbf{I} - \mathbf{P}\left(\mathbf{P}'\mathbf{P}\right)^+\mathbf{P}'\right]\mathbf{\Pi}_i,\ \mathbf{S}_f^* = \left[\mathbf{I} - \mathbf{P}\left(\mathbf{P}'\mathbf{P}\right)^+\mathbf{P}'\right]\mathbf{S}_f, \tag{A.39}$$

$\mathbf{\Pi}_i = (\mathbf{A}'_i, \mathbf{\Gamma}'_i)'$, $\mathbf{S}_f = \left(\mathbf{0}_{m_f \times m_d}, \mathbf{I}_{m_f}\right)'$, *and* $\mathbf{\Sigma}_g = E\left(\mathbf{g}_t \mathbf{g}'_t\right)$.

Proof. Since $\mathbf{X}_i = \mathbf{G}\mathbf{\Pi}_i + \mathbf{V}_i$ then

$$\begin{aligned} \frac{\mathbf{X}'_i \overline{\mathbf{M}}_q \mathbf{X}_i}{T} &= \frac{\mathbf{V}'_i \overline{\mathbf{M}}_q \mathbf{V}_i}{T} + \frac{\mathbf{V}'_i \overline{\mathbf{M}}_q \mathbf{G}\mathbf{\Pi}_i}{T} \\ &\quad + \frac{\mathbf{\Pi}'_i \mathbf{G}' \overline{\mathbf{M}}_q \mathbf{V}_i}{T} + \frac{\mathbf{\Pi}'_i \mathbf{G}' \overline{\mathbf{M}}_q \mathbf{G}\mathbf{\Pi}_i}{T}. \end{aligned} \tag{A.40}$$

Consider the first term and note that,

$$\frac{\mathbf{V}'_i \overline{\mathbf{M}}_q \mathbf{V}_i}{T} = \frac{\mathbf{V}'_i \mathbf{V}_i}{T} - \frac{\mathbf{V}'_i \mathbf{Q}}{T} \left(\frac{\mathbf{Q}'\mathbf{Q}}{T}\right)^+ \frac{\mathbf{Q}'\mathbf{V}_i}{T} \xrightarrow{p} \mathbf{\Sigma}_{vi} \text{ uniformly in } i, \tag{A.41}$$

where the convergence directly follows from Lemma A.2 (the first part of (A.16), the first part of (A.17), and the second part of (A.18)) . Next we examine the second and the third elements (the latter is transpose of the former). We have

$$\frac{\mathbf{V}'_i \overline{\mathbf{M}}_q \mathbf{G}\mathbf{\Pi}_i}{T} = \frac{\mathbf{V}'_i \mathbf{G}}{T}\mathbf{\Pi}_i - \frac{\mathbf{V}'_i \mathbf{Q}}{T} \left(\frac{\mathbf{Q}'\mathbf{Q}}{T}\right)^+ \frac{\mathbf{Q}'\mathbf{G}}{T}\mathbf{\Pi}_i \xrightarrow{p} \mathbf{0} \text{ uniformly in } i, \tag{A.42}$$

where we have used Lemma A.2, in particular the second part of (A.16), the first part of (A.17), and both parts of (A.18). Finally, we examine the last summand on the right side of (A.40). Let $Col\left(\overline{\mathbf{P}}_w\right)$ denote a linear space spanned by the column vectors of $\overline{\mathbf{P}}_w$ and consider the following decomposition of matrix $\mathbf{\Pi}_i$,

$$\mathbf{\Pi}_i = \overline{\mathbf{\Pi}}^*_i + \widetilde{\mathbf{\Pi}}_i, \tag{A.43}$$

where $\widetilde{\mathbf{\Pi}}_i \in Col\left(\overline{\mathbf{P}}_w\right)$, and $\overline{\mathbf{\Pi}}^*_i$ belongs to the orthogonal complement of the space spanned by the column vectors in $\overline{\mathbf{P}}_w$. The decomposition (A.43) is unique. Note that matrix $\overline{\mathbf{M}}_q$ has the property $\overline{\mathbf{M}}_q \mathbf{G}\overline{\mathbf{\Pi}}^*_i = \mathbf{G}\overline{\mathbf{\Pi}}^*_i$ and $\overline{\mathbf{M}}_q \mathbf{G}\widetilde{\mathbf{\Pi}}_i = \mathbf{0}$. It follows that

$$\mathbf{\Pi}'_i \frac{\mathbf{G}'\overline{\mathbf{M}}_q \mathbf{G}}{T}\mathbf{\Pi}_i = \mathbf{\Pi}'_i \frac{\mathbf{G}'\overline{\mathbf{M}}'_q \overline{\mathbf{M}}_q \mathbf{G}}{T}\mathbf{\Pi}_i = \overline{\mathbf{\Pi}}^{*\prime}_i \frac{\mathbf{G}'\mathbf{G}}{T}\overline{\mathbf{\Pi}}^*_i.$$

Using now the second part of (A.17) yields

$$\mathbf{\Pi}'_i \frac{\mathbf{G}'\overline{\mathbf{M}}_q \mathbf{G}}{T}\mathbf{\Pi}_i - \overline{\mathbf{\Pi}}^{*\prime}_i \mathbf{\Sigma}_g \overline{\mathbf{\Pi}}^*_i \xrightarrow{p} \mathbf{0} \text{ uniformly in } i.$$

But according to Lemma A.2, the first part of (A.19), $\overline{\mathbf{\Pi}}^*_i - \mathbf{\Pi}^*_i \xrightarrow{p} \mathbf{0}$, uniformly in i, and therefore

$$\mathbf{\Pi}'_i \frac{\mathbf{G}'\overline{\mathbf{M}}_q \mathbf{G}}{T}\mathbf{\Pi}_i - \mathbf{\Pi}^{*\prime}_i \mathbf{\Sigma}_g \mathbf{\Pi}^*_i \xrightarrow{p} \mathbf{0} \text{ uniformly in } i. \tag{A.44}$$

Using (A.41), (A.42), and (A.44) in (A.40) establishes (A.35), as desired. $\mathbf{\Sigma}_{vi}$ is positive definite by Assumption 7 and matrix $\mathbf{\Sigma}_g = E\left(\mathbf{g}_t \mathbf{g}'_t\right)$ is nonnegative definite. It follows $\mathbf{\Pi}^{*\prime}_i \mathbf{\Sigma}_g \mathbf{\Pi}^*_i$ is nonnegative definite. Sum of positive definite and positive semi-definite matrices is a positive definite matrix and therefore $\mathbf{\Sigma}_{iq} = \mathbf{\Sigma}_{vi} + \mathbf{\Pi}^{*\prime}_i \mathbf{\Sigma}_g \mathbf{\Pi}^*_i$ is positive definite.

Similarly to the proof of result (A.35), consider

$$\begin{aligned} \frac{\mathbf{X}_i'\overline{\mathbf{M}}_q\mathbf{F}}{T} &= \frac{\mathbf{V}_i'\overline{\mathbf{M}}_q\mathbf{F}}{T} + \frac{\mathbf{\Pi}_i'\mathbf{G}'\overline{\mathbf{M}}_q\mathbf{F}}{T}, \\ &= \frac{\mathbf{V}_i'\overline{\mathbf{M}}_q\mathbf{G}\mathbf{S}_f}{T} + \frac{\mathbf{\Pi}_i'\mathbf{G}'\overline{\mathbf{M}}_q\mathbf{G}\mathbf{S}_f}{T} \end{aligned} \tag{A.45}$$

where $\mathbf{F} = \mathbf{G}\mathbf{S}_f$, and $\mathbf{S}_f = \left(\mathbf{0}_{m_f \times m_d}, \mathbf{I}_{m_f}\right)'$ is the corresponding selection matrix. Using similar arguments as in (A.42), and (A.44), we obtain

$$\frac{\mathbf{V}_i'\overline{\mathbf{M}}_q\mathbf{G}\mathbf{S}_f}{T} \xrightarrow{p} \mathbf{0} \text{ uniformly in } i, \tag{A.46}$$

and

$$\frac{\mathbf{\Pi}_i'\mathbf{G}'\overline{\mathbf{M}}_q\mathbf{G}\mathbf{S}_f}{T} - \mathbf{\Pi}_i^{*\prime}\mathbf{\Sigma}_g\mathbf{S}_f^* \xrightarrow{p} \mathbf{0} \text{ uniformly in } i, \tag{A.47}$$

where $\mathbf{\Pi}_i^*$ and $\mathbf{S}_f^*$ is defined by (A.39). Using (A.46), and (A.47) in (A.45) completes the proof of (A.36). ■

Lemma A.6 *Suppose Assumptions 5-9 hold,* $(N,T) \xrightarrow{j} \infty$*, and there exist constants* α *and* K *such that* $0 \leq \alpha < 1/2$ *and conditions (36) and (37) hold. Then,*

$$\sqrt{N}\frac{\mathbf{X}_i'\overline{\mathbf{M}}_w\mathbf{X}_i}{T} - \sqrt{N}\frac{\mathbf{X}_i'\overline{\mathbf{M}}_q\mathbf{X}_i}{T} \xrightarrow{p} \mathbf{0} \textit{ uniformly in } i, \tag{A.48}$$

$$\sqrt{N}\frac{\mathbf{X}_i'\overline{\mathbf{M}}_w\boldsymbol{\varepsilon}_i}{T} - \sqrt{N}\frac{\mathbf{X}_i'\overline{\mathbf{M}}_q\boldsymbol{\varepsilon}_i}{T} \xrightarrow{p} \mathbf{0} \textit{ uniformly in } i, \tag{A.49}$$

$$\sqrt{N}\frac{\mathbf{X}_i'\overline{\mathbf{M}}_w\mathbf{F}}{T} - \sqrt{N}\frac{\mathbf{X}_i'\overline{\mathbf{M}}_q\mathbf{F}}{T} \xrightarrow{p} \mathbf{0} \textit{ uniformly in } i, \tag{A.50}$$

and

$$\sqrt{N}\frac{\mathbf{X}_i'\overline{\mathbf{M}}_w\boldsymbol{\vartheta}_i}{T} - \sqrt{N}\frac{\mathbf{X}_i'\overline{\mathbf{M}}_q\boldsymbol{\vartheta}_i}{T} \xrightarrow{p} \mathbf{0} \textit{ uniformly in } i, \tag{A.51}$$

where $\boldsymbol{\vartheta}_i = (\vartheta_{i1}, \vartheta_{i2}, ..., \vartheta_{iT})'$*, and* $\vartheta_{it} = \sum_{\ell=1}^{m_n(N)} \lambda_{i\ell} n_{\ell t}$.

Proof. We have

$$\begin{aligned} \frac{\sqrt{N}}{T}\mathbf{X}_i'\overline{\mathbf{M}}_w\mathbf{X}_i - \frac{\sqrt{N}}{T}\mathbf{X}_i'\mathbf{M}_q\mathbf{X}_i &= \sqrt{N}\frac{\mathbf{X}_i'\overline{\mathbf{H}}_w}{T}\left(\frac{\overline{\mathbf{H}}_w'\overline{\mathbf{H}}_w}{T}\right)^+\frac{\overline{\mathbf{H}}_w'\mathbf{X}_i}{T} - \sqrt{N}\frac{\mathbf{X}_i'\mathbf{Q}}{T}\left(\frac{\mathbf{Q}'\mathbf{Q}}{T}\right)^+\frac{\mathbf{Q}'\mathbf{X}_i}{T} \\ &= \frac{\sqrt{N}\mathbf{X}_i'\left(\overline{\mathbf{H}}_w - \mathbf{Q}\right)}{T}\left(\frac{\overline{\mathbf{H}}_w'\overline{\mathbf{H}}_w}{T}\right)^+\frac{\overline{\mathbf{H}}_w'\mathbf{X}_i}{T} + \\ &\quad + \frac{\mathbf{X}_i'\mathbf{Q}}{T}\left(\frac{\mathbf{Q}'\mathbf{Q}}{T}\right)^+\frac{\sqrt{N}\left(\mathbf{Q}' - \overline{\mathbf{H}}_w'\right)\mathbf{X}_i}{T} + \\ &\quad + \frac{\mathbf{X}_i'\mathbf{Q}}{T}\sqrt{N}\left[\left(\frac{\overline{\mathbf{H}}_w'\overline{\mathbf{H}}_w}{T}\right)^+ - \left(\frac{\mathbf{Q}'\mathbf{Q}}{T}\right)^+\right]\frac{\overline{\mathbf{H}}_w'\mathbf{X}_i}{T}. \end{aligned} \tag{A.52}$$

We focus on the individual elements on the right side of (A.52). The second part of (A.19), the second part of (A.21) and the first part of (A.26) imply

$$\underbrace{\frac{\sqrt{N}\mathbf{X}_i'\left(\overline{\mathbf{H}}_w-\mathbf{Q}\right)}{T}}_{o_p(1)}\underbrace{\left(\frac{\overline{\mathbf{H}}_w'\overline{\mathbf{H}}_w}{T}\right)^+}_{O_p(1)}\underbrace{\frac{\overline{\mathbf{H}}_w'\mathbf{X}_i}{T}}_{O_p(1)}\xrightarrow{p}\mathbf{0}\text{ uniformly in } i.$$

The second part of (A.20), the second part of (A.18), and the first part of (A.26) imply

$$\underbrace{\frac{\mathbf{X}_i'\mathbf{Q}}{T}}_{O_p(1)}\underbrace{\left(\frac{\mathbf{Q}'\mathbf{Q}}{T}\right)^+}_{O_p(1)}\underbrace{\frac{\sqrt{N}\left(\mathbf{Q}'-\overline{\mathbf{H}}_w'\right)\mathbf{X}_i}{T}}_{o_p(1)}\xrightarrow{p}\mathbf{0}\text{ uniformly in } i.$$

Finally, the second part of (A.20), the second part of (A.21) and result (A.27) imply

$$\underbrace{\frac{\mathbf{X}_i'\mathbf{Q}}{T}}_{O_p(1)}\underbrace{\sqrt{N}\left[\left(\frac{\overline{\mathbf{H}}_w'\overline{\mathbf{H}}_w}{T}\right)^+-\left(\frac{\mathbf{Q}'\mathbf{Q}}{T}\right)^+\right]}_{o_p(1)}\underbrace{\frac{\overline{\mathbf{H}}_w'\mathbf{X}_i}{T}}_{O_p(1)}\xrightarrow{p}\mathbf{0}\text{ uniformly in } i,$$

which completes the proof of (A.48).

To establish result (A.49), consider

$$\begin{aligned}\frac{\sqrt{N}}{T}\mathbf{X}_i'\overline{\mathbf{M}}_w\boldsymbol{\varepsilon}_i-\frac{\sqrt{N}}{T}\mathbf{X}_i'\mathbf{M}_q\boldsymbol{\varepsilon}_i \quad = \quad & \frac{\sqrt{N}\mathbf{X}_i'\left(\overline{\mathbf{H}}_w-\mathbf{Q}\right)}{T}\left(\frac{\overline{\mathbf{H}}_w'\overline{\mathbf{H}}_w}{T}\right)^+\frac{\overline{\mathbf{H}}_w'\boldsymbol{\varepsilon}_i}{T}\\ &+\frac{\mathbf{X}_i'\mathbf{Q}}{T}\left(\frac{\mathbf{Q}'\mathbf{Q}}{T}\right)^+\frac{\sqrt{N}\left(\mathbf{Q}'-\overline{\mathbf{H}}_w'\right)\boldsymbol{\varepsilon}_i}{T}\\ &+\frac{\mathbf{X}_i'\mathbf{Q}}{T}\sqrt{N}\left[\left(\frac{\overline{\mathbf{H}}_w'\overline{\mathbf{H}}_w}{T}\right)^+-\left(\frac{\mathbf{Q}'\mathbf{Q}}{T}\right)^+\right]\frac{\overline{\mathbf{H}}_w'\boldsymbol{\varepsilon}_i}{T} \qquad \text{(A.53)}\\ &\xrightarrow{p}\mathbf{0}\text{ uniformly in } i,\end{aligned}$$

where, similarly to the proof of (A.48), Lemmas A.2 and A.3 can be used repeatedly to establish the convergence of the elements on the right side of (A.53).

Results (A.50) and (A.51) can also be established in a similar way. In particular, Lemmas A.2 and A.3 imply

$$\begin{aligned}\frac{\sqrt{N}}{T}\mathbf{X}_i'\overline{\mathbf{M}}_w\mathbf{F}-\frac{\sqrt{N}}{T}\mathbf{X}_i'\mathbf{M}_q\mathbf{F} &= \frac{\sqrt{N}\mathbf{X}_i'\left(\overline{\mathbf{H}}_w-\mathbf{Q}\right)}{T}\left(\frac{\overline{\mathbf{H}}_w'\overline{\mathbf{H}}_w}{T}\right)^+\frac{\overline{\mathbf{H}}_w'\mathbf{F}}{T}\\ &+\frac{\mathbf{X}_i'\mathbf{Q}}{T}\left(\frac{\mathbf{Q}'\mathbf{Q}}{T}\right)^+\frac{\sqrt{N}\left(\mathbf{Q}'-\overline{\mathbf{H}}_w'\right)\mathbf{F}}{T}\\ &+\frac{\mathbf{X}_i'\mathbf{Q}}{T}\sqrt{N}\left[\left(\frac{\overline{\mathbf{H}}_w'\overline{\mathbf{H}}_w}{T}\right)^+-\left(\frac{\mathbf{Q}'\mathbf{Q}}{T}\right)^+\right]\frac{\overline{\mathbf{H}}_w'\mathbf{F}}{T}\\ &\xrightarrow{p}\mathbf{0}\text{ uniformly in } i,\end{aligned}$$

and

$$\begin{aligned}\frac{\sqrt{N}}{T}\mathbf{X}_i'\overline{\mathbf{M}}_w\boldsymbol{\vartheta}_i-\frac{\sqrt{N}}{T}\mathbf{X}_i'\mathbf{M}_q\boldsymbol{\vartheta}_i &= \frac{\sqrt{N}\mathbf{X}_i'\left(\overline{\mathbf{H}}_w-\mathbf{Q}\right)}{T}\left(\frac{\overline{\mathbf{H}}_w'\overline{\mathbf{H}}_w}{T}\right)^+\frac{\overline{\mathbf{H}}_w'\boldsymbol{\vartheta}_i}{T}\\ &+\frac{\mathbf{X}_i'\mathbf{Q}}{T}\left(\frac{\mathbf{Q}'\mathbf{Q}}{T}\right)^+\frac{\sqrt{N}\left(\mathbf{Q}'-\overline{\mathbf{H}}_w'\right)\boldsymbol{\vartheta}_i}{T}\\ &+\frac{\mathbf{X}_i'\mathbf{Q}}{T}\sqrt{N}\left[\left(\frac{\overline{\mathbf{H}}_w'\overline{\mathbf{H}}_w}{T}\right)^+-\left(\frac{\mathbf{Q}'\mathbf{Q}}{T}\right)^+\right]\frac{\overline{\mathbf{H}}_w'\boldsymbol{\vartheta}_i}{T}\\ &\xrightarrow{p}\mathbf{0}\text{ uniformly in } i.\end{aligned}$$

■

Lemma A.7 *Suppose Assumptions 5-10 hold, $(N,T)\xrightarrow{j}\infty$, and there exist constants α and K such that $0\leq\alpha<1/2$ and conditions (36) and (37) hold. Then,*

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{\mathbf{X}_i'\overline{\mathbf{M}}_q\boldsymbol{\varepsilon}_i}{T}\xrightarrow{L_1}\mathbf{0}, \tag{A.54}$$

and

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{\mathbf{X}_i'\overline{\mathbf{M}}_q\boldsymbol{\vartheta}_i}{T}\xrightarrow{p}\mathbf{0}, \tag{A.55}$$

where $\boldsymbol{\vartheta}_i=(\vartheta_{i1},\vartheta_{i2},...,\vartheta_{iT})'$, and $\vartheta_{it}=\sum_{\ell=1}^{m_n(N)}\lambda_{i\ell}n_{\ell t}$.

Proof. Proof of Lemma A.7 is similar to the proof of Lemma A.4. Let $T_N=T(N)$ be any non-decreasing integer-valued function of N such that $\lim_{N\to\infty}T_N=\infty$. Consider the following two-dimensional vector array $\{\boldsymbol{\kappa}_{Nt}\}$ defined by

$$\boldsymbol{\kappa}_{Nt}=\frac{1}{T_N\sqrt{N}}\sum_{i=1}^{N}\widetilde{\mathbf{x}}_{it}\varepsilon_{it}.$$

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 $N \in \mathbb{N}$. Using independence of $\widetilde{\mathbf{x}}_{it}$, and ε_{jt} for any $i, j \in \mathbb{N}$, and independence of ε_{it} and ε_{jt} for any $i \neq j$, we have

$$E\left(\frac{\boldsymbol{\kappa}_{Nt}\boldsymbol{\kappa}'_{Nt}}{c_{Nt}^2}\right) = \frac{1}{N}\sum_{i=1}^{N} E\left(\widetilde{\mathbf{x}}_{it}\widetilde{\mathbf{x}}'_{it}\right) E\left(\varepsilon_{it}^2\right),$$

and

$$\left\|E\left(\frac{\boldsymbol{\kappa}_{Nt}\boldsymbol{\kappa}'_{Nt}}{c_{Nt}^2}\right)\right\| \leq \sup_{i\in\mathbb{N}} \left\|E\left(\widetilde{\mathbf{x}}_{it}\widetilde{\mathbf{x}}'_{it}\right)\right\| \frac{1}{N}\sum_{i=1}^{N} E\left(\varepsilon_{it}^2\right) < K, \tag{A.56}$$

where $\sup_i E\left(\varepsilon_{it}^2\right) < K$ by Assumption 7, and $\sup_{i\in N}\left\|E\left(\widetilde{\mathbf{x}}_{it}\widetilde{\mathbf{x}}'_{it}\right)\right\| < K$ by Assumption 10. (A.56) implies uniform integrability of $\{\boldsymbol{\kappa}_{Nt}/c_{Nt}\}$. Since ε_{it} is covariance stationary process with absolute summable autocovariances, it follows that array $\boldsymbol{\kappa}_{Nt}$ is uniformly integrable L_1-mixingale array with respect to the constant array c_{Nt}, and using a mixingale weak law yields

$$\sum_{t=1}^{T_N}\boldsymbol{\kappa}_{Nt} = \frac{1}{T_N\sqrt{N}}\sum_{t=1}^{T_N}\sum_{i=1}^{N}\widetilde{\mathbf{x}}_{it}\varepsilon_{it} \overset{L_1}{\to} \mathbf{0}.$$

This completes the proof of result (A.54). Result (A.55) is established in a similar way. This time, we define

$$\boldsymbol{\kappa}_{Nt} = \frac{1}{T_N\sqrt{N}}\sum_{i=1}^{N}\widetilde{\mathbf{x}}_{it}\vartheta_{it}.$$

We have

$$E\left(\frac{\boldsymbol{\kappa}_{Nt}\boldsymbol{\kappa}'_{Nt}}{c_{Nt}^2}\right) = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N} E\left(\widetilde{\mathbf{x}}_{it}\widetilde{\mathbf{x}}'_{jt}\right) E\left(\vartheta_{it}\vartheta_{jt}\right).$$

Noting that $\sup_{i,j\in N}\left\|E\left(\widetilde{\mathbf{x}}_{it}\widetilde{\mathbf{x}}'_{jt}\right)\right\| < K$ (by Assumption 10), and that $\vartheta_{it} = \sum_{\ell=1}^{m_n(N)}\lambda_{i\ell}n_{\ell t}$, we obtain

$$\begin{aligned}\left\|E\left(\frac{\boldsymbol{\kappa}_{Nt}\boldsymbol{\kappa}'_{Nt}}{c_{Nt}^2}\right)\right\| &< \frac{K}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{\ell=1}^{m_n(N)}\lambda_{i\ell}\lambda_{j\ell} \\ &< \frac{K}{N}\sum_{\ell=1}^{m_n(N)}\left(\sum_{i=1}^{N}\lambda_{i\ell}\right)^2\end{aligned}$$

Using conditions (36) and (37), and noting that $0 \leq \alpha < 1/2$ imply

$$\left\|E\left(\frac{\boldsymbol{\kappa}_{Nt}\boldsymbol{\kappa}'_{Nt}}{c_{Nt}^2}\right)\right\| < K\ N^{2\alpha-1}\ m_n(N) < K.$$

Hence $\left\|E\left(\boldsymbol{\kappa}_{Nt}\boldsymbol{\kappa}'_{Nt}/c_{Nt}^2\right)\right\|$ is bounded in $N \in \mathbb{N}$. Using now the same arguments as in derivation of (A.54), we have

$$\sum_{t=1}^{T_N}\boldsymbol{\kappa}_{Nt} = \frac{1}{T_N\sqrt{N}}\sum_{t=1}^{T_N}\sum_{i=1}^{N}\widetilde{\mathbf{x}}_{it}\vartheta_{it} \overset{L_1}{\to} \mathbf{0},$$

which completes the proof of result (A.55). ■

Lemma A.8 *Suppose Assumptions 5-10 hold,* $(N, T) \overset{j}{\to} \infty$*, and there exist constants* α *and* K *such that*

$0 \leq \alpha < 1$ *and conditions (32) and (33) hold. Then,*

$$\frac{\mathbf{X}_i'\overline{\mathbf{M}}_w\mathbf{X}_i}{T} - \frac{\mathbf{X}_i'\overline{\mathbf{M}}_q\mathbf{X}_i}{T} \xrightarrow{p} \mathbf{0} \text{ uniformly in } i, \tag{A.57}$$

$$\frac{\mathbf{X}_i'\overline{\mathbf{M}}_w\mathbf{F}}{T} - \frac{\mathbf{X}_i'\overline{\mathbf{M}}_q\mathbf{F}}{T} \xrightarrow{p} \mathbf{0} \text{ uniformly in } i, \tag{A.58}$$

$$\frac{\mathbf{X}_i'\overline{\mathbf{M}}_w\boldsymbol{\vartheta}_i}{T} - \frac{\mathbf{X}_i'\overline{\mathbf{M}}_q\boldsymbol{\vartheta}_i}{T} \xrightarrow{p} \mathbf{0} \text{ uniformly in } i, \tag{A.59}$$

and

$$\frac{\mathbf{X}_i'\overline{\mathbf{M}}_w\boldsymbol{\varepsilon}_i}{T} - \frac{\mathbf{X}_i'\overline{\mathbf{M}}_q\boldsymbol{\varepsilon}_i}{T} \xrightarrow{p} \mathbf{0} \text{ uniformly in } i, \tag{A.60}$$

where $\boldsymbol{\vartheta}_i = (\vartheta_{i1}, \vartheta_{i2}, ..., \vartheta_{iT})'$, *and* $\vartheta_{it} = \sum_{\ell=1}^{m_n(N)} \lambda_{i\ell} n_{\ell t}$.

Proof. Results (A.57)-(A.60) can be established in a similar way as results (A.48)-(A.51) of Lemma A.6, i.e. Lemmas A.2 and A.3 can be used repeatedly to work out orders of magnitude in probability of individual elements in (A.57)-(A.60). ■

Lemma A.9 *Suppose Assumptions 5-10 hold,* $(N,T) \xrightarrow{j} \infty$, *and there exist constants* α *and* K *such that* $0 \leq \alpha < 1$ *and conditions (32) and (33) hold. Then,*

$$\sum_{i=1}^{N} w_i \frac{\mathbf{X}_i'\overline{\mathbf{M}}_w\mathbf{X}_i}{T}\boldsymbol{\upsilon}_i \xrightarrow{p} \mathbf{0}, \tag{A.61}$$

$$\sum_{i=1}^{N} w_i \frac{\mathbf{X}_i'\overline{\mathbf{M}}_w\mathbf{F}}{T}\boldsymbol{\eta}_i \xrightarrow{p} \mathbf{0}, \tag{A.62}$$

and

$$\frac{1}{N}\sum_{i=1}^{N} \boldsymbol{\Sigma}_{iq}^{-1} \frac{\mathbf{X}_i'\mathbf{M}_q\mathbf{F}}{T}\boldsymbol{\eta}_i \xrightarrow{p} \mathbf{0}. \tag{A.63}$$

Proof. Granularity conditions (1) and (2) imply

$$|w_i| < \frac{K}{N}, \tag{A.64}$$

where constant K does not depend on $N \in \mathbb{N}$ nor on $i = 1, 2, ..., N$. Using (A.64) and result (A.57) of Lemma A.8 yields

$$\sum_{i=1}^{N} w_i \frac{\mathbf{X}_i'\overline{\mathbf{M}}_w\mathbf{X}_i}{T}\boldsymbol{\upsilon}_i - \sum_{i=1}^{N} w_i \frac{\mathbf{X}_i'\overline{\mathbf{M}}_q\mathbf{X}_i}{T}\boldsymbol{\upsilon}_i \xrightarrow{p} \mathbf{0}.$$

But,

$$\sum_{i=1}^{N} w_i \frac{\mathbf{X}_i'\overline{\mathbf{M}}_q\mathbf{X}_i}{T}\boldsymbol{\upsilon}_i = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{\Delta}_{NTi}\boldsymbol{\upsilon}_i,$$

where

$$\boldsymbol{\Delta}_{NTi} = w_i^* \left(\frac{\mathbf{X}_i'\overline{\mathbf{M}}_q\mathbf{X}_i}{T}\right),$$

$w_i^* = Nw_i$, and (A.64) imply $|w_i^*| < K$. Also $\left(\mathbf{X}_i'\overline{\mathbf{M}}_q\mathbf{X}_i/T\right)$ has bounded second moments by Assumption 10, and therefore $E\left(\boldsymbol{\Delta}_{NTi}^2\right) < K$. Furthermore, $\boldsymbol{\upsilon}_i$ is independently distributed across i and independently distributed of $\mathbf{X}_i'\overline{\mathbf{M}}_q\mathbf{X}_i/T$. It follows that

$$\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Delta}_{NTi}\boldsymbol{\upsilon}_i \xrightarrow{p} \mathbf{0},$$

and

$$\sum_{i=1}^{N} w_i \frac{\mathbf{X}_i'\overline{\mathbf{M}}_w\mathbf{X}_i}{T}\boldsymbol{\upsilon}_i \xrightarrow{p} \mathbf{0},$$

as required. Result (A.62) and (A.63) can be established in a similar way as (A.64). ■

Lemma A.10 *Suppose Assumptions 5-10 hold,* $(N,T) \xrightarrow{j} \infty$*, and there exist constants* α *and* K *such that* $0 \leq \alpha < 1$ *and conditions (32) and (33) hold. Then,*

$$\left(\sum_{i=1}^{N} w_i \frac{\mathbf{X}_i'\overline{\mathbf{M}}_w\mathbf{X}_i}{T}\right)^{-1} = O_p\left(1\right). \tag{A.65}$$

Proof. Results (A.57) of Lemma A.8 and result (A.35) of Lemma A.5 imply

$$\frac{\mathbf{X}_i'\overline{\mathbf{M}}_w\mathbf{X}_i}{T} \xrightarrow{p} \boldsymbol{\Sigma}_{iq} \text{ uniformly in } i,$$

and therefore for any weights $\{w_i\}$ satisfying granularity conditions (1)-(2) we have

$$\sum_{i=1}^{N} w_i \frac{\mathbf{X}_i'\overline{\mathbf{M}}_w\mathbf{X}_i}{T} - \sum_{i=1}^{N} w_i \boldsymbol{\Sigma}_{iq} \xrightarrow{p} \mathbf{0},$$

as $(N,T) \xrightarrow{j} \infty$. The limit $\lim_{N\to\infty}\sum_{i=1}^{N} w_i\boldsymbol{\Sigma}_{iq} = \boldsymbol{\Psi}^*$ exists by Assumption 10 and furthermore, by the same assumption, $\boldsymbol{\Psi}^*$ is nonsingular. This implies (A.65). ■

B Mathematical proofs

Proof of Theorem 2. We prove the theorem in two parts. First, we establish consistency of the CCEP estimator and in the second part we establish consistency of the CCEMG estimator. Consider

$$\widehat{\boldsymbol{\beta}}_P - \boldsymbol{\beta} = \left(\sum_{i=1}^{N} w_i \frac{\mathbf{X}_i'\overline{\mathbf{M}}_w\mathbf{X}_i}{T}\right)^{-1} \sum_{i=1}^{N} w_i \frac{\mathbf{X}_i'\overline{\mathbf{M}}_w\left(\mathbf{X}_i\boldsymbol{\upsilon}_i + \mathbf{F}\boldsymbol{\gamma}_i + \boldsymbol{\vartheta}_i + \boldsymbol{\varepsilon}_i\right)}{T}. \tag{B.1}$$

We focus on the individual elements on the right side of (B.1) below. Lemma A.10 established

$$\left(\sum_{i=1}^{N} w_i \frac{\mathbf{X}_i'\overline{\mathbf{M}}_w\mathbf{X}_i}{T}\right)^{-1} = O_p\left(1\right). \tag{B.2}$$

According to result (A.61) of Lemma A.9, we have

$$\sum_{i=1}^{N} w_i \frac{\mathbf{X}_i' \overline{\mathbf{M}}_w \mathbf{X}_i}{T} \boldsymbol{\upsilon}_i \xrightarrow{p} \mathbf{0}. \tag{B.3}$$

Noting that $\boldsymbol{\gamma}_i$ can be written as $\boldsymbol{\gamma}_i = \overline{\boldsymbol{\gamma}}_w + \boldsymbol{\eta}_i - \overline{\boldsymbol{\eta}}_w$, and that $\sum_{i=1}^{N} w_i \mathbf{X}_i' \overline{\mathbf{M}}_w = \overline{\mathbf{X}}_w' \overline{\mathbf{M}}_w = \mathbf{0}$, and using result (A.62) of Lemma A.9, we obtain

$$\sum_{i=1}^{N} w_i \frac{\mathbf{X}_i' \overline{\mathbf{M}}_w \mathbf{F}}{T} \boldsymbol{\gamma}_i = \sum_{i=1}^{N} w_i \frac{\mathbf{X}_i' \overline{\mathbf{M}}_w \mathbf{F}}{T} \boldsymbol{\eta}_i \xrightarrow{p} \mathbf{0}. \tag{B.4}$$

Result (A.59) of Lemma A.8 and result (A.32) of Lemma A.4 imply

$$\sum_{i=1}^{N} w_i \frac{\mathbf{X}_i' \overline{\mathbf{M}}_w \boldsymbol{\vartheta}_i}{T} \xrightarrow{p} \mathbf{0}. \tag{B.5}$$

Similarly, result (A.60) of Lemma A.8 and result (A.31) of Lemma A.4 yields

$$\sum_{i=1}^{N} w_i \frac{\mathbf{X}_i' \overline{\mathbf{M}}_w \boldsymbol{\varepsilon}_i}{T} \xrightarrow{p} \mathbf{0}. \tag{B.6}$$

Using (B.2)-(B.6) in (B.1) establishes (35).

Next we establish consistency of CCEMG estimator. Consider

$$\widehat{\boldsymbol{\beta}}_{MG} - \boldsymbol{\beta} = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{\upsilon}_i + \frac{1}{N}\sum_{i=1}^{N} \widehat{\boldsymbol{\Psi}}_{iT}^{-1} \frac{\mathbf{X}_i' \overline{\mathbf{M}}_w \mathbf{F}}{T} \boldsymbol{\gamma}_i + \frac{1}{N}\sum_{i=1}^{N} \widehat{\boldsymbol{\Psi}}_{iT}^{-1} \frac{\mathbf{X}_i' \overline{\mathbf{M}}_w \boldsymbol{\vartheta}_i}{T} + \frac{1}{N}\sum_{i=1}^{N} \widehat{\boldsymbol{\Psi}}_{iT}^{-1} \frac{\mathbf{X}_i' \overline{\mathbf{M}}_w \boldsymbol{\varepsilon}_i}{T}, \tag{B.7}$$

where $\widehat{\boldsymbol{\Psi}}_{iT} = T^{-1} \mathbf{X}_i' \overline{\mathbf{M}}_w \mathbf{X}_i$. $\boldsymbol{\upsilon}_i$ is identically and independently distributed across i with zero mean and bounded second moments, and therefore

$$\frac{1}{N}\sum_{i=1}^{N} \boldsymbol{\upsilon}_i \xrightarrow{p} \mathbf{0}. \tag{B.8}$$

Results (A.57) and (A.58) of Lemma A.8 imply

$$\frac{1}{N}\sum_{i=1}^{N} \widehat{\boldsymbol{\Psi}}_{iT}^{-1} \frac{\mathbf{X}_i' \overline{\mathbf{M}}_w \mathbf{F}}{T} \boldsymbol{\gamma}_i - \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{\Sigma}_{iq}^{-1} \frac{\mathbf{X}_i' \mathbf{M}_q \mathbf{F}}{T} \boldsymbol{\gamma}_i \xrightarrow{p} \mathbf{0}.$$

But $\mathbf{F}\overline{\boldsymbol{\gamma}}_w$ belongs to the space spanned by column vectors of $\mathbf{Q}$, and therefore $\mathbf{M}_q \mathbf{F} \boldsymbol{\gamma}_i = \mathbf{M}_q \mathbf{F} \left(\overline{\boldsymbol{\gamma}}_w + \boldsymbol{\eta}_i - \overline{\boldsymbol{\eta}}_w \right) = \mathbf{M}_q \mathbf{F} \left(\boldsymbol{\eta}_i - \overline{\boldsymbol{\eta}}_w \right)$, where $\overline{\boldsymbol{\eta}}_w = O_p\left(N^{-1/2}\right)$. Now using (A.63) of Lemma A.9 it follows that

$$\frac{1}{N}\sum_{i=1}^{N} \widehat{\boldsymbol{\Psi}}_{iT}^{-1} \left(\frac{\mathbf{X}_i' \overline{\mathbf{M}}_w \mathbf{F}}{T} \right) \boldsymbol{\gamma}_i \xrightarrow{p} \mathbf{0}. \tag{B.9}$$

Results (A.57) and (A.59) of Lemma A.8, and result (A.32) of Lemma A.4 imply

$$\frac{1}{N}\sum_{i=1}^{N}\widehat{\mathbf{\Psi}}_{iT}^{-1}\left(\frac{\mathbf{X}_i'\overline{\mathbf{M}}_w\boldsymbol{\vartheta}_i}{T}\right)\xrightarrow{p}\mathbf{0}. \tag{B.10}$$

Similarly, results (A.57) and (A.60) of Lemma A.8, and result (A.31) of Lemma A.4 imply

$$\frac{1}{N}\sum_{i=1}^{N}\widehat{\mathbf{\Psi}}_{iT}^{-1}\left(\frac{\mathbf{X}_i'\overline{\mathbf{M}}_w\boldsymbol{\varepsilon}_i}{T}\right)\xrightarrow{p}\mathbf{0}. \tag{B.11}$$

Using (B.8)-(B.11) in (B.7) establish (34). ■

Proof of Theorem 3. We prove the theorem in two parts. First, we establish asymptotic distribution of the CCEP estimator and in the second part we establish asymptotic distribution of the CCEMG estimator. Consider

$$\left(\sum_{i=1}^{N}w_i^2\right)^{-1/2}\left(\widehat{\boldsymbol{\beta}}_P-\boldsymbol{\beta}\right)=\left(\sum_{i=1}^{N}w_i\frac{\mathbf{X}_i'\overline{\mathbf{M}}_w\mathbf{X}_i}{T}\right)^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\widetilde{w}_i\frac{\mathbf{X}_i'\overline{\mathbf{M}}_w\left(\mathbf{X}_i\mathbf{v}_i+\mathbf{F}\boldsymbol{\gamma}_i+\boldsymbol{\vartheta}_i+\boldsymbol{\varepsilon}_i\right)}{T}, \tag{B.12}$$

where $\widetilde{w}_i=\sqrt{N}w_i\left(\sum_{i=1}^{N}w_i^2\right)^{-1/2}$, and, by granularity conditions (1)-(2) there exists a real constant $K<\infty$ (independent of i and N), such that

$$|\widetilde{w}_i|=\left|\sqrt{N}w_i\left(\sum_{i=1}^{N}w_i^2\right)^{-1/2}\right|<K. \tag{B.13}$$

We focus on the individual terms on the right side of (B.12) below. Results (A.48) of Lemma A.6 and result (A.35) of Lemma A.5 imply

$$\frac{\mathbf{X}_i'\overline{\mathbf{M}}_w\mathbf{X}_i}{T}\xrightarrow{p}\mathbf{\Sigma}_{iq}\text{ uniformly in }i,$$

and therefore for any weights $\{w_i\}$ satisfying granularity conditions (1)-(2) we have

$$\sum_{i=1}^{N}w_i\frac{\mathbf{X}_i'\overline{\mathbf{M}}_w\mathbf{X}_i}{T}-\sum_{i=1}^{N}w_i\mathbf{\Sigma}_{iq}\xrightarrow{p}\mathbf{0},$$

as $(N,T)\xrightarrow{j}\infty$. The limit $\lim_{N\to\infty}\sum_{i=1}^{N}w_i\mathbf{\Sigma}_{iq}=\mathbf{\Psi}^*$ exists by Assumption 10 and furthermore, by the same assumption, $\mathbf{\Psi}^*$ is nonsingular. It follows that

$$\left(\sum_{i=1}^{N}w_i\frac{\mathbf{X}_i'\overline{\mathbf{M}}_w\mathbf{X}_i}{T}\right)^{-1}\xrightarrow{p}\mathbf{\Psi}^{*-1}, \tag{B.14}$$

as $(N,T)\xrightarrow{j}\infty$. Next we focus on the individual elements in the second summation on the right side of equation (B.12). Noting that $\boldsymbol{\gamma}_i$ can be written as $\boldsymbol{\gamma}_i=\overline{\boldsymbol{\gamma}}_w+\boldsymbol{\eta}_i-\overline{\boldsymbol{\eta}}_w$, and that $\sum_{i=1}^{N}w_i\mathbf{X}_i'\overline{\mathbf{M}}_w=\overline{\mathbf{X}}_w'\overline{\mathbf{M}}_w=\mathbf{0}$, we have

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\widetilde{w}_i\mathbf{X}_i'\overline{\mathbf{M}}_w\mathbf{F}\boldsymbol{\gamma}_i=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\widetilde{w}_i\mathbf{X}_i'\overline{\mathbf{M}}_w\mathbf{F}\boldsymbol{\eta}_i. \tag{B.15}$$

(B.13), (B.15) and result (A.50) of Lemma A.6 imply

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\widetilde{w}_i\frac{\mathbf{X}_i'\overline{\mathbf{M}}_w\mathbf{F}}{T}\boldsymbol{\gamma}_i-\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\widetilde{w}_i\frac{\mathbf{X}_i'\overline{\mathbf{M}}_q\mathbf{F}}{T}\boldsymbol{\eta}_i\overset{p}{\to}\mathbf{0}. \tag{B.16}$$

(B.13) and result (A.51) of Lemma A.6 imply

$$\frac{1}{N}\sum_{i=1}^{N}\widetilde{w}_i\left(\sqrt{N}\frac{\mathbf{X}_i'\overline{\mathbf{M}}_w\boldsymbol{\vartheta}_i}{T}\right)-\frac{1}{N}\sum_{i=1}^{N}\widetilde{w}_i\left(\sqrt{N}\frac{\mathbf{X}_i'\overline{\mathbf{M}}_q\boldsymbol{\vartheta}_i}{T}\right)\overset{p}{\to}\mathbf{0},$$

and, using result (A.55) of Lemma A.7, we have

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\widetilde{w}_i\frac{\mathbf{X}_i'\overline{\mathbf{M}}_w\boldsymbol{\vartheta}_i}{T}\overset{p}{\to}\mathbf{0}, \tag{B.17}$$

as $(N,T)\overset{j}{\to}\infty$. Similarly, result (A.49) of Lemma A.6 and result (A.54) of Lemma A.7 establish

$$\sqrt{N}\frac{\mathbf{X}_i'\overline{\mathbf{M}}_w\boldsymbol{\varepsilon}_i}{T}\overset{p}{\to}\mathbf{0}\text{ uniformly in }i,$$

and therefore (noting that $\widetilde{w}_i$ is uniformly bounded in i, see (B.13)),

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\widetilde{w}_i\frac{\mathbf{X}_i'\overline{\mathbf{M}}_w\boldsymbol{\varepsilon}_i}{T}=\frac{1}{N}\sum_{i=1}^{N}\widetilde{w}_i\left(\sqrt{N}\frac{\mathbf{X}_i'\overline{\mathbf{M}}_w\boldsymbol{\varepsilon}_i}{T}\right)\overset{p}{\to}\mathbf{0}. \tag{B.18}$$

Using (B.14), (B.16), (B.17), (B.18) and result (A.48) of Lemma A.6 in (B.12), we obtain

$$\left(\sum_{i=1}^{N}w_i^2\right)^{-1/2}\left(\widehat{\boldsymbol{\beta}}_P-\boldsymbol{\beta}\right)\overset{d}{\sim}\boldsymbol{\Psi}^{*-1}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\widetilde{w}_i\frac{\mathbf{X}_i'\overline{\mathbf{M}}_q\left(\mathbf{X}_i\mathbf{v}_i+\mathbf{F}\boldsymbol{\eta}_i\right)}{T}.$$

Assumption 10 is sufficient for the bounded second moments of $\mathbf{X}_i'\overline{\mathbf{M}}_q\mathbf{X}_i/T$ and $\mathbf{X}_i'\overline{\mathbf{M}}_q\mathbf{F}/T$. In particular, condition $E\left(\widetilde{x}_{ist}^4\right)<K$, for $s=1,2,..,k$, is sufficient for the bounded second moment of $\mathbf{X}_i'\overline{\mathbf{M}}_q\mathbf{X}_i/T$. To see this note that

$$\frac{\mathbf{X}_i'\overline{\mathbf{M}}_q\mathbf{X}_i}{T}=\frac{1}{T}\sum_{t=1}^{T}\widetilde{\mathbf{x}}_{it}\widetilde{\mathbf{x}}_{it}',$$

and, by Minkowski's inequality,

$$\left\|\frac{1}{T}\sum_{t=1}^{T}\widetilde{x}_{ist}\widetilde{x}_{ipt}'\right\|_{L_2}\leq\frac{1}{T}\sum_{t=1}^{T}\left\|\widetilde{x}_{ist}\widetilde{x}_{ipt}'\right\|_{L_2},$$

for any $s,p=1,2,..,k$. But by Cauchy-Schwarz inequality, we have $E\left(\widetilde{x}_{ist}^2\widetilde{x}_{ipt}^2\right)\leq\left[E\left(\widetilde{x}_{ist}^4\right)E\left(\widetilde{x}_{ipt}^4\right)\right]^{1/2}$, and therefore bounded fourth moments of the elements of $\widetilde{\mathbf{x}}_{it}$ are sufficient for the existence of an upper bound for the second moments of $\mathbf{X}_i'\overline{\mathbf{M}}_q\mathbf{X}_i/T$. Similar arguments can be used to establish that $\mathbf{X}_i'\overline{\mathbf{M}}_q\mathbf{F}/T$ has bounded second moments. It therefore follows from Lemma 4 of Pesaran (2006) and Lemma A.5 that

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

as $(N,T) \xrightarrow{j} \infty$, where

$$\boldsymbol{\Sigma}_P = \boldsymbol{\Psi}^{*-1}\mathbf{R}^*\boldsymbol{\Psi}^{*-1}, \tag{B.19}$$

in which

$$\boldsymbol{\Psi}^* = \lim_{N\to\infty}\sum_{i=1}^{N} w_i \boldsymbol{\Sigma}_{iq},\ \mathbf{R}^* = \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\widetilde{w}_i^2\left(\boldsymbol{\Sigma}_{iq}\boldsymbol{\Omega}_\beta\boldsymbol{\Sigma}_{iq} + \mathbf{Q}_{if}\boldsymbol{\Omega}_\gamma\mathbf{Q}_{if}'\right),$$

$\boldsymbol{\Omega}_\beta = Var\left(\boldsymbol{\beta}_i\right)$, $\boldsymbol{\Omega}_\gamma = Var\left(\boldsymbol{\gamma}_i\right)$, $\boldsymbol{\Sigma}_{iq}$ is defined in Assumption 10 and $\mathbf{Q}_{if}$ is defined by (A.38). Next, we consider asymptotic distribution of CCEMG estimator. Consider

$$\begin{aligned}\sqrt{N}\left(\widehat{\boldsymbol{\beta}}_{MG} - \boldsymbol{\beta}\right) &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\boldsymbol{\upsilon}_i + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\widehat{\boldsymbol{\Psi}}_{iT}^{-1}\frac{\mathbf{X}_i'\overline{\mathbf{M}}_w\mathbf{F}}{T}\boldsymbol{\gamma}_i + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\widehat{\boldsymbol{\Psi}}_{iT}^{-1}\frac{\mathbf{X}_i'\overline{\mathbf{M}}_w\boldsymbol{\vartheta}_i}{T} + \\ &+ \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\widehat{\boldsymbol{\Psi}}_{iT}^{-1}\frac{\mathbf{X}_i'\overline{\mathbf{M}}_w\boldsymbol{\varepsilon}_i}{T},\end{aligned} \tag{B.20}$$

where $\widehat{\boldsymbol{\Psi}}_{iT} = T^{-1}\mathbf{X}_i'\overline{\mathbf{M}}_w\mathbf{X}_i$. It follows from result (A.48) of Lemma A.6 and result (A.35) of Lemma A.5 that

$$\widehat{\boldsymbol{\Psi}}_{iT} - \boldsymbol{\Sigma}_{iq} = o_p\left(N^{-1/2}\right) \text{ uniformly in } i. \tag{B.21}$$

Using (B.21), result (A.51) of Lemma A.6, and result (A.55) of Lemma A.7, we have

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\widehat{\boldsymbol{\Psi}}_{iT}^{-1}\frac{\mathbf{X}_i'\overline{\mathbf{M}}_w\boldsymbol{\vartheta}_i}{T} \xrightarrow{p} \mathbf{0}. \tag{B.22}$$

Similarly, (B.21), result (A.49) of Lemma A.6, and result (A.54) of Lemma A.7 imply

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\widehat{\boldsymbol{\Psi}}_{iT}^{-1}\frac{\mathbf{X}_i'\overline{\mathbf{M}}_w\boldsymbol{\varepsilon}_i}{T} \xrightarrow{p} \mathbf{0}. \tag{B.23}$$

Noting that $\mathbf{F}\overline{\boldsymbol{\gamma}}_w$belongs to the linear space spanned by the column vectors of $\mathbf{Q} = \mathbf{G}\overline{\mathbf{P}}_w$, we have $\overline{\mathbf{M}}_q\mathbf{F}\overline{\boldsymbol{\gamma}}_w = \mathbf{0}$, and $\mathbf{X}_i'\overline{\mathbf{M}}_q\mathbf{F}\boldsymbol{\gamma}_i = \mathbf{X}_i'\overline{\mathbf{M}}_q\mathbf{F}\left(\boldsymbol{\eta}_i - \overline{\boldsymbol{\eta}}_w\right)$. Using results (A.48) and (A.50) of Lemma A.6 and noting that

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\frac{\mathbf{X}_i'\overline{\mathbf{M}}_q\mathbf{X}_i}{T}\right)^{-1}\frac{\mathbf{X}_i'\overline{\mathbf{M}}_q\mathbf{F}}{T}\overline{\boldsymbol{\eta}}_w \xrightarrow{p} \mathbf{0},$$

yields

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\widehat{\boldsymbol{\Psi}}_{iT}^{-1}\frac{\mathbf{X}_i'\overline{\mathbf{M}}_w\mathbf{F}}{T}\boldsymbol{\gamma}_i - \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\frac{\mathbf{X}_i'\overline{\mathbf{M}}_q\mathbf{X}_i}{T}\right)^{-1}\frac{\mathbf{X}_i'\overline{\mathbf{M}}_q\mathbf{F}}{T}\boldsymbol{\eta}_i \xrightarrow{p} \mathbf{0}. \tag{B.24}$$

Using (B.22)-(B.24) in (B.20) yields

$$\sqrt{N}\left(\widehat{\boldsymbol{\beta}}_{MG}-\boldsymbol{\beta}\right)\overset{d}{\sim}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\boldsymbol{\upsilon}_i+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\frac{\mathbf{X}_i'\overline{\mathbf{M}}_q\mathbf{X}_i}{T}\right)^{-1}\frac{\mathbf{X}_i'\overline{\mathbf{M}}_q\mathbf{F}}{T}\boldsymbol{\eta}_i.$$

It now follows that $\sqrt{N}\left(\widehat{\boldsymbol{\beta}}_{MG}-\boldsymbol{\beta}\right)\to N\left(\mathbf{0},\boldsymbol{\Sigma}_{MG}\right)$, where

$$\boldsymbol{\Sigma}_{MG}=\boldsymbol{\Omega}_{\beta}+\lim_{N\to\infty}\left[\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Sigma}_{iq}^{-1}\mathbf{Q}_{if}\boldsymbol{\Omega}_{\gamma}\mathbf{Q}_{if}'\boldsymbol{\Sigma}_{iq}^{-1}\right], \tag{B.25}$$

in which $\boldsymbol{\Omega}_{\beta}=Var\left(\boldsymbol{\beta}_i\right)$, $\boldsymbol{\Omega}_{\gamma}=Var\left(\boldsymbol{\gamma}_i\right)$, $\boldsymbol{\Sigma}_{iq}$ is defined in Assumption 10 and $\mathbf{Q}_{if}$ is defined by (A.38). ■