
Volume 160
Issue 2
February 2011
ISSN 0304-4076
ELSEVIER
JOURNAL OF
Econometrics
Editors:
TAKESHI AMEMIYA
A. RONALD GALLANT
JOHN F. GEWEKE
CHENG HSIAO
PETER ROBINSON
ARNOLD ZELLNER
Associate Editors:
YACINE AÏT-SAHALIA
BADI H. BALTAGI
RAVI BANSAL
MARCUS J. CHAMBERS
SONGNIAN CHEN
XIAOHONG CHEN
MIKHAIL CHERNOV
VICTOR CHERNOZHUKOV
MANFRED DEISTLER
MIGUEL A. DELGADO
YANQIN FAN
SYLVIA FRUHWIRTH-
-SCHNATTER
ERIC GHYSELS
JOHN C. HAM
JAVIER HIDALGO
HAN HONG
MICHAEL KEANE
YUICHI KITAMURA
GARY M. KOOP
NAOTO KUNITOMO
KAJAL LAHIRI
QI LI
TONG LI
ROSA L. MATZKIN
FRANCESCA MOLINARI
FRANZ C. PALM
DALE J. POIRIER
BENEDIKT PÖTSCHER
INGMAR PRUCHA
PETER C. REISS
ERIC RENAULT
FRANK SCHORFHEIDE
ROBIN SICKLES
MARK STEEL
FALLAW SOWELL
DAG BJARNE TJOESTHEIM
HERMAN VAN DIJK
QUANG H. VUONG
EDWARD VYTLACIL
TOM WANSBEEK
TAO ZHA
CONTENTS
J. Hahn, J.C. Ham and H.R. Moon, The Hausman test and weak instruments 289
G. Montes-Rojas and W. Sosa-Escudero, Robust tests for heteroskedasticity in the one-way error components model 300
M.J. Dueker, Z. Psaradakis, M. Sola and F. Spagnolo, Multivariate contemporaneous-threshold autoregressive models 311
CONTENTS CONTINUED ON BACK COVER

Journal of Econometrics 160 (2011) 326–348

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Panels with non-stationary multifactor error structures[☆]

G. Kapetanios[a], M. Hashem Pesaran[b,c], T. Yamagata[d,*]

[a] Queen Mary, University of London, United Kingdom
[b] Cambridge University, United Kingdom
[c] University of Southern California, CA, United States
[d] University of York, United Kingdom

A R T I C L E I N F O

Article history:
Received 15 June 2009
Received in revised form
14 April 2010
Accepted 5 October 2010
Available online 16 October 2010

JEL classification:
C12
C13
C33

Keywords:
Cross-section dependence
Large panels
Unit roots
Principal components
Common correlated effects

A B S T R A C T

The presence of cross-sectionally correlated error terms invalidates much inferential theory of panel data models. Recently, work by Pesaran (2006) has suggested a method which makes use of cross-sectional averages to provide valid inference in the case of stationary panel regressions with a multifactor error structure. This paper extends this work and examines the important case where the unobservable common factors follow unit root processes. The extension to $I(1)$ processes is remarkable on two counts. First, it is of great interest to note that while intermediate results needed for deriving the asymptotic distribution of the panel estimators differ between the $I(1)$ and $I(0)$ cases, the final results are surprisingly similar. This is in direct contrast to the standard distributional results for $I(1)$ processes that radically differ from those for $I(0)$ processes. Second, it is worth noting the significant extra technical demands required to prove the new results. The theoretical findings are further supported for small samples via an extensive Monte Carlo study. In particular, the results of the Monte Carlo study suggest that the cross-sectional-average-based method is robust to a wide variety of data generation processes and has lower biases than the alternative estimation methods considered in the paper.

1. Introduction

Panel data sets have been increasingly used in economics to analyze complex economic phenomena. One of their attractions is the ability to use an extended data set to obtain information about parameters of interest which are assumed to have common values across panel units. Most of the work carried out on panel data has usually assumed some form of cross-sectional independence to derive the theoretical properties of various inferential procedures. However, such assumptions are often suspect, and as a result recent advances in the literature have focused on estimation of panel data models subject to error cross-sectional dependence.

A number of different approaches have been advanced for this purpose. In the case of spatial data sets where a natural immutable distance measure is available, the dependence is often captured through "spatial lags" using techniques familiar from the time series literature. In economic applications, spatial techniques are often adapted using alternative measures of "economic distance". This approach is exemplified in work by Lee and Pesaran (1993), Conley and Dupor (2003), Conley and Topa (2002) and Pesaran et al. (2004), as well as the literature on spatial econometrics recently surveyed by Anselin (2001). In the case of panel data models where the cross-section dimension (N) is small (typically $N < 10$) and the time series dimension (T) is large, the standard approach is to treat the equations from the different cross-section units as a system of seemingly unrelated regression equations (SURE) and then estimate the system by generalized least squares (GLS) techniques.

The SURE approach is not applicable if the errors are correlated with the regressors and/or if the panels under consideration have a large cross-sectional dimension. This has led a number of investigators to consider unobserved factor models, where the cross-section error correlations are defined in terms of the factor loadings. The use of unobserved factors also allows for a certain degree of correlation between the idiosyncratic errors and the unobserved factors. Use of factor models is not new in economics, and dates back to the pioneering work of Stone (1947), who applied the principal component (PC) analysis of Hotelling

☆ The authors thank Vanessa Smith and Elisa Tosetti, and seminar participants at Kyoto University, University of Amsterdam, University of Nottingham and the annual conference at the Granger Centre for most helpful comments on a previous version of this paper. This version has also benefited greatly from constructive comments and suggestions by the Editor (Cheng Hsiao), an associate editor and three anonymous referees. Hashem Pesaran and Takashi Yamagata acknowledge financial support from the ESRC (Grant No. RES-000-23-0135).

* Corresponding author. Tel.: +44 0 1904 433758.
E-mail address: ty509@york.ac.uk (T. Yamagata).

0304-4076/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2010.10.001

to US macroeconomic time series over the period 1922–1938 and was able to demonstrate that three factors (namely total income, its rate of change and a time trend) explained over 97% of the total variations of all the 17 macro variables that he had considered. Until recently, subsequent applications of the PC approach to economic times series has been primarily in finance. See, for example, Chamberlain and Rothschild (1983), Connor and Korajzcyk (1986) and Connor (1988). But more recently the unobserved factor models have gained popularity for forecasting with a large number of variables, as advocated by Stock and Watson (2002). The factor model is used very much in the spirit of the original work by Stone, in order to summarize the empirical content of a large number of macroeconomics variables by a small set of factors which, when estimated using principal components, is then used for further modelling and/or forecasting. Related literature on dynamic factor models has also been put forward by Forni and Reichlin (1998) and Forni et al. (2000).

Recent uses of factor models in forecasting focus on the consistent estimation of unobserved factors and their loadings. Related theoretical advances by Bai and Ng (2002) and Bai (2003) are also concerned with the estimation and selection of unobserved factors and do not consider the estimation and inference problems in standard panel data models, where the objects of interest are slope coefficients of the conditioning variables (regressors). In such panels, the unobserved factors are viewed as nuisance variables, introduced primarily to model the cross-section dependences of the error terms in a parsimonious manner relative to the SURE formulation.

Despite these differences, knowledge of factor models could still be useful for the analysis of panel data models if it is believed that the errors might be cross-sectionally correlated. Disregarding the possible factor structure of the errors in panel data models can lead to inconsistent parameter estimates and incorrect inference. Coakley et al. (2002) suggest a possible solution to the problem using the method of Stock and Watson (2002). But, as Pesaran (2006) shows, the PC approach proposed by Coakley et al. (2002) can still yield inconsistent estimates. Pesaran (2006) suggests a new approach by noting that linear combinations of the unobserved factors can be well approximated by cross-section averages of the dependent variable and the observed regressors. This leads to a new set of estimators, referred to as the Common Correlated Effects (CCE) estimators, that can be computed by running standard panel regressions augmented with the cross-section averages of the dependent and independent variables. The CCE procedure is applicable to panels with a single factor or multiple unobserved factors, and it does not necessarily require the number of unobserved factors to be smaller than the number of observed cross-section averages.

In this paper, we extend the analysis of Pesaran (2006) to the case where the unobserved common factors are integrated of order 1, or $I(1)$. Our analysis does not require an *a priori* knowledge of the number of unobserved factors. It is only required that the number of unobserved factors remains fixed as the sample size is increased. The extension of the results of Pesaran (2006) to the $I(1)$ case is far from straightforward, and it involves the development of new intermediate results that could be of relevance to the analysis of panels with unit roots. It is also remarkable in the sense that, whilst the intermediate results needed for deriving the asymptotic distribution of the panel estimators differ between the $I(1)$ and $I(0)$ cases, the final results are surprisingly similar. This is in direct contrast to the usual phenomenon whereby distributional results for $I(1)$ processes are radically different to those for $I(0)$ processes and involve functionals of Brownian motion whose use requires separate tabulations of critical values.

It is very important to appreciate that our primary focus is on estimating the coefficients of the panel regression model. We do not wish to investigate the (co-)integration properties of the unobserved factors. Rather, our focus is robustness to the properties of the unobserved factors, for the estimation of the coefficients of the observed regressors that vary over time as well as over the cross-section units. In this sense, the extension provided by our work is of great importance in empirical applications where the integration properties of the unobserved common factors are typically unknown. In the CCE approach, the nature of the factors does not matter for inferential analysis of the coefficients of the observed variables. The theoretical findings of the paper are further supported for small samples via an extensive Monte Carlo study. In particular, the results of the Monte Carlo study clearly show that the CCE estimator is robust to a wide variety of data generation processes and has lower biases than all of the alternative estimation methods considered in the paper.

The structure of the paper is as follows. Section 2 provides an overview of the method suggested by Pesaran (2006) in the case of stationary factor processes. Section 3 provides the theoretical framework of the analysis of non-stationarity. In this section, the theoretical properties of the various estimators are presented. Section 4 presents an extensive Monte Carlo study, and Section 5 concludes. The Appendices contain proofs of the theoretical results.

Notation: K stands for a finite positive constant, $\|\boldsymbol{A}\| = [\text{Tr}(\boldsymbol{AA}')]^{1/2}$ is the Frobenius norm of the $m \times n$ matrix $\boldsymbol{A}$, and $\boldsymbol{A}^+$ denotes the Moore–Penrose inverse of $\boldsymbol{A}$. $rk(\boldsymbol{A})$ denotes the rank of $\boldsymbol{A}$. $\sup_i W_i$ is the supremum of W_i over i. $a_n = O(b_n)$ states that the deterministic sequence $\{a_n\}$ is at most of order b_n, $\boldsymbol{x}_n = O_p(\boldsymbol{y}_n)$ states that the vector of random variables, $\boldsymbol{x}_n$, is at most of order $\boldsymbol{y}_n$ in probability, and $\boldsymbol{x}_n = o_p(\boldsymbol{y}_n)$ is of smaller order in probability than $\boldsymbol{y}_n$; $\stackrel{q.m.}{\to}$ denotes convergence in quadratic mean (or mean square error), $\stackrel{p}{\to}$ convergence in probability, $\stackrel{d}{\to}$ convergence in distribution, and $\stackrel{d}{\sim}$ asymptotic equivalence of probability distributions. All asymptotics are carried out under $N \to \infty$, either with a fixed T, or *jointly* with $T \to \infty$. Joint convergence of N and T will be denoted by $(N, T) \stackrel{j}{\to} \infty$. Restrictions (if any) on the relative rates of convergence of N and T will be specified separately.

2. Panel data models with observed and unobserved common effects

In this section, we review the methodology introduced in Pesaran (2006). Let y_{it} be the observation on the ith cross-section unit at time t for $i = 1, 2, \ldots, N$; $t = 1, 2, \ldots, T$, and suppose that it is generated according to the following linear heterogeneous panel data model:

$$y_{it} = \boldsymbol{\alpha}_i'\boldsymbol{d}_t + \boldsymbol{\beta}_i'\boldsymbol{x}_{it} + \boldsymbol{\gamma}_i'\boldsymbol{f}_t + \varepsilon_{it}, \tag{1}$$

where $\boldsymbol{d}_t$ is an $n \times 1$ vector of observed common effects, which is partitioned as $\boldsymbol{d}_t = (\boldsymbol{d}_{1t}', \boldsymbol{d}_{2t}')'$, where $\boldsymbol{d}_{1t}$ is an $n_1 \times 1$ vector of deterministic components such as intercepts or seasonal dummies and $\boldsymbol{d}_{2t}$ is an $n_2 \times 1$ vector of unit root stochastic observed common effects, with $n = n_1 + n_2$, $\boldsymbol{x}_{it}$ is a $k \times 1$ vector of observed individual-specific regressors on the ith cross-section unit at time t, $\boldsymbol{f}_t$ is the $m \times 1$ vector of unobserved common effects, and ε_{it} are the individual-specific (idiosyncratic) errors assumed to be independently distributed of $(\boldsymbol{d}_t, \boldsymbol{x}_{it})$. The unobserved factors, $\boldsymbol{f}_t$, could be correlated with $(\boldsymbol{d}_t, \boldsymbol{x}_{it})$, and to allow for such a possibility the following specification for the individual specific regressors will be considered:

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

where $\boldsymbol{A}_i$ and $\boldsymbol{\Gamma}_i$ are $n \times k$ and $m \times k$ factor loading matrices with fixed and bounded components, and $\boldsymbol{v}_{it} = (v_{i1t}, \ldots, v_{ikt})'$ are the specific components of $\boldsymbol{x}_{it}$ distributed independently of the common effects and across i, but assumed to follow general covariance stationary processes. In our set-up, ε_{it} is assumed to be stationary, which implies that, in the case where $\boldsymbol{f}_t$ and/or $\boldsymbol{d}_t$ contain unit root processes, then y_{it}, $\boldsymbol{x}_{it}$, $\boldsymbol{d}_t$ and $\boldsymbol{f}_t$ must be cointegrated.[1] Some of the implications of this property are explored further in Remark 6.

Combining (1) and (2), we now have

$$\underset{(k+1)\times 1}{\boldsymbol{z}_{it}} = \begin{pmatrix} y_{it} \\ \boldsymbol{x}_{it} \end{pmatrix} = \underset{(k+1)\times n}{\boldsymbol{B}_i'}\ \underset{n\times 1}{\boldsymbol{d}_t} + \underset{(k+1)\times m}{\boldsymbol{C}_i'}\ \underset{m\times 1}{\boldsymbol{f}_t} + \underset{(k+1)\times 1}{\boldsymbol{u}_{it}}, \tag{3}$$

where

$$\boldsymbol{u}_{it} = \begin{pmatrix} \varepsilon_{it} + \boldsymbol{\beta}_i'\boldsymbol{v}_{it} \\ \boldsymbol{v}_{it} \end{pmatrix} = \begin{pmatrix} 1 & \boldsymbol{\beta}_i' \\ \boldsymbol{0} & \boldsymbol{I}_k \end{pmatrix} \begin{pmatrix} \varepsilon_{it} \\ \boldsymbol{v}_{it} \end{pmatrix}, \tag{4}$$

$$\begin{aligned} \boldsymbol{B}_i &= \begin{pmatrix} \boldsymbol{\alpha}_i & \boldsymbol{A}_i \end{pmatrix} \begin{pmatrix} 1 & \boldsymbol{0} \\ \boldsymbol{\beta}_i & \boldsymbol{I}_k \end{pmatrix}, \\ \boldsymbol{C}_i &= \begin{pmatrix} \boldsymbol{\gamma}_i & \boldsymbol{\Gamma}_i \end{pmatrix} \begin{pmatrix} 1 & \boldsymbol{0} \\ \boldsymbol{\beta}_i & \boldsymbol{I}_k \end{pmatrix}, \end{aligned} \tag{5}$$

$\boldsymbol{I}_k$ is an identity matrix of order k, and the rank of $\boldsymbol{C}_i$ is determined by the rank of the $m \times (k+1)$ matrix of the unobserved factor loadings

$$\tilde{\boldsymbol{\Gamma}}_i = \begin{pmatrix} \boldsymbol{\gamma}_i & \boldsymbol{\Gamma}_i \end{pmatrix}. \tag{6}$$

As discussed in Pesaran (2006), the above set-up is sufficiently general and renders a variety of panel data models as special cases. In the panel literature with T small and N large, the primary parameters of interest are the means of the individual specific slope coefficients, $\boldsymbol{\beta}_i$, $i = 1, 2, \ldots, N$. The common factor loadings, $\boldsymbol{\alpha}_i$ and $\boldsymbol{\gamma}_i$, are generally treated as nuisance parameters. In cases where both N and T are large, it is also possible to consider consistent estimation of the factor loadings, but this topic will not be pursued here. The presence of unobserved factors in (1) implies that estimation of $\boldsymbol{\beta}_i$ and its cross-sectional mean cannot be undertaken using standard methods. Pesaran (2006) has suggested using cross-section averages of y_{it} and $\boldsymbol{x}_{it}$ to deal with the effects of proxies for the unobserved factors in (1). To see why such an approach could work, consider simple cross-section averages of the equations in (3)[2]:

$$\bar{\boldsymbol{z}}_t = \bar{\boldsymbol{B}}'\boldsymbol{d}_t + \bar{\boldsymbol{C}}'\boldsymbol{f}_t + \bar{\boldsymbol{u}}_t, \tag{7}$$

where

$$\bar{\boldsymbol{z}}_t = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{z}_{it}, \qquad \bar{\boldsymbol{u}}_t = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{u}_{it},$$

and

$$\bar{\boldsymbol{B}} = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{B}_i, \qquad \bar{\boldsymbol{C}} = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{C}_i. \tag{8}$$

We distinguish between two important cases: when the rank condition

$$rk(\bar{\boldsymbol{C}}) = m \le k+1, \quad \text{for all } N, \text{ and as } N \to \infty, \tag{9}$$

holds, and when it does not. Under the former, the analysis simplifies considerably, since it is possible to proxy the unobserved factors by linear combinations of cross-section averages, $\bar{\boldsymbol{z}}_t$, and the observed common components, $\boldsymbol{d}_t$. But if the rank condition is not satisfied, this is not possible, although as we shall see it is still possible to consistently estimate the mean of the regression coefficients, $\boldsymbol{\beta}$, by the CCE procedure.

In the case where the rank condition is met, we have

$$\boldsymbol{f}_t = (\bar{\boldsymbol{C}}\bar{\boldsymbol{C}}')^{-1}\bar{\boldsymbol{C}}(\bar{\boldsymbol{z}}_t - \bar{\boldsymbol{B}}'\boldsymbol{d}_t - \bar{\boldsymbol{u}}_t). \tag{10}$$

But since

$$\bar{\boldsymbol{u}}_t \overset{q.m.}{\to} \boldsymbol{0}, \quad \text{as } N \to \infty, \text{ for each } t, \tag{11}$$

and

$$\bar{\boldsymbol{C}} \overset{p}{\to} \boldsymbol{C} = \tilde{\boldsymbol{\Gamma}} \begin{pmatrix} 1 & \boldsymbol{0} \\ \boldsymbol{\beta} & \boldsymbol{I}_k \end{pmatrix}, \quad \text{as } N \to \infty, \tag{12}$$

where

$$\tilde{\boldsymbol{\Gamma}} = (E(\boldsymbol{\gamma}_i), E(\boldsymbol{\Gamma}_i)) = (\boldsymbol{\gamma}, \boldsymbol{\Gamma}), \tag{13}$$

it follows, assuming that Rank $(\tilde{\boldsymbol{\Gamma}}) = m$, that

$$\boldsymbol{f}_t - (\boldsymbol{C}\boldsymbol{C}')^{-1}\boldsymbol{C}(\bar{\boldsymbol{z}}_t - \bar{\boldsymbol{B}}'\boldsymbol{d}_t) \overset{q.m.}{\to} \boldsymbol{0}, \quad \text{as } N \to \infty.$$

This suggests that, for sufficiently large N, it is valid to use $\bar{\boldsymbol{h}}_t = (\boldsymbol{d}_t', \bar{\boldsymbol{z}}_t')'$ as observable proxies for $\boldsymbol{f}_t$. This result holds irrespective of whether the unobserved factor loadings, $\boldsymbol{\gamma}_i$ and $\boldsymbol{\Gamma}_i$, are fixed or random.

When the rank condition is not satisfied, the use of cross-section averages alone does not allow consistent estimation of all of the unobserved factors, and as a result the estimation of the individual coefficients $\boldsymbol{\beta}_i$ by means of the cross-section averages alone will not be possible. But, interestingly enough, consistent estimates of the mean of the slope coefficients, $\boldsymbol{\beta}$, and their asymptotic distribution can be obtained if it is further assumed that the factor loadings are distributed independently of the factors and the individual-specific error processes.

2.1. The CCE estimators

We now discuss the two estimators for the means of the individual specific slope coefficients proposed by Pesaran (2006). One is the Mean Group (MG) estimator proposed in Pesaran and Smith (1995) and the other is a generalization of the fixed effects estimator that allows for the possibility of cross-section dependence. The former is referred to as the ''Common Correlated Effects Mean Group'' (CCEMG) estimator, and the latter as the ''Common Correlated Effects Pooled'' (CCEP) estimator.

The CCEMG estimator is a simple average of the individual CCE estimators, $\hat{\boldsymbol{b}}_i$ of $\boldsymbol{\beta}_i$,

$$\hat{\boldsymbol{b}}_{MG} = N^{-1}\sum_{i=1}^{N} \hat{\boldsymbol{b}}_i, \tag{14}$$

where

$$\hat{\boldsymbol{b}}_i = (\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{X}_i)^{-1}\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{y}_i, \tag{15}$$

$\boldsymbol{X}_i = (\boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}, \ldots, \boldsymbol{x}_{iT})'$, $\boldsymbol{y}_i = (y_{i1}, y_{i2}, \ldots, y_{iT})'$, $\bar{\boldsymbol{M}}$ is defined by

$$\bar{\boldsymbol{M}} = \boldsymbol{I}_T - \bar{\boldsymbol{H}}(\bar{\boldsymbol{H}}'\bar{\boldsymbol{H}})^{-1}\bar{\boldsymbol{H}}', \tag{16}$$

$\bar{\boldsymbol{H}} = (\boldsymbol{D}, \bar{\boldsymbol{Z}})$, $\boldsymbol{D}$ and $\bar{\boldsymbol{Z}}$ being, respectively, the $T \times n$ and $T \times (k+1)$ matrices of observations on $\boldsymbol{d}_t$ and $\bar{\boldsymbol{z}}_t$. We also define for later use

$$\boldsymbol{M}_g = \boldsymbol{I}_T - \boldsymbol{G}(\boldsymbol{G}'\boldsymbol{G})^{-1}\boldsymbol{G}', \tag{17}$$

and

[1] However, as will be shown later, our results on the estimators of $\boldsymbol{\beta}$ hold even if the factor loadings $\boldsymbol{\gamma}_i$ and/or $\boldsymbol{\Gamma}_i$ are zero (or weak in the sense of Chudik et al. (forthcoming)), and it is not necessary that $\boldsymbol{x}_{it}$ and $\boldsymbol{f}_t$ are cointegrated. What is required for our results is that, conditional on $\boldsymbol{d}_t$ and $\boldsymbol{f}_t$, the idiosyncratic errors ε_{it} and $\boldsymbol{v}_{it}$ are stationary.

[2] Pesaran (2006) considers cross-section weighted averages that are more general. But to simplify the exposition we confine our discussion to simple averages throughout.

$$\boldsymbol{M}_q = \boldsymbol{I}_T - \boldsymbol{Q}(\boldsymbol{Q}'\boldsymbol{Q})^+\boldsymbol{Q}', \quad \text{with } \boldsymbol{Q} = \boldsymbol{G}\bar{\boldsymbol{P}}, \tag{18}$$

where $\boldsymbol{G} = (\boldsymbol{D}, \boldsymbol{F}), \boldsymbol{D} = (\boldsymbol{d}_1, \boldsymbol{d}_2, \ldots, \boldsymbol{d}_T)', \boldsymbol{F} = (\boldsymbol{f}_1, \boldsymbol{f}_2, \ldots, \boldsymbol{f}_T)'$ are $T \times n$ and $T \times m$ data matrices on observed and unobserved common factors, respectively, $(\boldsymbol{A})^+$ denotes the Moore–Penrose inverse of $\boldsymbol{A}$, and

$$\underset{(n+m)\times(n+k+1)}{\bar{\boldsymbol{P}}} = \begin{pmatrix} \boldsymbol{I}_n & \bar{\boldsymbol{B}} \\ \boldsymbol{0} & \bar{\boldsymbol{C}} \end{pmatrix}, \qquad \bar{\boldsymbol{U}}^* = (\boldsymbol{0}, \bar{\boldsymbol{U}}), \tag{19}$$

where $\bar{\boldsymbol{U}}^*$ has the same dimension as $\bar{\boldsymbol{H}}$ and $\bar{\boldsymbol{U}} = (\bar{\boldsymbol{u}}_1, \bar{\boldsymbol{u}}_2, \ldots, \bar{\boldsymbol{u}}_T)'$ is a $T \times (k+1)$ matrix of observations on $\bar{\boldsymbol{u}}_t$. Efficiency gains from pooling of observations over the cross-section units can be achieved when the individual slope coefficients, $\boldsymbol{\beta}_i$, are the same. Such a pooled estimator of $\boldsymbol{\beta}$, denoted by CCEP, is given by

$$\hat{\boldsymbol{b}}_P = \left(\sum_{i=1}^N \boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{X}_i\right)^{-1} \sum_{i=1}^N \boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{y}_i, \tag{20}$$

which can also be viewed as a generalized fixed effects (GFE) estimator, and reduces to the standard FE estimator if $\bar{\boldsymbol{H}} = \boldsymbol{\tau}_T$ with $\boldsymbol{\tau}_T$ being a $T \times 1$ vector of ones.

3. Theoretical properties of CCE estimators in non-stationary panel data models

The following assumptions will be used in the derivation of the asymptotic properties of the CCE estimators.

Assumption 1 (*Non-Stationary Common Effects*)**.** The $(n_2+m) \times 1$ vector of stochastic common effects, $\mathfrak{g}_t = (\boldsymbol{d}_{2t}', \boldsymbol{f}_t')'$, follows the multivariate unit root process

$$\mathfrak{g}_t = \mathfrak{g}_{t-1} + \boldsymbol{\zeta}_{\mathfrak{g}t},$$

where $\boldsymbol{\zeta}_{\mathfrak{g}t}$ is an $(n_2+m) \times 1$ vector of $L_{2+\delta}$, $\delta > 0$, stationary near epoque dependent (NED) processes of size $1/2$, on some α-mixing process of size $-(2+\delta)/\delta$, distributed independently of the individual-specific errors, $\varepsilon_{it'}$ and $\boldsymbol{v}_{it'}$ for all i, t and t'.

Assumption 2 (*Individual-Specific Errors*)**.** (i) The individual-specific errors ε_{it} and $\boldsymbol{v}_{jt}$ are distributed independently of each other, for all i, j and t. ε_{it} have uniformly bounded positive variance, $\sup_i \sigma_i^2 < K$, for some constant K, and uniformly bounded fourth-order cumulants. $\boldsymbol{v}_{it}$ have covariance matrices, $\boldsymbol{\Sigma}_{\boldsymbol{v}_i}$, which are non-singular and satisfy $\sup_i \|\boldsymbol{\Sigma}_{\boldsymbol{v}_i}\| < K < \infty$, autocovariance matrices, $\boldsymbol{\Gamma}_{iv}(s)$, such that $\sup_i \sum_{s=-\infty}^{\infty} \|\boldsymbol{\Gamma}_{iv}(s)\| < K < \infty$, and have uniformly bounded fourth-order cumulants. (ii) For each i, $(\varepsilon_{it}, \boldsymbol{v}_{it}')'$ is an $(k+1) \times 1$ vector of $L_{2+\delta}$, $\delta > 0$, stationary near epoque dependent (NED) processes of size $\frac{2\delta}{2\delta-4}$ on some α-mixing process $\boldsymbol{\psi}_{it}$ of size $-(2+\delta)/\delta$ which is partitioned conformably to $(\varepsilon_{it}, \boldsymbol{v}_{it}')'$ as $(\psi_{\varepsilon it}, \boldsymbol{\psi}_{\boldsymbol{v}it}')'$, where $\psi_{\varepsilon it}$ and $\boldsymbol{\psi}_{\boldsymbol{v}jt}$ are independent for all i and j.

Assumption 3. The coefficient matrices, $\boldsymbol{B}_i$ and $\boldsymbol{C}_i$, are independently and identically distributed across i, and independent of the individual specific errors, ε_{jt} and $\boldsymbol{v}_{jt}$, the common factors, $\boldsymbol{\zeta}_{\mathfrak{g}t}$, for all i, j and t with fixed means $\boldsymbol{B}$ and $\boldsymbol{C}$, and uniformly bounded second-order moments. In particular,

$$\begin{aligned} &\mathrm{vec}(\boldsymbol{B}_i) = \mathrm{vec}(\boldsymbol{B}) + \boldsymbol{\eta}_{B,i}, \\ &\boldsymbol{\eta}_{B,i} \backsim \mathrm{IID}(\boldsymbol{0}, \boldsymbol{\Omega}_{B\eta}), \quad \text{for } i = 1, 2, \ldots, N, \end{aligned} \tag{21}$$

and

$$\begin{aligned} &\mathrm{vec}(\boldsymbol{C}_i) = \mathrm{vec}(\boldsymbol{C}) + \boldsymbol{\eta}_{C,i}, \\ &\boldsymbol{\eta}_{C,i} \backsim \mathrm{IID}(\boldsymbol{0}, \boldsymbol{\Omega}_{C\eta}), \quad \text{for } i = 1, 2, \ldots, N, \end{aligned} \tag{22}$$

where $\boldsymbol{\Omega}_{B\eta}$ and $\boldsymbol{\Omega}_{C\eta}$ are $(k+1)n \times (k+1)n$ and $(k+1)m \times (k+1)m$ symmetric non-negative definite matrices, $\|\boldsymbol{B}\| < K$, $\|\boldsymbol{C}\| < K$, $\|\boldsymbol{\Omega}_{B\eta}\| < K$ and $\|\boldsymbol{\Omega}_{C\eta}\| < K$, for some constant K.

Assumption 4 (*Random Slope Coefficients*)**.** The slope coefficients, $\boldsymbol{\beta}_i$, follow the random coefficient model

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{\varkappa}_i, \qquad \boldsymbol{\varkappa}_i \backsim \mathrm{IID}(\boldsymbol{0}, \boldsymbol{\Omega}_{\varkappa}), \quad \text{for } i = 1, 2, \ldots, N, \tag{23}$$

where $\|\boldsymbol{\beta}\| < K$, $\|\boldsymbol{\Omega}_{\varkappa}\| < K$, for some constant K, $\boldsymbol{\Omega}_{\varkappa}$ is a $k \times k$ symmetric non-negative definite matrix, and the random deviations, $\boldsymbol{\varkappa}_i$, are distributed independently of $\boldsymbol{\gamma}_j$, $\boldsymbol{\Gamma}_j$, ε_{jt}, $\boldsymbol{v}_{jt}$, and $\boldsymbol{\zeta}_{\mathfrak{g}t}$ for all i, j and t. $\boldsymbol{\varkappa}_i$ has finite fourth moments uniformly over i.

Assumption 5 (*Identification of $\boldsymbol{\beta}_i$ and $\boldsymbol{\beta}$*)**.** $\left(\frac{\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{X}_i}{T}\right)^{-1}$ exists for all i and T, and $\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^N \boldsymbol{\Sigma}_{\boldsymbol{v}_i}$ is non-singular.

Assumption 6. $\left(\frac{\boldsymbol{X}_i'\boldsymbol{M}_g\boldsymbol{X}_i}{T}\right)^{-1}$ exists for all i and T, and $\sup_i E\left\|\frac{\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{X}_i}{T}\right\|^2 < K < \infty$.

Assumption 7. When rank condition (9) is not satisfied, (i) $\frac{1}{N}\sum_{i=1}^N \frac{\boldsymbol{X}_i'\boldsymbol{M}_q\boldsymbol{X}_i}{T^2}$ and $\boldsymbol{\Theta} = \lim_{N,T\to\infty}\left(\frac{1}{N}\sum_{i=1}^N \boldsymbol{\Theta}_{iT}\right)$, where $\boldsymbol{\Theta}_{iT} = E(T^{-2}\boldsymbol{X}_i'\boldsymbol{M}_q\boldsymbol{X}_i)$, are non-singular; (ii) if $m \geq 2k+1$, then $\left(\frac{\boldsymbol{X}_i'\boldsymbol{M}_q\boldsymbol{X}_i}{T^2}\right)^{-1}$ exists for all i and T and $\sup_i E\left\|\left(\frac{\boldsymbol{X}_i'\boldsymbol{M}_q\boldsymbol{X}_i}{T^2}\right)^{-1}\left(\frac{\boldsymbol{X}_i'\boldsymbol{M}_q\boldsymbol{F}}{T^2}\right)\right\|^2 < \infty$; and (iii) if $m < 2k+1$, then $E\left\|\frac{\boldsymbol{F}'\boldsymbol{F}}{T^2}\right\|^2 < \infty$ and $E\left\|\left(\frac{\boldsymbol{F}'\boldsymbol{F}}{T^2}\right)^{-1}\right\|^2 < \infty$.

Remark 1. Assumption 1 departs from the standard practice in the analysis of large panels with common factors and specifies that the factors are non-stationary. Assumption 2 concerns the individual specific errors and relaxes the assumption that ε_{it} are serially uncorrelated, often adopted in the literature (see, e.g., Pesaran (2006)). Assumptions 2–6 are standard in large panels with random coefficients. But some comments on Assumption 7 seems to be in order. This assumption is only used when the rank condition (9) is not satisfied. It is made up of three regularity conditions.[3] The last two are of greater significance and only relate to the Mean Group estimator presented in the next section. In effect, these assumptions ensure that the individual slope coefficient estimators possess second-order moments asymptotically, which seems plausible in most economic applications.

Remark 2. Note that Assumption 3 implies that $\boldsymbol{\gamma}_i$ are independently and identically distributed across i, and

$$\begin{aligned} &\boldsymbol{\gamma}_i = \boldsymbol{\gamma} + \boldsymbol{\eta}_i, \\ &\boldsymbol{\eta}_i \backsim \mathrm{IID}(\boldsymbol{0}, \boldsymbol{\Omega}_\eta), \quad \text{for } i = 1, 2, \ldots, N, \end{aligned} \tag{24}$$

where $\boldsymbol{\Omega}_\eta$ is an $m \times m$ symmetric non-negative definite matrix, and $\|\boldsymbol{\gamma}\| < K$, and $\|\boldsymbol{\Omega}_\eta\| < K$, for some constant K.

For each i and $t = 1, 2, \ldots, T$, writing the model in matrix notation, we have

$$\boldsymbol{y}_i = \boldsymbol{D}\boldsymbol{\alpha}_i + \boldsymbol{X}_i\boldsymbol{\beta}_i + \boldsymbol{F}\boldsymbol{\gamma}_i + \boldsymbol{\varepsilon}_i, \tag{25}$$

where $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{iT})'$. Using (25) in (15), we have

$$\hat{\boldsymbol{b}}_i - \boldsymbol{\beta}_i = \left(\frac{\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{X}_i}{T}\right)^{-1}\left(\frac{\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{F}}{T}\right)\boldsymbol{\gamma}_i + \left(\frac{\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{X}_i}{T}\right)^{-1}\left(\frac{\boldsymbol{X}_i'\bar{\boldsymbol{M}}\varepsilon_i}{T}\right), \tag{26}$$

[3] $E\|T^{-2}\boldsymbol{F}'\boldsymbol{F}\|^2 < \infty$, which is part of Assumption 7(iii), can be established under mild regularity conditions (see Lemma 4 of Phillips and Moon (1999)).

which shows the direct dependence of $\hat{\boldsymbol{b}}_i$ on the unobserved factors through $T^{-1}\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{F}$. To examine the properties of this component, we first note that (2) and (7) can be written in matrix notation as

$$\boldsymbol{X}_i = \boldsymbol{G}\boldsymbol{\Pi}_i + \boldsymbol{V}_i, \tag{27}$$

and

$$\bar{\boldsymbol{H}} = (\boldsymbol{D}, \bar{\boldsymbol{Z}}) = (\boldsymbol{D}, \boldsymbol{D}\bar{\boldsymbol{B}} + \boldsymbol{F}\bar{\boldsymbol{C}} + \bar{\boldsymbol{U}}) = \boldsymbol{G}\bar{\boldsymbol{P}} + \bar{\boldsymbol{U}}^*, \tag{28}$$

where $\boldsymbol{\Pi}_i = (\boldsymbol{A}_i', \boldsymbol{\Gamma}_i')'$, $\boldsymbol{V}_i = (\boldsymbol{v}_{i1}, \boldsymbol{v}_{i2}, \ldots, \boldsymbol{v}_{iT})'$, $\boldsymbol{G} = (\boldsymbol{D}, \boldsymbol{F})$, and $\bar{\boldsymbol{P}}$ and $\bar{\boldsymbol{U}}^*$ are defined by (19).

Using Lemmas 3 and 4 in Appendix A, and assuming that rank condition (9) is satisfied, it follows that

$$\frac{\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{F}}{T} = O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{N}\right), \quad \text{uniformly over } i, \tag{29}$$

$$\frac{\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{X}_i}{T} - \frac{\boldsymbol{X}_i'\boldsymbol{M}_g\boldsymbol{X}_i}{T} = O_p\left(\frac{1}{\sqrt{N}}\right), \quad \text{uniformly over } i, \tag{30}$$

and

$$\frac{\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{\varepsilon}_i}{T} - \frac{\boldsymbol{X}_i'\boldsymbol{M}_g\boldsymbol{\varepsilon}_i}{T} = O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{N}\right), \quad \text{uniformly over } i. \tag{31}$$

If the rank condition does not hold, then by Lemma 6 in Appendix A it follows that

$$\frac{\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{F}}{T} - \frac{\boldsymbol{X}_i'\boldsymbol{M}_q\boldsymbol{F}}{T} = O_p\left(\frac{1}{\sqrt{N}}\right), \quad \text{uniformly over } i, \tag{32}$$

$$\frac{\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{X}_i}{T} - \frac{\boldsymbol{X}_i'\boldsymbol{M}_q\boldsymbol{X}_i}{T} = O_p\left(\frac{1}{\sqrt{N}}\right), \quad \text{uniformly over } i, \tag{33}$$

and

$$\frac{\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{\varepsilon}_i}{T} - \frac{\boldsymbol{X}_i'\boldsymbol{M}_q\boldsymbol{\varepsilon}_i}{T} = O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{N}\right), \quad \text{uniformly over } i. \tag{34}$$

In the following subsections we discuss our main theoretical results.

3.1. Results for pooled estimators

We now examine the asymptotic properties of the pooled estimators. Focusing first on the MG estimator, and using (26), we have

$$\sqrt{N}(\hat{\boldsymbol{b}}_{MG} - \boldsymbol{\beta}) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\boldsymbol{\varkappa}_i + \frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{\Psi}}_{iT}^{-1}\left(\frac{\sqrt{N}\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{F}}{T}\right)\boldsymbol{\gamma}_i + \frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{\Psi}}_{iT}^{-1}\left(\frac{\sqrt{N}\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{\varepsilon}_i}{T}\right), \tag{35}$$

where $\hat{\boldsymbol{\Psi}}_{iT} = T^{-1}\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{X}_i$. In the case where rank condition (9) is satisfied, by (29), we have

$$\frac{\sqrt{N}(\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{F})}{T} = O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\frac{1}{\sqrt{N}}\right). \tag{36}$$

Using this, we can formally show that

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

Hence

$$\sqrt{N}(\hat{\boldsymbol{b}}_{MG} - \boldsymbol{\beta}) \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{\Sigma}_{MG}), \quad \text{as } (N, T) \xrightarrow{j} \infty. \tag{37}$$

The variance estimator for $\boldsymbol{\Sigma}_{MG}$ suggested by Pesaran (2006) is given by

$$\hat{\boldsymbol{\Sigma}}_{MG} = \frac{1}{N-1}\sum_{i=1}^{N}(\hat{\boldsymbol{b}}_i - \hat{\boldsymbol{b}}_{MG})(\hat{\boldsymbol{b}}_i - \hat{\boldsymbol{b}}_{MG})', \tag{38}$$

which can be used here as well. The following theorem summarizes the results for the Mean Group estimator. The result is proved in Appendix C.

Theorem 1. *Consider the panel data models (1) and (2). Let Assumptions 1–6 and 7(ii), (iii) hold. Then, for the Common Correlated Effects Mean Group estimator, $\hat{\boldsymbol{b}}_{MG}$, defined by (14), we have, as $(N, T) \xrightarrow{j} \infty$, that*

$$\sqrt{N}(\hat{\boldsymbol{b}}_{MG} - \boldsymbol{\beta}) \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{\Sigma}_{MG}),$$

where

$$\boldsymbol{\Sigma}_{MG} = \boldsymbol{\Omega}_{\varkappa} + \boldsymbol{\Lambda}, \tag{39}$$

$$\boldsymbol{\Lambda} = \lim_{N,T\to\infty}\left[\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Sigma}_{iqT}\right] \tag{40}$$

and $\boldsymbol{\Sigma}_{iqT}$ is defined in (C.67). $\boldsymbol{\Sigma}_{MG}$ can be consistently estimated by (38).

Note that this theorem does not require that the rank condition, (9), holds for any number, m, of unobserved factors so long as m is fixed. Also, it does not impose any restrictions on the relative rates of expansion of N and T. The following theorem summarizes the results for the second pooled estimator, $\hat{\boldsymbol{b}}_P$. The proof is provided in Appendix C.

Theorem 2. *Consider the panel data models (1) and (2), and suppose that Assumptions 1–6 and 7(i) hold. Then, for the Common Correlated Effects Pooled estimator, $\hat{\boldsymbol{b}}_P$, defined by (20), as $(N, T) \xrightarrow{j} \infty$, we have that*

$$\sqrt{N}(\hat{\boldsymbol{b}}_P - \boldsymbol{\beta}) \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{\Sigma}_P^*),$$

where $\boldsymbol{\Sigma}_P^*$ is given by

$$\boldsymbol{\Sigma}_P^* = \boldsymbol{\Theta}^{-1}(\boldsymbol{\Xi} + \boldsymbol{\Phi})\boldsymbol{\Theta}^{-1}, \tag{41}$$

where

$$\boldsymbol{\Phi} = \lim_{N,T\to\infty}\left(\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Phi}_{Ti}\right), \qquad \boldsymbol{\Xi} = \lim_{N,T\to\infty}\left(\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Xi}_{Ti}\right),$$

$$\boldsymbol{\Theta} = \lim_{N,T\to\infty}\left(\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Theta}_{Ti}\right)$$

$\boldsymbol{\Xi}_{Ti} = \mathrm{Var}[T^{-2}\boldsymbol{X}_i'\boldsymbol{M}_q\boldsymbol{X}_i\boldsymbol{\varkappa}_i]$, and $\boldsymbol{\Phi}_{Ti}$ and $\boldsymbol{\Theta}_{Ti}$ are given by (C.87) and (C.84), respectively. $\boldsymbol{\Sigma}_P^*$ can be estimated consistently by

$$\hat{\boldsymbol{\Sigma}}_P^* = \hat{\boldsymbol{\Psi}}^{*-1}\hat{\boldsymbol{R}}^*\hat{\boldsymbol{\Psi}}^{*-1}, \tag{42}$$

where

$$\hat{\boldsymbol{\Psi}}^* = N^{-1}\sum_{i=1}^{N}\frac{\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{X}_i}{T}, \tag{43}$$

$$\hat{\boldsymbol{R}}^* = \frac{1}{(N-1)}\sum_{i=1}^{N}\left(\frac{\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{X}_i}{T}\right)(\hat{\boldsymbol{b}}_i - \hat{\boldsymbol{b}}_{MG})(\hat{\boldsymbol{b}}_i - \hat{\boldsymbol{b}}_{MG})'\left(\frac{\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{X}_i}{T}\right). \tag{44}$$

Overall we see that, despite a number of differences in the above analysis, especially in terms of the results given in (29)–(34), compared to the results in Pesaran (2006), the conclusions are remarkably similar when the factors are assumed to follow unit root processes.

Remark 3. The formal analysis in the Appendices focuses on the case where the factor is an $I(1)$ process and no cointegration is present among the factors. But, as shown by Johansen (1995, pp. 40), when the factor process is cointegrated and there are $l < m$ cointegrating vectors, we have that $\boldsymbol{F}\boldsymbol{\gamma}_i = \boldsymbol{F}_1\boldsymbol{\delta}_{1i} + \boldsymbol{F}_2\boldsymbol{\delta}_{2i}$, where $\boldsymbol{F}_1$ is an $m - l$-dimensional $I(1)$ process with no cointegration, whereas $\boldsymbol{F}_2$ is an l-dimensional $I(0)$ process. This implies that the cointegration case is equivalent to a case where the model contains a mix of non-cointegrated $I(1)$ and $I(0)$ factor processes. Since we know that the results of the paper hold for both non-cointegrated $I(1)$ and, by Pesaran (2006), $I(0)$ factor processes, we conjecture that they hold for the cointegrated case, as well. However, we feel that a formal proof of this statement is beyond the scope of the present paper. We consider a case of cointegrated factors in the Monte Carlo study. The results clearly support the above claim.

Remark 4. In the case of standard linear panel data models with strictly exogenous regressors and homogeneous slopes, and without unobserved common factors, Pesaran et al. (1996) show that in general the fixed effect estimator is asymptotically at least as efficient as the Mean Group estimator. It is reasonable to expect that this result also applies to the CCE type estimators, namely that, under $\boldsymbol{\beta}_i = \boldsymbol{\beta}$ for all i, the CCEP estimator would be at least as efficient as the CCEMG estimator. Although a formal proof is beyond the scope of the present paper, the Monte Carlo results reported below provide some evidence in favour of this conjecture.

As we noted above, the whole analysis does not depend on whether the rank condition holds or not. But in the case where the rank condition is satisfied, a number of simplifications arise. In particular, the technical Assumption 7 is not needed, and Assumption 3 can be relaxed. Namely the factor loadings, $\boldsymbol{\gamma}_i$, need not follow the random coefficient model. It would be sufficient that they are bounded. Also, the expressions for the theoretical covariance matrices of the estimators change, although crucially the estimators of these covariance matrices do not. For completeness, we present corollaries on the theoretical properties of the pooled estimators when the rank condition holds, below. Proofs are provided in Appendix D.

Corollary 1. *Consider the panel data models (1) and (2). Assume that the rank condition, (9), is met and suppose that Assumptions 1–6 hold. Then, for the Common Correlated Effects Mean Group estimator, $\hat{\boldsymbol{b}}_{MG}$, defined by (14), we have, as $(N, T) \xrightarrow{j} \infty$, that*

$$\sqrt{N}(\hat{\boldsymbol{b}}_{MG} - \boldsymbol{\beta}) \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{\Sigma}_{MG}),$$

where $\boldsymbol{\Sigma}_{MG}$ is given by $\boldsymbol{\Omega}_{\varkappa}$. $\boldsymbol{\Sigma}_{MG}$ can be consistently estimated by (38).

Corollary 2. *Consider the panel data models (1) and (2), and suppose that the rank condition, (9), is met and that Assumptions 1–6 hold. Then, for the Common Correlated Effects Pooled estimator, $\hat{\boldsymbol{b}}_P$, defined by (20), as $(N, T) \xrightarrow{j} \infty$, we have that*

$$\sqrt{N}(\hat{\boldsymbol{b}}_P - \boldsymbol{\beta}) \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{\Sigma}_P^*),$$

where

$$\boldsymbol{\Sigma}_P^* = \boldsymbol{\Psi}^{*-1}\boldsymbol{R}^*\boldsymbol{\Psi}^{*-1}, \tag{45}$$

$$\boldsymbol{R}^* = \lim_{N,T\to\infty}\left[N^{-1}\sum_{i=1}^{N}\boldsymbol{\Sigma}_{\boldsymbol{v}\boldsymbol{\Omega}iT}\right], \tag{46}$$

$$\boldsymbol{\Psi}^* = \lim_{N\to\infty}\left(N^{-1}\sum_{i=1}^{N}\boldsymbol{\Sigma}_{\boldsymbol{v}_i}\right), \tag{47}$$

and $\boldsymbol{\Sigma}_{\boldsymbol{v}\boldsymbol{\Omega}iT}$ denotes the variance of $\frac{\boldsymbol{X}_i'\boldsymbol{M}_g\boldsymbol{X}_i}{T}\boldsymbol{\varkappa}_i$. $\boldsymbol{\Sigma}_P^*$ can be estimated consistently by (42).

3.2. Estimation of individual slope coefficients

In panel data models where N is large, the estimation of the individual slope coefficients is likely to be of secondary importance as compared to establishing the properties of pooled estimators. However, it might still be of interest to consider conditions under which they can be consistently estimated. In the case of our set-up, the following further assumption is needed.

Assumption 8. For each i, ε_{it} is a martingale difference sequence. For each i, $\boldsymbol{v}_{it}$ is a $k \times 1$ vector of $L_{2+\delta}$, $\delta > 0$, stationary near epoque dependent (NED) processes of size $1/2$, on some α-mixing process of size $-(2+\delta)/\delta$.

Then, we have the following result. The proof is provided in Appendix E.

Theorem 3. *Consider the panel data models (1) and (2) and suppose that Assumptions 1, 2(i) and 3–8 hold. Let $\sqrt{T}/N \to 0$, as $(N, T) \xrightarrow{j} \infty$, and assume that the rank condition (9) is satisfied. As $(N, T) \xrightarrow{j} \infty$, $\hat{\boldsymbol{b}}_i$, defined by (15), is a consistent estimator of $\boldsymbol{\beta}_i$. Further,*

$$\sqrt{T}(\hat{\boldsymbol{b}}_i - \boldsymbol{\beta}_i) \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{\Sigma}_{b_i}). \tag{48}$$

A consistent estimator of $\boldsymbol{\Sigma}_{b_i}$ is given by

$$\hat{\boldsymbol{\Sigma}}_{b_i} = \mathring{\sigma}_i^2\left(\frac{\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{X}_i}{T}\right)^{-1}, \tag{49}$$

where

$$\mathring{\sigma}_i^2 = \frac{\left(\boldsymbol{y}_i - \boldsymbol{X}_i\hat{\boldsymbol{b}}_i\right)'\bar{\boldsymbol{M}}(\boldsymbol{y}_i - \boldsymbol{X}_i\hat{\boldsymbol{b}}_i)}{T - (n + 2k + 1)}. \tag{50}$$

Remark 5. Parts of the above result hold under weaker versions of Assumption 8. In particular, we note that the central limit theorem in (E.110) holds if Assumption 2(ii) holds. However, in this case, the asymptotic variance has a different form, as autocovariances of $\varepsilon_{it}\boldsymbol{v}_{it}$ enter the asymptotic variance expression. If, then, a consistent estimate of the asymptotic variance is required, a Newey and West type correction (Newey and West, 1987) needs to be used. Consistency of this variance estimator requires more stringent assumptions than the NED Assumption 2(ii). It is sufficient to assume that $(\varepsilon_{it}, \boldsymbol{v}_{it}')'$ is a strongly mixing process for this consistency to hold.

Remark 6. It is worth noting that despite the fact that, under our assumptions, $\boldsymbol{f}_t$, y_{it} and $\boldsymbol{x}_{it}$ are $I(1)$ and cointegrated, implying that ε_{it} is an $I(0)$ process, in the results of Theorem 3, the rate of convergence of $\hat{\boldsymbol{b}}_i$ to $\boldsymbol{\beta}_i$ as $(N, T) \xrightarrow{j} \infty$ is $\sqrt{T}$ and not T. It is helpful to develop some intuition behind this result. Since for N sufficiently large $\boldsymbol{f}_t$ can be well approximated by the cross-section averages, for pedagogic purposes we might as well consider the case where $\boldsymbol{f}_t$ is observed. Without loss of generality, we also abstract from $\boldsymbol{d}_t$, and substitute (2) in (1) to obtain

$$y_{it} = \boldsymbol{\beta}_i'\left(\boldsymbol{\Gamma}_i'\boldsymbol{f}_t + \boldsymbol{v}_{it}\right) + \boldsymbol{\gamma}_i'\boldsymbol{f}_t + \varepsilon_{it} = \boldsymbol{\vartheta}_i'\boldsymbol{f}_t + \zeta_{it}, \quad (51)$$

where $\boldsymbol{\vartheta}_i = \boldsymbol{\Gamma}_i\boldsymbol{\beta}_i + \boldsymbol{\gamma}_i$ and $\zeta_{it} = \varepsilon_{it} + \boldsymbol{\beta}_i'\boldsymbol{v}_{it}$. First, it is clear that, under our assumptions, and for all values of $\boldsymbol{\beta}_i$, ζ_{it} is $I(0)$ irrespective of whether $\boldsymbol{f}_t$ is $I(0)$ or $I(1)$. But, if $\boldsymbol{f}_t$ is $I(1)$, since $\zeta_{it} \backsim I(0)$, then y_{it} will also be $I(1)$ and cointegrated with $\boldsymbol{f}_t$. Hence, it follows that $\boldsymbol{\vartheta}_i$ can be estimated superconsistently. However, the ordinary least squares (OLS) estimator of $\boldsymbol{\beta}_i$ need not be superconsistent. To see this, note that $\boldsymbol{\beta}_i$ can be estimated equivalently by regressing the residuals from the regressions of y_{it} on $\boldsymbol{f}_t$ on the residuals from the regressions of $\boldsymbol{x}_{it}$ on $\boldsymbol{f}_t$. Both these sets of residuals are stationary processes, and the resulting estimator of $\boldsymbol{\beta}_i$ will be at most $\sqrt{T}$-consistent.

Remark 7. An issue related to the above remark concerns the probability limit of the OLS estimator of the coefficients of $\boldsymbol{x}_{it}$ in a regression of y_{it} on $\boldsymbol{x}_{it}$ alone. In general, such a regression will be subject to the omitted variable problem and hence misspecified. Also, the asymptotic properties of such OLS estimators cannot be derived without further assumptions. However, there is a special case which illustrates the utility of our method. Abstracting from $\boldsymbol{d}_t$, assuming that $k = m$ and that $\boldsymbol{\Gamma}_i$ is invertible, then, similarly to (51), write the model for y_{it} as

$$y_{it} = \boldsymbol{\beta}_i'\boldsymbol{x}_{it} + \boldsymbol{\gamma}_i'\boldsymbol{\Gamma}_i'^{-1}(\boldsymbol{x}_{it} - \boldsymbol{v}_{it}) + \varepsilon_{it} = \boldsymbol{\varrho}_i'\boldsymbol{x}_{it} + \varsigma_{it}, \quad (52)$$

where $\boldsymbol{\varrho}_i' = \boldsymbol{\beta}_i' + \boldsymbol{\gamma}_i'\boldsymbol{\Gamma}_i'^{-1}$ and $\varsigma_{it} = \varepsilon_{it} - \boldsymbol{\gamma}_i'\boldsymbol{\Gamma}_i'^{-1}\boldsymbol{v}_{it}$. Note that ς_{it} is, by construction, correlated with $\boldsymbol{v}_{it}$. The question is whether estimating a regression of the form (52) provides a consistent estimate of $\boldsymbol{\varrho}_i$. For stationary processes this would not be case, due to the correlation between ς_{it} and $\boldsymbol{v}_{it}$. However, in the case of non-stationary data this is not clear, and consistency would depend on the exact specification of the model. Under the assumptions we have made in this remark, the estimator of $\boldsymbol{\varrho}_i$ would be consistent. However, even in this case it is clear that the application of the least squares method to (52) can only lead to a consistent estimator of $\boldsymbol{\varrho}_i$ and not of $\boldsymbol{\beta}_i$. To consistently estimate the latter we need to augment the regressions of y_{it} on $\boldsymbol{x}_{it}$ with their cross-section averages.

4. Monte Carlo design and evidence

In this section, we provide Monte Carlo evidence on the small-sample properties of the CCEMG and the CCEP estimators, which are defined by (14) and (20), respectively. We consider nine alternative estimators. The first one is the CupBC estimator proposed by Bai et al. (2009), which is a bias-corrected version of a continuously updated estimator that estimates both the slope parameters and the unobserved factors iteratively. The CupBC estimator, as analyzed by Bai et al. (2009), assumes that the number of unobserved factors is known and only considers the case where the slopes are homogeneous.[4] In addition, we consider two alternative principal component (PC) augmentation approaches discussed in Kapetanios and Pesaran (2007). The first PC approach applies the Bai and Ng (2002) procedure to $\boldsymbol{z}_{it} = (y_{it}, \boldsymbol{x}_{it}')'$ to obtain consistent estimates of the unobserved factors, and then uses the estimated factors to augment the regression (1), and thus produces consistent estimator of $\boldsymbol{\beta}$. We consider both pooled and mean group versions of this estimator, which we refer to as *PC1POOL* and *PC1MG*. The second PC approach begins with extracting the principal component estimates of the unobserved factors from y_{it} and $\boldsymbol{x}_{it}$ separately. In the second step, y_{it} and $\boldsymbol{x}_{it}$ are regressed on their respective factor estimates, and in the third step the residuals from these regressions are used to compute the standard pooled and mean group estimators, with no cross-sectional dependence adjustments. We refer to the estimators based on this approach as *PC2POOL* and *PC2MG*, respectively. On top of these principal component estimators, we consider two sets of benchmark estimators. The first set consists of infeasible mean group and pooled estimators, which are obtained assuming that the factors are observable (i.e., $\bar{\boldsymbol{z}}_t$ for the CCE estimators is replaced by true factor $\boldsymbol{f}_t$). The other set consists of naive mean group and pooled estimators, which ignore the factor structure. The naive estimators are expected to illustrate the extent of bias and size distortions that can occur if the error cross-section dependence that is induced by the factor structure is ignored.

We report summaries of the performance of the estimators in the Monte Carlo experiments in terms of average biases, root mean square errors and rejection probabilities of the t-test for slope parameters under both the null hypothesis and an alternative hypothesis. For computing the t-statistics, the standard errors of mean group and pooled CCE estimators are estimated using (38) and (42), respectively. The standard errors of PC1, PC2, infeasible and naive estimators are estimated similarly to those of the CCE estimators. The standard errors of the CupBC estimator are computed following Bai et al. (2009).

4.1. Baseline design

The experimental design of the Monte Carlo study closely follows the one used in Pesaran (2006). Consider the following data generating process (DGP):

$$y_{it} = \alpha_{i1}d_{1t} + \beta_{i1}x_{1it} + \beta_{i2}x_{2it} + \gamma_{i1}f_{1t} + \gamma_{i2}f_{2t} + \varepsilon_{it}, \quad (53)$$

and

$$x_{ijt} = a_{ij1}d_{1t} + a_{ij2}d_{2t} + \gamma_{ij1}f_{1t} + \gamma_{ij3}f_{3t} + \mathrm{v}_{ijt}, \quad j = 1, 2, \quad (54)$$

for $i = 1, 2, \ldots, N$, and $t = 1, 2, \ldots, T$. This DGP is a restricted version of the general linear model considered in Pesaran (2006), and sets $n = k = 2$, and $m = 3$, with $\boldsymbol{\alpha}_i' = (\alpha_{i1}, 0)$, $\boldsymbol{\beta}_i' = (\beta_{i1}, \beta_{i2})$, and $\boldsymbol{\gamma}_i' = (\gamma_{i1}, \gamma_{i2}, 0)$, and

$$\boldsymbol{A}_i' = \begin{pmatrix} a_{i11} & a_{i12} \\ a_{i21} & a_{i22} \end{pmatrix}, \qquad \boldsymbol{\Gamma}_i' = \begin{pmatrix} \gamma_{i11} & 0 & \gamma_{i13} \\ \gamma_{i21} & 0 & \gamma_{i23} \end{pmatrix}.$$

The observed common factors and the individual-specific errors of $\boldsymbol{x}_{it}$ are generated as independent stationary AR(1) processes with zero means and unit variances:

$$d_{1t} = 1, \qquad d_{2t} = \rho_d d_{2,t-1} + v_{dt}, \quad t = -49, \ldots, 1, \ldots, T,$$
$$v_{dt} \sim \mathrm{IIDN}(0, 1 - \rho_d^2), \quad \rho_d = 0.5, \; d_{2,-50} = 0,$$
$$\mathrm{v}_{ijt} = \rho_{vij}\mathrm{v}_{ijt-1} + \varkappa_{ijt}, \quad t = -49, \ldots, 1, \ldots, T,$$
$$\varkappa_{ijt} \sim \mathrm{IIDN}\left(0, 1 - \rho_{vij}^2\right), \quad \mathrm{v}_{ji,-50} = 0,$$

and

$$\rho_{vij} \sim \mathrm{IIDU}[0.05, 0.95], \quad \text{for } j = 1, 2.$$

But the unobserved common factors are generated as non-stationary processes:

$$f_{jt} = f_{jt-1} + v_{fj,t}, \quad \text{for } j = 1, 2, 3, \; t = -49, \ldots, 0, \ldots, T, \quad (55)$$
$$v_{fj,t} \sim \mathrm{IIDN}(0, 1), \qquad f_{j,-50} = 0, \quad \text{for } j = 1, 2, 3.$$

The first 50 observations are discarded.

[4] See Bai et al. (2009), for more details.

To illustrate the robustness of the CCE estimators and others to the dynamics of the individual-specific errors of y_{it}, these are generated as the (cross-sectional) mixture of stationary heterogeneous AR(1) and MA(1) errors. Namely,

$$\varepsilon_{it} = \rho_{i\varepsilon}\varepsilon_{i,t-1} + \sigma_i\sqrt{1-\rho_{i\varepsilon}^2}\omega_{it},$$
$$i = 1, 2, \ldots, N_1,\ t = -49, \ldots, 0, \ldots, T,$$

and

$$\varepsilon_{it} = \frac{\sigma_i}{\sqrt{1+\theta_{i\varepsilon}^2}}(\omega_{it} + \theta_{i\varepsilon}\omega_{i,t-1}),$$
$$i = N_1 + 1, \ldots, N,\ t = -49, \ldots, 0, \ldots, T,$$

where N_1 is the nearest integer to $N/2$,

$$\omega_{it} \sim \text{IIDN}(0, 1), \qquad \sigma_i^2 \sim \text{IIDU}[0.5, 1.5],$$
$$\rho_{i\varepsilon} \sim \text{IIDU}[0.05, 0.95], \qquad \theta_{i\varepsilon} \sim \text{IIDU}[0, 1].$$

$\rho_{vij}, \rho_{i\varepsilon}, \theta_{i\varepsilon}$ and σ_i are not changed across replications. The first 49 observations are discarded. The factor loadings of the observed common effects, α_{i1} and $\text{vec}(\boldsymbol{A}_i) = (a_{i11}, a_{i21}, a_{i12}, a_{i22})'$ are generated as IIDN(1, 1) and $\text{IIDN}(0.5\boldsymbol{\tau}_4, 0.5\boldsymbol{I}_4)$ with $\boldsymbol{\tau}_4 = (1, 1, 1, 1)'$, respectively, which are not changed across replications. The parameters of the unobserved common effects in the $\boldsymbol{x}_{it}$ equation are generated independently across replications as

$$\boldsymbol{\Gamma}_i' = \begin{pmatrix} \gamma_{i11} & 0 & \gamma_{i13} \\ \gamma_{i21} & 0 & \gamma_{i23} \end{pmatrix} \sim \text{IID}\begin{pmatrix} N(0.5, 0.50) & 0 & N(0, 0.50) \\ N(0, 0.50) & 0 & N(0.5, 0.50) \end{pmatrix}.$$

For the parameters of the unobserved common effects in the y_{it} equation, $\boldsymbol{\gamma}_i$, we considered two different sets, which we denote by $\mathcal{A}$ and $\mathcal{B}$. Under set $\mathcal{A}$, the $\boldsymbol{\gamma}_i$ are drawn such that the rank condition is satisfied, namely

$$\gamma_{i1} \sim \text{IIDN}(1, 0.2), \qquad \gamma_{i2\mathcal{A}} \sim \text{IIDN}(1, 0.2), \qquad \gamma_{i3} = 0,$$

and

$$E\left(\tilde{\boldsymbol{\Gamma}}_{i\mathcal{A}}\right) = \left(E(\boldsymbol{\gamma}_{i\mathcal{A}}), \quad E(\boldsymbol{\Gamma}_i)\right) = \begin{pmatrix} 1 & 0.5 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0.5 \end{pmatrix}.$$

Under set $\mathcal{B}$,

$$\gamma_{i1} \sim \text{IIDN}(1, 0.2), \qquad \gamma_{i2\mathcal{B}} \sim \text{IIDN}(0, 1), \qquad \gamma_{i3} = 0,$$

so

$$E\left(\tilde{\boldsymbol{\Gamma}}_{i\mathcal{B}}\right) = \left(E(\boldsymbol{\gamma}_{i\mathcal{B}}), \quad E(\boldsymbol{\Gamma}_i)\right) = \begin{pmatrix} 1 & 0.5 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0.5 \end{pmatrix},$$

and the rank condition is *not* satisfied. For each set, we conducted two different experiments.

- *Experiment* 1 examines the case of heterogeneous slopes with $\beta_{ij} = 1 + \eta_{ij}, j = 1, 2$, and $\eta_{ij} \sim \text{IIDN}(0, 0.04)$, across replications.
- *Experiment* 2 considers the case of homogeneous slopes with $\boldsymbol{\beta}_i = \boldsymbol{\beta} = (1, 1)'$.

The two versions of experiment 1 will be denoted by $1\mathcal{A}$ and $1\mathcal{B}$, and those of experiment 2 by $2\mathcal{A}$ and $2\mathcal{B}$.

Concerning the infeasible pooled estimator, it is important to note that, although this estimator is unbiased under all four sets of experiments, it need not be efficient, since in these experiments the slope coefficients, $\boldsymbol{\beta}_i$, and/or error variances, σ_i^2, differ across i. As a result, the CCE or PC augmented estimators may in fact dominate the infeasible estimator in terms of root mean square error (RMSE), particularly in the case of experiments $1\mathcal{A}$ and $1\mathcal{B}$, where the slopes as well as the error variances are allowed to vary across i.

Another important consideration worth bearing in mind when comparing the CCE and the principal component type estimators is the fact that the computation of the CupBC, PC1 and PC2 estimators assumes that $m = 3$, namely that the number of unobserved factors is known. In practice, m might be difficult to estimate accurately, particularly when N or T happen to be smaller than 50. By contrast, the CCE type estimators are valid for any fixed m and do not require an *a priori* estimate for m.

Each experiment was replicated 2000 times for the (N, T) pairs with $N, T = 20, 30, 50, 100, 200$. In what follows, we shall focus on β_1 (the cross-section mean of β_{i1}), and the results for β_2, which are very similar to those for β_1, will not be reported. The results for all the estimators considered are reported in Table 1. Since the performance of CCE and CupBC estimators dominates other feasible estimators in most of the designs considered, to save space we do not report the results of these estimators for the remaining experiments.

4.2. *Designs for robustness checks*

In this subsection, we consider a number of Monte Carlo experiment designs that aim to check the robustness of the estimators to a variety of empirical settings.

4.2.1. *The number of factors exceeds $k + 1$*

In order to show the effect of a different type of violation of the rank condition from experiment $\mathcal{B}$, we consider the DGP $1\mathcal{A}$, but an extra factor term $\gamma_{i4}f_{4t}$ is added to the right-hand side (RHS) of the y equation (53), where $\gamma_{i4} \sim \text{IIDN}(0.5, 0.2), f_{4t} = f_{4t-1} + v_{f4,t}, v_{f4,t} \sim \text{IIDN}(0, 1), f_{4,-50} = 0$. In this case, observe that

$$E(\boldsymbol{\gamma}_i, \boldsymbol{\Gamma}_i)' = \begin{pmatrix} 1 & 1 & 0 & 0.5 \\ 0.5 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0 \end{pmatrix}$$

whose rank is $k + 1 = 3$, which is less than the number of unobserved factors, $m = 4$. Under this experiment, the number of factors is treated as unknown and is estimated, using the information criterion 'PC_{p2}' which is proposed by Bai and Ng (2002, pp. 201).[5] The information criterion is applied to the first differenced variables with the maximum number of factors set to six. The results are reported in Table 5. However, recall that the CCE type estimators do not make use of the number of the factors and are valid irrespective of whether $k + 1$ is more or less than m.

4.2.2. *Cointegrating factors*

In this design, the unobserved common factors are generated as cointegrated non-stationary processes. There are two underlying stochastic trends, given by

$$f_{jt}^t = f_{jt-1}^t + v_{fj,t}^t, \quad \text{for } j = 1, 2,\ t = -49, \ldots, 0, \ldots, T, \tag{56}$$
$$v_{fj,t}^t \sim \text{IIDN}(0, 1), \qquad f_{j,-50}^t = 0, \quad \text{for } j = 1, 2.$$

Then, this experiment uses the same design as $1\mathcal{A}$, but the $I(1)$ factors in (53) and (54) are replaced by

$$f_{1t} = f_{1t}^t + 0.5f_{2t}^t + v_{f1,t}, \quad t = -49, \ldots, 0, \ldots, T,$$
$$f_{2t} = 0.5f_{1t}^t + f_{2t}^t + v_{f2,t}, \quad t = -49, \ldots, 0, \ldots, T,$$
$$f_{3t} = 0.75f_{1t}^t + 0.25f_{2t}^t + v_{f3,t}, \quad t = -49, \ldots, 0, \ldots, T,$$
$$v_{fj,t} \sim \text{IIDN}(0, 1), \qquad f_{j,-50} = 0, \quad \text{for } j = 1, 2, 3.$$

The first 50 observations are discarded. The results are reported in Table 6.

[5] PC_{p2} is one of the information criteria which performed well in the finite sample investigations reported in Bai and Ng (2002).

Table 1
Small-sample properties of common correlated effects type estimators in the case of experiment 1$\mathcal{A}$ (heterogeneous slopes + full rank).

	Bias (×100)					Root mean square error (×100)					Size (5% level, $H_0:\beta_1=1.00$)					Power (5% level, $H_1:\beta_1=0.95$)				
(N, T)	20	30	50	100	200	20	30	50	100	200	20	30	50	100	200	20	30	50	100	200
CCE type estimators																				
	CCEMG																			
20	0.05	−0.10	−0.03	0.06	−0.07	9.67	7.89	6.74	5.87	5.54	7.20	6.90	7.15	7.90	7.55	11.65	13.00	16.10	17.50	20.10
30	0.09	−0.01	−0.01	−0.13	0.10	7.69	6.09	5.11	4.54	4.22	6.95	5.30	5.90	6.25	6.35	11.40	14.25	18.05	22.05	26.85
50	−0.19	0.22	−0.11	0.14	−0.04	5.88	4.61	4.01	3.44	3.13	5.70	5.05	6.65	6.20	5.95	15.10	20.40	25.60	34.10	36.65
100	0.00	0.04	0.04	0.03	0.04	4.25	3.46	2.89	2.33	2.27	5.75	5.85	5.25	4.90	6.20	23.35	34.30	44.40	56.00	63.25
200	−0.05	−0.02	−0.03	0.05	0.00	3.07	2.49	2.01	1.72	1.51	4.40	5.15	4.90	5.60	5.10	35.55	52.65	68.70	83.65	90.50
	CCEP																			
20	0.18	0.00	−0.05	−0.01	−0.13	8.75	7.67	6.85	6.32	6.21	7.70	8.10	7.30	8.05	7.15	12.75	13.50	16.05	16.80	18.30
30	−0.17	−0.12	0.09	−0.15	0.13	7.10	5.99	5.32	4.78	4.46	7.55	6.25	6.75	6.65	6.45	12.40	15.00	19.30	20.65	26.90
50	0.00	0.18	−0.07	0.12	−0.01	5.33	4.51	3.97	3.47	3.22	6.80	6.20	5.90	6.35	6.45	17.45	22.15	26.40	32.90	36.25
100	0.00	0.09	0.03	0.00	0.02	3.78	3.25	2.85	2.34	2.28	5.70	5.65	5.60	5.15	6.25	28.15	37.40	44.80	55.20	61.75
200	−0.07	−0.04	−0.05	0.05	0.00	2.71	2.29	1.95	1.70	1.53	5.10	4.35	5.05	4.70	4.75	44.75	56.80	70.30	83.55	89.75
Bai, Kao and Ng principal component estimator																				
	CupBC																			
20	0.62	0.70	0.81	0.77	0.87	11.16	9.86	8.35	7.46	6.95	67.25	64.40	57.90	60.95	65.75	72.05	68.75	65.00	68.50	74.60
30	0.35	0.42	0.73	0.59	0.83	8.91	7.70	6.51	5.66	5.28	66.80	61.05	55.95	55.40	63.30	71.85	69.95	66.35	70.75	77.60
50	0.53	0.67	0.33	0.63	0.54	6.77	6.01	5.05	4.20	3.83	64.45	58.85	51.95	51.35	56.70	77.20	72.65	69.55	76.90	83.45
100	0.21	0.34	0.35	0.28	0.33	4.83	4.15	3.39	2.76	2.55	64.60	56.40	47.85	43.35	52.65	80.70	80.35	82.50	87.90	92.50
200	0.10	0.10	0.08	0.23	0.17	3.55	2.94	2.45	2.00	1.69	62.85	52.85	45.00	44.50	48.00	86.65	88.70	90.75	96.60	99.10
Infeasible estimators (including f_{1t} and f_{2t})																				
	Infeasible MG																			
20	0.01	−0.19	−0.08	0.15	−0.08	7.21	6.33	5.62	4.98	4.76	6.40	6.20	6.80	5.95	6.50	12.75	15.35	16.85	19.70	20.40
30	0.02	−0.14	0.01	−0.02	0.12	5.91	4.95	4.43	3.97	3.87	6.50	5.80	6.05	5.30	5.90	16.15	18.05	23.35	25.20	28.80
50	−0.10	0.07	−0.06	0.14	−0.04	4.48	3.75	3.39	3.09	2.94	6.45	5.25	5.90	5.25	5.20	21.70	27.35	31.45	38.45	40.25
100	0.01	0.07	0.02	0.00	0.04	3.16	2.78	2.49	2.15	2.14	5.50	5.15	5.45	4.70	5.45	36.85	46.15	55.10	62.50	66.65
200	−0.07	0.04	−0.07	0.06	0.01	2.22	1.93	1.69	1.57	1.44	4.85	5.00	5.00	5.60	4.70	59.15	72.85	82.25	90.40	92.75
	Infeasible pooled																			
20	0.15	−0.13	−0.15	−0.26	−0.21	7.30	6.96	6.92	7.11	7.40	6.40	6.80	6.60	7.00	5.10	13.70	13.75	14.55	14.10	12.65
30	−0.20	−0.15	0.22	−0.07	0.27	6.23	5.78	5.79	5.89	6.61	7.05	5.90	7.00	5.25	5.70	15.70	15.35	18.95	16.70	16.60
50	0.12	0.07	−0.08	0.21	0.02	4.61	4.40	4.31	4.71	5.02	5.70	5.80	5.50	6.25	5.00	22.20	22.55	23.65	25.50	21.00
100	−0.05	0.07	0.09	0.06	0.00	3.30	3.26	3.12	3.30	3.52	5.25	5.60	5.20	5.20	5.30	33.45	38.20	38.85	36.75	32.30
200	−0.08	0.06	−0.12	0.07	−0.02	2.35	2.22	2.20	2.45	2.49	4.95	4.70	4.50	5.85	4.70	56.15	62.10	59.50	59.05	52.20
Naive estimators (excluding f_{1t} and f_{2t})																				
	Naive MG																			
20	22.18	23.13	26.82	29.96	32.62	31.76	32.97	37.37	41.49	47.04	32.05	32.95	34.85	35.45	31.50	41.00	42.65	43.50	41.95	38.05
30	22.23	25.06	28.36	31.33	34.01	30.51	33.31	37.87	41.46	45.32	40.45	44.10	46.65	43.85	39.45	51.00	53.95	57.45	52.20	47.15
50	22.21	23.91	25.65	29.61	33.64	29.75	31.12	32.75	37.73	42.66	55.80	59.30	58.00	59.25	54.75	68.30	70.85	70.30	69.20	65.05
100	21.97	23.92	26.76	30.04	32.88	28.40	30.02	32.97	36.39	40.06	71.20	75.25	77.90	78.60	75.25	81.05	84.35	85.95	85.85	83.20
200	22.15	24.09	27.49	30.09	33.23	27.87	29.44	32.80	35.71	39.34	81.85	86.00	87.85	88.05	87.95	88.75	91.95	92.30	92.90	92.05
	Naive pooled																			
20	25.25	26.60	31.27	33.59	34.84	35.30	37.01	42.66	45.42	47.67	42.15	43.65	47.75	45.20	44.50	52.50	52.65	55.95	53.40	51.95
30	25.76	29.39	32.45	35.37	35.46	35.48	39.13	42.70	45.97	46.81	51.55	56.70	57.65	59.55	56.20	61.05	66.60	66.55	67.75	64.55
50	26.54	28.75	30.39	34.01	35.88	35.61	37.39	39.05	44.04	45.93	64.75	67.15	69.25	70.35	69.35	73.55	76.25	78.25	78.65	77.45
100	25.81	28.47	31.30	33.15	34.91	34.39	36.76	39.90	41.79	44.27	75.85	78.90	81.35	79.30	80.15	85.10	86.55	88.05	86.65	86.40
200	25.95	28.32	31.89	33.65	34.11	34.20	36.21	39.63	42.39	42.68	83.45	86.25	87.70	87.40	87.20	89.95	91.90	93.55	92.20	92.20
Principal component estimators, augmented																				
	PC1MG																			
20	−12.27	−11.15	−10.30	−8.87	−8.90	17.09	14.81	13.24	11.51	11.55	22.55	25.35	30.05	33.40	37.40	12.15	12.95	13.30	12.70	13.75
30	−9.25	−7.86	−6.46	−5.72	−5.25	13.55	10.84	8.98	7.80	7.15	20.60	20.90	21.65	24.75	24.70	10.75	8.25	7.35	7.40	6.75
50	−6.84	−5.05	−3.89	−3.01	−3.12	10.10	7.79	5.86	4.67	4.47	19.95	17.65	16.25	14.95	17.90	8.70	8.20	7.65	11.40	9.75
100	−4.78	−3.21	−2.03	−1.57	−1.45	7.44	5.34	3.68	2.87	2.72	20.10	16.80	11.45	9.75	11.10	9.55	12.15	20.25	28.85	36.75
200	−4.31	−2.54	−1.39	−0.81	−0.78	6.39	4.19	2.60	1.93	1.71	25.20	17.95	10.95	8.15	7.65	13.85	21.95	42.85	67.65	77.15
	PC1POOL																			
20	−11.97	−11.04	−10.35	−9.09	−9.23	15.88	14.38	13.07	11.59	12.07	25.50	28.35	32.05	34.45	38.95	12.05	14.10	14.90	14.55	14.90
30	−8.86	−7.66	−6.34	−5.73	−5.37	12.48	10.45	8.89	7.80	7.34	21.45	23.75	22.05	24.70	25.50	11.00	8.80	7.55	7.95	6.35
50	−6.20	−4.86	−3.81	−3.07	−3.19	9.06	7.52	5.72	4.73	4.54	21.40	18.75	16.00	16.05	18.90	8.55	9.55	8.10	10.90	9.65
100	−4.36	−3.00	−2.01	−1.60	−1.49	6.61	5.01	3.61	2.88	2.74	21.05	16.85	11.25	9.35	10.80	11.25	14.55	20.85	27.90	36.30
200	−3.62	−2.32	−1.36	−0.81	−0.79	5.39	3.81	2.51	1.91	1.73	25.15	17.60	10.50	7.80	7.80	16.35	26.75	45.45	68.00	76.15

(*continued on next page*)

4.2.3. *Semi-strong factor structure*

Chudik et al. (forthcoming) introduce the notions of weak, semi-strong and strong factor structures and prove that these different factor structures do not affect the consistency of the CCE type estimators with $I(0)$ factors. Here we consider the effect of having a semi-strong factor structure when the factors are $I(1)$. For this purpose, the same DGP of the experiment 1$\mathcal{A}$ is used, but all factor loadings in (53) and (54) are multiplied by $N^{-1/2}$. The results are reported in Table 7. It is easily seen that when the factors are weak or semi-strong they cannot be consistently estimated by

Table 1 (*continued*)

(N, T)	Bias ($\times 100$)					Root mean square error ($\times 100$)					Size (5% level, $H_0 : \beta_1 = 1.00$)					Power (5% level, $H_1 : \beta_1 = 0.95$)				
	20	30	50	100	200	20	30	50	100	200	20	30	50	100	200	20	30	50	100	200
Principal component estimators, orthogonalized																				
	PC2MG																			
20	−31.26	−27.06	−24.01	−22.67	−23.11	32.83	28.34	25.00	23.44	23.83	86.50	88.45	91.25	95.20	97.40	74.10	73.95	75.80	82.05	88.20
30	−25.50	−21.21	−18.27	−16.69	−16.33	26.82	22.25	19.13	17.35	16.92	86.85	87.10	89.10	93.35	95.95	70.15	67.80	66.10	69.25	74.70
50	−20.65	−16.23	−13.32	−11.41	−10.89	21.68	17.06	13.98	11.95	11.37	90.15	88.35	88.80	89.05	91.70	70.80	60.25	52.20	45.80	46.10
100	−16.17	−12.44	−9.69	−7.61	−6.60	16.87	12.97	10.18	7.99	7.02	93.65	93.30	89.75	87.50	83.30	72.35	56.20	37.60	19.30	13.60
200	−14.61	−10.78	−8.12	−5.79	−4.59	15.11	11.19	8.45	6.08	4.85	98.95	97.85	95.45	90.75	83.75	79.65	60.20	33.30	10.00	6.75
	PC2POOL																			
20	−31.97	−27.47	−24.27	−23.18	−24.19	33.39	28.69	25.23	23.99	24.99	91.00	90.70	93.20	95.55	98.50	80.65	78.60	78.80	83.35	90.45
30	−26.32	−21.51	−18.24	−16.83	−16.75	27.53	22.48	19.13	17.51	17.37	91.35	90.40	89.70	93.35	96.15	78.50	71.80	66.65	70.65	76.90
50	−21.22	−16.35	−13.17	−11.35	−10.99	22.10	17.15	13.82	11.91	11.48	95.05	90.90	88.95	88.20	91.70	79.65	63.80	52.95	46.20	48.25
100	−16.77	−12.52	−9.62	−7.55	−6.60	17.43	13.06	10.11	7.95	7.03	97.95	95.05	90.50	86.45	82.30	80.90	60.80	38.10	18.30	14.25
200	−15.16	−10.91	−8.00	−5.66	−4.53	15.67	11.33	8.34	5.96	4.79	99.75	98.45	95.95	89.35	82.50	88.65	65.85	33.35	8.40	6.30

Notes: The DGP is $y_{it} = \alpha_{i1}d_{1t} + \beta_{i1}x_{1it} + \beta_{i2}x_{2it} + \gamma_{i1}f_{1t} + \gamma_{i2}f_{2t} + \varepsilon_{it}$, with $\varepsilon_{it} = \rho_{i\varepsilon}\varepsilon_{i,t-1} + \sigma_i(1-\rho_{i\varepsilon}^2)^{1/2}\omega_{it}$, $i = 1, 2, \dots, [N/2]$, and $\varepsilon_{it} = \sigma_i(1+\theta_{i\varepsilon}^2)^{-1/2}(\omega_{it} + \theta_{i\varepsilon}\omega_{i,t-1})$, $i = [N/2]+1, \dots, N$, $\omega_{it} \sim \text{IIDN}(0, 1)$, $\sigma_i^2 \sim \text{IIDU}[0.5, 1.5]$, $\rho_{i\varepsilon} \sim \text{IIDU}[0.05, 0.95]$, $\theta_{i\varepsilon} \sim \text{IIDU}[0, 1]$. Regressors are generated by $x_{ijt} = a_{ij1}d_{1t} + a_{ij2}d_{2t} + \gamma_{ij1}f_{1t} + \gamma_{ij3}f_{3t} + v_{ijt}$, $j = 1, 2$, for $i = 1, 2, \dots, N$. $d_{1t} = 1$, $d_{2t} = 0.5d_{2,t-1} + v_{dt}$, $v_{dt} \sim \text{IIDN}(0, 1 - 0.5^2)$, $d_{2,-50} = 0$; $f_{jt} = f_{jt-1} + v_{fj,t}$, $v_{fj,t} \sim \text{IIDN}(0, 1)$, $f_{j,-50} = 0$, for $j = 1, 2, 3$; $v_{ijt} = \rho_{vij}v_{ijt-1} + \upsilon_{ijt}$, $\upsilon_{ijt} \sim \text{IIDN}(0, 1 - \rho_{vij}^2)$, $v_{ij,-50} = 0$ and $\rho_{vij} \sim \text{IIDU}[0.05, 0.95]$ for $j = 1, 2$, for $t = -49, \dots, T$, with the first 50 observations discarded; $\alpha_{i1} \sim \text{IIDN}(1, 1)$; $a_{ij\ell} \sim \text{IIDN}(0.5, 0.5)$ for $j = 1, 2$, $\ell = 1, 2$; γ_{i11} and $\gamma_{i23} \sim \text{IIDN}(0.5, 0.50)$, γ_{i13} and $\gamma_{i21} \sim \text{IIDN}(0, 0.50)$; γ_{i1} and $\gamma_{i2} \sim \text{IIDN}(1, 0.2)$; $\beta_{ij} = 1 + \eta_{ij}$, with $\eta_{ij} \sim \text{IIDN}(0, 0.04)$ for $j = 1, 2$. $\rho_{vij}, \rho_{i\varepsilon}, \theta_{i\varepsilon}, \sigma_i^2, \alpha_{i1}, a_{ij\ell}$ for $j = 1, 2$, $\ell = 1, 2$ are fixed across replications. CCEMG and CCEP are defined by (14) and (20). CupBC is the bias-corrected iterated principal component estimator of Bai et al. (2009). The PC1 and PC2 estimators are from Kapetanios and Pesaran (2007). The variance estimators of all mean group and pooled estimators (except that of CupBC) are defined by (38) and (42), respectively. The PC type estimators are computed assuming that the number of unobserved factors, $m = 3$, is known. All experiments are based on 2000 replications.

Table 2
Small-sample properties of common correlated effects type estimators in the case of experiment $2\mathcal{A}$ (homogeneous slopes + full rank).

(N, T)	Bias ($\times 100$)					Root mean square error ($\times 100$)					Size (5% level, $H_0 : \beta_1 = 1.00$)					Power (5% level, $H_1 : \beta_1 = 0.95$)				
	20	30	50	100	200	20	30	50	100	200	20	30	50	100	200	20	30	50	100	200
	CCEMG																			
20	0.05	−0.15	0.02	−0.15	0.09	8.45	6.29	5.10	3.78	3.14	7.15	6.40	6.80	6.75	6.85	11.70	13.80	21.75	31.25	47.90
30	−0.14	0.12	0.04	0.03	0.00	6.44	5.11	3.80	2.67	2.07	6.05	6.75	7.25	6.40	6.45	12.70	20.45	30.70	50.90	71.60
50	0.08	−0.06	0.02	0.05	0.03	5.08	3.79	2.80	1.94	1.39	6.10	5.90	4.85	5.40	5.35	18.00	26.90	44.45	75.65	95.00
100	−0.04	−0.08	0.06	−0.04	−0.01	3.59	2.76	2.02	1.35	0.98	4.55	5.50	6.05	5.10	6.10	28.30	43.00	72.35	95.20	99.90
200	0.06	−0.02	0.03	0.01	0.00	2.83	2.05	1.52	1.00	0.68	5.60	4.45	6.35	5.20	5.70	44.20	67.95	91.90	99.90	100.00
	CCEP																			
20	0.18	0.00	0.03	−0.14	0.08	6.95	5.56	4.94	3.98	3.74	6.60	6.75	7.30	6.75	6.80	14.25	16.25	25.25	33.70	46.25
30	−0.14	0.14	0.07	0.01	0.01	5.20	4.50	3.55	2.67	2.26	5.10	5.90	7.25	6.25	6.40	15.25	24.55	34.90	52.95	70.70
50	0.05	0.07	−0.02	0.04	0.03	4.08	3.29	2.56	1.84	1.39	5.40	5.40	5.45	6.20	5.30	24.60	34.35	51.70	78.65	95.00
100	−0.02	−0.04	0.06	−0.04	−0.01	2.87	2.37	1.78	1.24	0.93	5.60	6.20	6.40	5.25	5.95	41.65	58.35	81.85	97.80	100.00
200	0.07	−0.03	0.01	0.02	0.00	2.17	1.63	1.32	0.92	0.65	5.60	3.95	5.70	5.60	5.35	65.25	84.40	96.95	100.00	100.00
	CupBC																			
20	0.12	0.10	0.08	−0.01	0.01	8.25	6.13	4.14	2.32	1.29	64.00	52.40	38.20	25.15	18.85	70.40	66.65	65.90	84.75	98.35
30	0.04	0.08	0.07	0.02	−0.01	6.40	4.73	3.08	1.72	0.96	61.85	50.00	35.40	23.25	19.15	71.30	71.35	79.30	95.00	99.90
50	−0.04	0.22	−0.06	0.04	0.03	4.89	3.56	2.31	1.27	0.70	59.90	49.25	34.45	21.90	15.40	77.20	81.60	88.35	98.85	100.00
100	0.03	0.01	0.02	−0.05	0.01	3.27	2.43	1.66	0.86	0.48	60.30	48.40	34.40	20.25	17.15	87.15	91.65	97.40	100.00	100.00
200	0.07	0.01	0.03	0.03	0.00	2.43	1.73	1.16	0.63	0.33	59.95	46.60	32.60	20.70	14.90	94.70	97.70	99.80	100.00	100.00

Notes: The DGP is the same as that of Table 1, except that $\beta_{ij} = 1$ for all i and j, $i = 1, 2, \dots, N$, $j = 1, 2$. See notes to Table 1.

the principal components, and this could adversely impact the estimators of $\boldsymbol{\beta}$ that rely on the PCs as estimators of the unobserved factors.

4.2.4. *A structural break in the means of the unobserved factors*

Finally, the results of recent research by Stock and Watson (2008) suggest that the possible structural breaks in the means of the unobserved factors will not affect the consistency of the CCE type estimators, as well as the principal component type estimators. In view of this, we considered another set of experiments, corresponding to the DGPs specified as $1\mathcal{A}$, but now the unobserved factors are generated subject to mean shifts. Specifically, under these experiments the unobserved factors are generated as $f_{jt} = \varphi_{jt}$ for $t < [2T/3]$ and $f_{jt} = 1 + \varphi_{jt}$ for $t \geq [2T/3]$, with $[A]$ being the greatest integer less than or equal to A, where $\varphi_{jt} = \varphi_{j,t-1} + \zeta_{jt}$, and $\zeta_{jt} \sim \text{IIDN}(0, 1)$, for $j = 1, 2, 3$. Results are reported in Table 8.

4.3. *Results*

The results of experiments $1\mathcal{A}$, $2\mathcal{A}$, $1\mathcal{B}$, $2\mathcal{B}$ are summarized in Tables 1–4, respectively. We also provide results for the naive estimator (which excludes the unobserved factors or their estimates) and the infeasible estimator (which includes the unobserved factors as additional regressors) for comparison purposes. But for the sake of brevity we include the simulation results for these estimators only for experiment $1\mathcal{A}$.

As can be seen from Table 1, the naive estimator is substantially biased, performs very poorly, and is subject to large size distortions: this is an outcome that continues to apply in the case of other experiments (not reported here). In contrast, the feasible CCE estimators perform well, have biases that are close to the bias of the infeasible estimators, show little size distortions even for relatively small values of N and T, and their RMSE falls steadily with increases in N and/or T. These results are quite similar to the results

Table 3
Small-sample properties of common correlated effects type estimators in the case of experiment 1$\mathcal{B}$ (heterogeneous slopes + rank deficient).

	Bias (×100)					Root mean square error (×100)					Size (5% level, $H_0 : \beta_1 = 1.00$)					Power (5% level, $H_1 : \beta_1 = 0.95$)				
(N, T)	20	30	50	100	200	20	30	50	100	200	20	30	50	100	200	20	30	50	100	200
	CCEMG																			
20	0.33	−0.19	0.20	0.14	0.23	15.02	13.90	12.61	13.35	13.78	6.80	6.90	6.75	6.60	7.20	9.40	8.95	10.15	10.15	10.15
30	0.30	0.14	0.09	−0.17	0.35	12.91	12.03	10.70	10.07	10.59	5.50	6.80	5.25	6.15	4.80	8.40	10.05	9.45	10.35	11.65
50	−0.15	0.63	−0.20	−0.17	0.02	9.82	8.46	7.87	7.42	7.34	5.80	5.10	6.10	5.75	5.90	9.75	12.90	13.40	14.00	15.20
100	0.25	0.13	0.27	0.00	0.06	7.01	6.55	5.85	5.25	5.01	5.75	5.95	5.45	5.45	6.10	14.50	17.75	21.65	22.65	27.30
200	0.05	−0.11	−0.17	−0.07	−0.05	5.35	4.65	4.15	3.61	3.31	4.80	5.05	4.75	5.15	4.55	19.45	23.70	29.75	37.25	43.45
	CCEP																			
20	0.48	0.06	−0.04	0.16	0.10	13.13	12.81	12.21	13.57	15.30	6.75	7.40	7.00	6.65	6.75	9.90	10.20	10.40	10.35	10.25
30	−0.23	−0.06	0.18	−0.25	0.43	11.48	10.70	10.39	9.95	11.04	6.10	6.90	5.70	6.00	5.50	9.05	9.95	10.55	10.25	10.60
50	0.00	0.48	−0.18	−0.17	−0.02	8.42	7.57	7.23	7.22	7.22	5.25	5.90	6.25	5.30	5.50	11.40	14.05	14.15	14.35	15.20
100	0.11	0.18	0.24	−0.06	0.05	5.87	5.72	5.27	4.87	4.98	5.10	6.00	5.40	4.95	6.00	17.25	19.60	23.50	23.55	27.00
200	0.04	−0.10	−0.16	−0.04	−0.03	4.35	3.99	3.75	3.30	3.15	5.40	4.70	5.25	4.10	3.95	25.75	28.50	34.50	41.10	46.05
	CupBC																			
20	1.34	0.83	1.07	1.12	1.35	11.24	9.52	8.24	7.59	7.24	67.35	60.20	56.70	60.85	66.80	70.45	66.05	66.05	71.25	76.85
30	0.51	0.85	1.14	0.86	1.23	8.97	7.52	6.47	5.78	5.60	67.40	59.80	55.35	56.95	65.35	72.35	68.95	69.15	72.95	80.20
50	0.57	0.70	0.62	0.91	0.81	6.77	5.85	4.98	4.32	4.05	64.65	57.35	52.40	52.00	59.70	74.90	72.25	70.40	78.65	84.50
100	0.30	0.44	0.45	0.42	0.46	4.86	4.20	3.44	2.76	2.61	66.40	56.55	48.20	44.00	53.00	79.35	80.40	83.10	89.40	93.60
200	0.14	0.14	0.13	0.27	0.26	3.53	2.99	2.45	2.00	1.69	64.80	53.35	45.15	43.95	46.95	86.90	88.45	91.20	96.90	99.35

Notes: The DGP is the same as that of Table 1, except that $\gamma_{i2} \sim$ IIDN(0, 1), so the rank condition is not satisfied. See notes to Table 1.

Table 4
Small-sample properties of common correlated effects type estimators in the case of experiment 2$\mathcal{B}$ (homogeneous slopes + rank deficient).

	Bias (×100)					Root mean square error (×100)					Size (5% level, $H_0 : \beta_1 = 1.00$)					Power (5% level, $H_1 : \beta_1 = 0.95$)				
(N, T)	20	30	50	100	200	20	30	50	100	200	20	30	50	100	200	20	30	50	100	200
	CCEMG																			
20	−0.28	−0.26	0.41	−0.31	0.73	14.45	12.85	12.02	12.07	13.47	7.35	5.45	6.40	6.70	6.00	9.35	9.15	10.95	11.55	10.90
30	−0.11	0.07	0.09	0.45	−0.05	11.99	10.78	9.82	9.52	10.33	5.20	5.90	5.95	6.50	6.55	7.85	10.50	12.40	14.35	14.90
50	0.00	0.23	−0.07	−0.02	0.00	9.01	7.97	7.62	6.79	6.72	5.05	4.80	5.00	5.45	4.95	9.40	12.20	15.75	17.60	21.15
100	0.14	−0.08	−0.12	−0.03	0.06	6.66	5.92	5.16	4.78	4.56	4.65	5.40	5.60	4.60	6.35	15.10	18.15	23.95	28.50	34.85
200	0.14	0.11	0.01	−0.17	−0.07	5.13	4.45	3.88	3.27	3.34	5.45	5.10	5.45	4.65	5.15	22.35	28.80	36.60	44.75	56.70
	CCEP																			
20	−0.12	−0.19	0.35	−0.26	0.66	12.66	11.53	11.56	12.12	15.07	7.45	7.00	7.55	6.35	6.50	9.85	10.00	12.60	12.65	11.50
30	−0.09	0.05	0.06	0.39	0.03	10.00	9.57	9.26	9.36	11.05	5.55	5.75	6.80	6.70	6.75	9.90	11.70	13.30	15.20	14.50
50	−0.14	0.39	−0.08	0.01	0.03	7.29	6.92	6.84	6.58	6.79	4.95	5.25	5.45	5.60	4.85	11.25	15.60	16.65	19.95	20.40
100	0.20	−0.13	−0.11	−0.05	0.04	5.44	4.97	4.55	4.45	4.39	4.80	5.35	5.40	4.95	6.05	20.60	22.65	28.35	31.40	36.80
200	0.19	0.11	−0.08	−0.13	−0.07	3.97	3.71	3.35	2.96	3.09	5.25	5.15	5.05	5.00	5.60	31.95	38.45	44.30	50.70	60.40
	CupBC																			
20	0.44	0.33	0.29	0.26	0.20	8.11	6.02	4.11	2.41	1.33	59.65	48.25	34.40	26.15	19.00	69.45	65.90	68.15	85.85	99.30
30	0.18	0.22	0.23	0.14	0.09	6.33	4.64	3.04	1.72	1.00	60.05	48.75	33.85	21.60	20.00	71.45	72.15	79.20	95.10	100.00
50	0.12	0.36	0.03	0.13	0.07	4.90	3.62	2.32	1.29	0.70	60.90	47.10	32.85	20.00	14.75	77.00	82.25	88.35	98.95	100.00
100	0.18	0.02	0.09	−0.01	0.04	3.23	2.48	1.65	0.86	0.48	59.65	48.70	33.20	19.80	16.90	87.85	91.10	97.85	100.00	100.00
200	0.10	0.03	0.06	0.05	0.02	2.39	1.72	1.17	0.63	0.33	59.50	45.65	32.15	21.50	15.80	95.05	98.50	99.65	100.00	100.00

Notes: The DGP is the same as that of Table 1, except that $\gamma_{i2} \sim$ IIDN(0, 1), so the rank condition is not satisfied, and $\beta_{ij} = 1$ for all i and j, $i = 1, 2, \ldots, N$, $j = 1, 2$. See notes to Table 1.

presented in Pesaran (2006), and illustrate the robustness of the CCE estimators to the presence of unit roots in the unobserved common factors. This is important since it obviates the need for pretesting of unobserved common factors for the possibility of non-stationary components.

The CCE estimators perform well, in both heterogeneous and homogeneous slope cases, and irrespective of whether the rank condition is satisfied, although the CCE estimators with rank deficiency have sightly higher RMSEs than those under the full rank condition. The RMSEs of the CCE estimators of Tables 1 and 3 (heterogeneous case) are higher than those reported in Tables 2 and 4 for the homogeneous case. The sizes of the t-test based on the CCE estimators are very close to the nominal 5% level. In the case of full rank, the powers of the tests for the CCE estimators are much higher than in the rank-deficient case. Finally, not surprisingly, the power of the tests for the CCE estimators in the homogeneous case is higher than that in the heterogeneous case.

It is also important to note that the small-sample properties of the CCE estimator do not seem to be much affected by the residual serial correlation of the idiosyncratic errors, ε_{it}. The robustness of the CCE estimator to the short-run dynamics is particularly helpful in practice where typically little is known about such dynamics. In fact, a comparison of the results for the CCEP estimator with the infeasible counterpart given in Table 1 shows that the former can even be more efficient (in the RMSE sense). For example, the RMSE of the CCEP for $N = T = 50$ is 3.97 whilst the RMSE of the infeasible pooled estimator is 4.31. This might seem counter-intuitive at first, but, as indicated above, the infeasible estimator does not take account of the residual serial correlation of the idiosyncratic errors, but the CCE estimator does allow for such possibilities indirectly through the use of the cross-section averages that partly embody the serial correlation properties of $\boldsymbol{f}_t$ and the ε_{it}.

Consider now the PC augmented estimators and recall that they are computed assuming that the true number of common factors is known. The results in Table 1 bear some resemblance to those presented in Kapetanios and Pesaran (2007). The biases and RMSEs of the PC1POOL and PC1MG estimators improve as both N and T increase, but the t-tests based on these estimators

Table 5
Small-sample properties of common correlated effects type estimators, in the case of heterogeneous slopes; the number of factors $m = 4$ exceeds $k + 1 = 3$.

	Bias ($\times 100$)					Root mean square error ($\times 100$)				
(N, T)	20	30	50	100	200	20	30	50	100	200
	CCEMG									
20	0.23	0.29	0.06	−0.23	−0.16	10.97	9.59	8.29	7.61	7.70
30	0.20	0.08	−0.07	0.14	−0.03	8.98	7.65	6.84	6.42	6.29
50	−0.04	0.00	−0.16	−0.19	0.14	6.81	6.03	5.12	4.71	4.67
100	0.12	−0.06	0.01	−0.01	0.12	4.81	4.25	3.69	3.53	3.46
200	0.01	−0.04	0.03	−0.04	−0.10	3.78	3.08	2.84	2.61	2.53
	CCEP									
20	0.09	0.50	−0.02	−0.22	−0.11	9.57	8.94	8.07	7.70	7.83
30	0.03	−0.05	−0.08	0.04	−0.09	7.96	7.21	6.60	6.36	6.25
50	−0.04	−0.05	−0.13	−0.14	0.13	6.06	5.59	4.85	4.54	4.49
100	0.06	−0.07	−0.01	0.01	0.11	4.21	3.85	3.51	3.37	3.38
200	−0.04	−0.05	0.00	−0.03	−0.10	3.13	2.74	2.62	2.42	2.37
	CupBC									
20	0.49	0.32	0.06	0.11	0.11	11.56	10.26	8.94	7.09	6.30
30	0.01	0.12	0.12	0.21	0.07	9.38	7.98	6.68	5.58	4.62
50	−0.11	0.25	−0.08	−0.02	0.21	7.07	6.29	5.03	4.04	3.54
100	0.06	0.04	0.10	0.04	−0.04	4.81	4.32	3.58	2.82	2.54
200	0.05	−0.11	0.00	0.00	−0.03	3.55	3.14	2.61	2.00	1.67

Notes: The DGP is the same as that of Table 1, except that an extra term $\gamma_{i4}f_{4t}$ is added to the y equation, where $\gamma_{i4} \sim \text{IIDN}(0.5, 0.2)$, $f_{4t} = f_{4t-1} + v_{f4,t}$, $v_{f4,t} \sim \text{IIDN}(0, 1)$, $f_{4,-50} = 0$. For the CupBC estimator, the number of unobserved factors is treated as an unknown but is estimated by the information criterion PC_{P2}, which is proposed by Bai and Ng (2002). We set the maximum number of factors to six. See also the notes to Table 1.

substantially over-reject the null hypothesis. The PC2POOL and PC2MG estimators perform even worse. The biases of the PC estimators are always larger in absolute value than the respective biases of the CCE estimators. The size distortion of the PC augmented estimators is particularly pronounced. Finally, it is worth noting that the performance of the PC estimators actually gets worse when N is small and kept small but T rises. This may be related to the fact that the accuracy of the factor estimates depends on the minimum of N and T.

Now consider the CupBC estimator, and again recall that it is computed assuming that the true number of common factors is known. Let us begin with discussing results in the case in which the rank condition is satisfied, the results of which are reported in Tables 1 and 2. As is evident, the average bias and RMSEs of CupBC estimator are comparable to those of CCE estimators. Because of this, the results of CCEMG, CCEP and CupBC estimators only are reported in Table 2 onwards. In the case of heterogeneous slopes with the rank condition satisfied, the RMSEs of the CCE estimator are uniformly smaller than those of the CupBC estimator (as can be seen from Table 1). This might be expected, since the CupBC estimator is designed for the model with homogeneous slopes. In the case of homogeneous slopes with the rank condition satisfied, as can be seen from Table 2, the RMSEs of the CCEP estimator are smaller than those of the CupBC estimator when T is relatively small ($T = 20$ and 30). Turning our attention to the performance of the t-test, it is apparent that the size of the test based on the CupBC estimator is far from the nominal level across all experiments. This is especially so for experiments where the slopes are heterogeneous. In these cases, increases in N and T do not seem to help to improve the test performance. Even for homogeneous slope cases, the best rejection probability result is 14.90% for $T = N = 200$ in Table 2. In contrast, the size of the t-test based on the CCE estimators is close to 5% nominal level across all experiments. Tables 3 and 4 provide the summary of experimental results in the rank-deficient case. For this design, even though the size of the t-test based on the CupBC estimator is grossly oversized, the RMSEs of the estimator are smaller than those of the CCE estimators. However, note that in these experiments the number of factors is treated as known, which is rarely expected in a practical situation. We return to this issue below.

Tables 5–8 report the results of the experiments carried out as robustness checks.[6] Table 5 reports the results of the experiments where the number of unobserved factors is four ($m = 4$), which exceeds $k + 1 = 3$, in the case of heterogeneous slopes. In this experiment, CupBC estimates are obtained supposing that m is unknown but estimated using the information criterion PC_{P2}, which is proposed by Bai and Ng (2002), applied to the first differences of $(y_{it}, x_{1it}, x_{2it})$. We set the maximum number of factors to six.[7] First, despite the number of unobserved factors, $m = 4$, exceeding the number of regressors and regressand ($k + 1 = 3$), the RMSEs of the CCE estimators decrease as N and T are increased, which confirms the consistency of the estimators in the rank-deficient case. Furthermore, the RMSEs of the CCE estimators dominate those of the CupBC estimator, except only when T is very large ($\geq$100). We note that, although not reported for brevity, the size of the t-test based on the CCE estimators is very close to the nominal 5% level, whilst the size distortion of the CupBC estimators is acute for all cases considered. Tables 6–8 report the results of experiments with the same DGP as in Table 1 but where the unobserved factors are cointegrated, factor structures are semi-strong, and the unobserved factors are subject to mean shifts, respectively. In all of these designs the CCE estimators uniformly dominate the CupBC estimator in terms of both RMSEs and the size of the t-test (which is not reported in the tables). These are consistent with the findings of Chudik et al. (forthcoming) and Stock and Watson (2008).

5. Conclusions

Recently, there has been increased interest in the analysis of panel data models where the standard assumption that the errors of the panel regressions are cross-sectionally uncorrelated

[6] For brevity, the size and power of t-tests are not reported in Tables 5–8, since they are qualitatively similar to those in Tables 1–4. For similar reasons, the results for homogeneous slopes and/or rank-deficient cases (for Tables 6–8) are not reported. A full set of results is available upon request from the authors.

[7] For small N and T, the information criterion tends to overestimate the number of the factors in the first-differenced data $(y_{it}, x_{1it}, x_{2it})$, and the estimates tend to 4 as N and T get larger.

Table 6
Small-sample properties of common correlated effects type estimators, heterogeneous slopes and full rank, cointegrated factors, in the case of experiment 1$\mathcal{A}$ (heterogeneous slopes + full rank).

	Bias (×100)					Root mean square error (×100)				
(N, T)	20	30	50	100	200	20	30	50	100	200
	CCEMG									
20	0.05	−0.05	−0.22	0.08	0.00	9.26	7.87	6.58	5.69	5.29
30	−0.14	0.09	0.03	−0.02	0.02	7.35	6.02	5.18	4.54	4.16
50	−0.03	0.14	−0.05	0.11	0.11	5.85	4.70	4.06	3.49	3.14
100	−0.05	−0.01	0.03	−0.05	0.00	4.15	3.40	2.87	2.49	2.19
200	−0.05	0.14	0.03	0.04	−0.04	3.08	2.46	2.02	1.72	1.59
	CCEP									
20	−0.06	−0.01	−0.23	0.06	−0.01	8.52	7.54	6.65	5.95	5.68
30	−0.06	−0.07	−0.07	−0.02	0.01	6.78	5.90	5.25	4.70	4.29
50	−0.03	0.14	−0.09	0.12	0.13	5.35	4.54	4.05	3.55	3.19
100	−0.02	0.03	0.06	−0.03	−0.02	3.77	3.18	2.84	2.50	2.22
200	−0.04	0.10	−0.01	0.05	−0.04	2.70	2.33	1.99	1.72	1.60
	CupBC									
20	0.54	0.85	0.61	0.68	0.78	11.01	9.58	8.01	6.94	6.32
30	0.69	0.52	0.54	0.50	0.68	8.65	7.48	6.26	5.39	4.91
50	0.49	0.54	0.50	0.53	0.58	6.82	5.69	4.99	4.24	3.70
100	0.33	0.37	0.38	0.22	0.26	4.61	3.86	3.43	2.84	2.52
200	0.13	0.31	0.13	0.25	0.09	3.39	2.88	2.41	2.03	1.82

Notes: The DGP of the same as that of Table 1, except that the factors are generated as cointegrated non-stationary processes: $f_{1t} = f_{1t}^t + 0.5f_{2t}^t + v_{f1,t}, f_{2t} = 0.5f_{1t}^t + f_{2t}^t + v_{f2,t}, f_{3t} = 0.75f_{1t}^t + 0.25f_{2t}^t + v_{f3,t}$, with $v_{fj,t} \sim$ IIDN(0, 1), $f_{j,-50} = 0$, for $j = 1, 2, 3$, where $f_{\ell t}^t = f_{\ell t-1}^t + v_{f\ell,t}^t$, with $v_{f\ell,t}^t \sim$ IIDN(0, 1), for $\ell = 1, 2, t = -49, \ldots, 0, \ldots, T$. See also the notes to Table 1.

Table 7
Small-sample properties of common correlated effects type estimators, semi-strong factors, in the case of experiment 1 $\mathcal{A}$ (heterogeneous slopes + full rank).

	Bias (×100)					Root mean square error (×100)				
(N, T)	20	30	50	100	200	20	30	50	100	200
	CCEMG									
20	−0.09	−0.22	−0.07	0.09	−0.09	9.92	8.01	6.57	5.63	5.17
30	0.02	0.01	0.01	−0.11	0.10	7.74	6.21	5.14	4.43	4.10
50	−0.12	0.16	−0.11	0.14	−0.04	5.96	4.57	3.99	3.42	3.10
100	0.01	0.03	0.05	0.02	0.04	4.23	3.51	2.87	2.33	2.26
200	−0.06	0.01	−0.01	0.05	0.00	3.06	2.46	2.00	1.72	1.51
	CCEP									
20	0.09	−0.07	−0.06	0.04	−0.12	8.64	7.49	6.34	5.65	5.34
30	−0.19	−0.10	0.09	−0.08	0.13	7.12	5.90	5.12	4.49	4.21
50	0.01	0.13	−0.05	0.13	−0.02	5.27	4.46	3.93	3.43	3.16
100	0.04	0.08	0.02	0.00	0.03	3.77	3.28	2.84	2.35	2.28
200	−0.07	−0.03	−0.04	0.05	0.00	2.68	2.30	1.96	1.70	1.53
	CupBC									
20	0.23	0.46	0.17	0.43	0.45	12.29	10.55	8.09	6.75	5.80
30	−0.20	0.09	0.38	0.20	0.49	9.53	8.03	6.39	5.14	4.58
50	0.39	0.37	0.06	0.20	0.15	7.34	6.08	5.07	3.99	3.40
100	0.18	0.18	0.06	0.05	0.09	4.99	4.40	3.61	2.69	2.45
200	0.00	0.03	0.03	0.09	0.01	3.77	3.03	2.55	1.98	1.64

Notes: The DGP of the same as that of Table 1, except that the factor loadings matrix $\boldsymbol{\Gamma}_i'$ is multiplied by $N^{-1/2}$ for all i. See also the notes to Table 1.

is violated. When the errors of a panel regression are cross-sectionally correlated, then standard estimation methods do not necessarily produce consistent estimates of the parameters of interest. An influential strand of the relevant literature provides a convenient parameterization of the problem in terms of a factor model for the error terms.

Pesaran (2006) adopts an error multifactor structure and suggests new estimators that take into account cross-sectional dependence, making use of cross-sectional averages of the dependent and explanatory variables. However, he focuses on the case of weakly stationary factors that could be restrictive in some applications. This paper provides a formal extension of the results of Pesaran (2006) to the case where the unobserved factors are allowed to follow unit root processes. It is shown that the main results of Pesaran continue to hold in this more general case. This is certainly of interest, given the fact that usually there are major differences between results obtained for unit root and stationary processes. When we consider the small-sample properties of the new estimators, we observe that again the results accord with the conclusions reached in the stationary case, lending further support to the use of the CCE estimators irrespective of the order of integration of the data observed. The Monte Carlo experiments also show that the CCE type estimators are robust to a number of important departures from the theory developed in this paper, and in general have better small-sample properties than alternatives that are available in the literature. Most importantly, the tests based on CCE estimators have the correct size, whilst the factor-based estimators (including the one recently proposed by Bai et al. (2009)) show substantial size distortions even in the case of relatively large samples.

Appendix A. Lemmas

Proofs of the lemmas are provided in Appendix B.

Table 8
Small-sample properties of common correlated effects type estimators, one break in the means of unobserved factors, in the case of experiment 1A (heterogeneous slopes + full rank).

(N, T)	Bias ($\times 100$)					Root mean square error ($\times 100$)				
	20	30	50	100	200	20	30	50	100	200
	CCEMG									
20	0.01	−0.10	−0.02	0.06	−0.07	9.66	7.82	6.74	5.87	5.54
30	0.14	−0.03	−0.02	−0.13	0.10	7.68	6.08	5.11	4.54	4.22
50	−0.21	0.20	−0.11	0.14	−0.04	5.91	4.64	4.01	3.43	3.13
100	0.02	0.03	0.05	0.03	0.04	4.26	3.48	2.88	2.33	2.26
200	−0.08	−0.02	−0.02	0.06	0.00	3.08	2.49	2.01	1.72	1.51
	CCEP									
20	0.17	0.00	−0.05	0.00	−0.13	8.73	7.61	6.86	6.30	6.21
30	−0.15	−0.13	0.07	−0.14	0.14	7.10	5.98	5.31	4.78	4.46
50	−0.03	0.18	−0.06	0.11	−0.01	5.30	4.53	3.97	3.47	3.21
100	0.05	0.09	0.04	0.01	0.02	3.80	3.26	2.85	2.34	2.28
200	−0.06	−0.04	−0.05	0.05	0.00	2.72	2.29	1.95	1.71	1.53
	CupBC									
20	0.52	0.77	0.79	0.80	0.89	11.18	9.87	8.39	7.52	6.97
30	0.32	0.58	0.77	0.58	0.84	8.91	7.80	6.55	5.68	5.27
50	0.58	0.75	0.38	0.61	0.54	6.78	6.01	5.03	4.18	3.82
100	0.28	0.35	0.38	0.29	0.32	4.85	4.22	3.41	2.75	2.55
200	0.10	0.08	0.08	0.23	0.17	3.57	2.93	2.44	2.01	1.69

Notes: The DGP is the same as that of Table 1, except that $f_{jt} = \varphi_{jt}$ for $t < \lfloor 2T/3 \rfloor$ and $f_{jt} = 1 + \varphi_{jt}$ for $t \geq \lfloor 2T/3 \rfloor$, with $\lfloor A \rfloor$ being the greatest integer part of A, where $\varphi_{jt} = \varphi_{j,t-1} + \zeta_{jt}$, $\zeta_{jt} \sim \text{IIDN}(0, 1)$, $j = 1, 2, 3$. See also the notes to Table 1.

Lemma 1. *Under Assumptions 1–4,*

$$\frac{\bar{\boldsymbol{U}}'\bar{\boldsymbol{U}}}{T} = O_p\left(\frac{1}{N}\right) \tag{A.1}$$

$$\frac{\boldsymbol{V}_i'\bar{\boldsymbol{U}}}{T} = O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right), \quad \frac{\boldsymbol{\varepsilon}_i'\bar{\boldsymbol{U}}}{T} = O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right), \quad \text{uniformly over } i \tag{A.2}$$

$$\frac{\boldsymbol{F}'\bar{\boldsymbol{U}}}{T} = O_p\left(\frac{1}{\sqrt{N}}\right), \qquad \frac{\boldsymbol{D}'\bar{\boldsymbol{U}}}{T} = O_p\left(\frac{1}{\sqrt{N}}\right) \tag{A.3}$$

$$\frac{\boldsymbol{X}_i'\bar{\boldsymbol{U}}}{T} = O_p\left(\frac{1}{\sqrt{N}}\right), \quad \text{uniformly over } i \tag{A.4}$$

$$\frac{\boldsymbol{Q}'\bar{\boldsymbol{U}}}{T} = O_p\left(\frac{1}{\sqrt{N}}\right) \tag{A.5}$$

$$\frac{\boldsymbol{Q}'\boldsymbol{Q}}{T^2} = O_p(1) \tag{A.6}$$

$$\frac{\boldsymbol{Q}'\boldsymbol{X}_i}{T^2} = O_p(1), \quad \text{uniformly over } i \tag{A.7}$$

$$\frac{\boldsymbol{Q}'\boldsymbol{G}}{T^2} = O_p(1) \tag{A.8}$$

$$\frac{\bar{\boldsymbol{H}}'\bar{\boldsymbol{H}}}{T^2} = O_p(1) \tag{A.9}$$

$$\frac{\bar{\boldsymbol{H}}'\boldsymbol{G}}{T^2} = O_p(1) \tag{A.10}$$

$$\frac{\bar{\boldsymbol{H}}'\boldsymbol{\varepsilon}_i}{T} = O_p(1), \quad \text{uniformly over } i \tag{A.11}$$

$$\frac{\bar{\boldsymbol{H}}'\boldsymbol{V}_i}{T} = O_p(1), \quad \text{uniformly over } i \tag{A.12}$$

$$\frac{\bar{\boldsymbol{H}}'\boldsymbol{X}_i}{T^2} = O_p(1), \quad \text{uniformly over } i \tag{A.13}$$

$$\frac{\bar{\boldsymbol{H}}'\bar{\boldsymbol{U}}}{T} = O_p\left(\frac{1}{\sqrt{N}}\right). \tag{A.14}$$

Lemma 2. *Under Assumptions 1–4,*

$$\frac{\boldsymbol{V}_i'\bar{\boldsymbol{U}}}{T} = O_p\left(\frac{1}{\sqrt{TN}}\right) + O_p\left(\frac{1}{N}\right) \quad \text{uniformly over } i. \tag{A.15}$$

Lemma 3. *Under Assumptions 1–4, and assuming that rank condition (9) holds,*

$$\frac{\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{X}_i}{T} - \frac{\boldsymbol{X}_i'\boldsymbol{M}_g\boldsymbol{X}_i}{T} = O_p\left(\frac{1}{\sqrt{N}}\right), \quad \text{uniformly over } i, \tag{A.16}$$

$$\frac{\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{\varepsilon}_i}{T} - \frac{\boldsymbol{X}_i'\boldsymbol{M}_g\boldsymbol{\varepsilon}_i}{T} = O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{N}\right), \quad \text{uniformly over } i. \tag{A.17}$$

Lemma 4. *Assume that the rank condition (9) holds. Then, under Assumptions 1–4,*

$$\frac{\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{F}}{T} = O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right), \quad \text{uniformly over } i. \tag{A.18}$$

Lemma 5. *Under Assumptions 1–4,*

$$\frac{\boldsymbol{X}_i'\boldsymbol{M}_g\boldsymbol{X}_i}{T} - \boldsymbol{\Sigma}_{\boldsymbol{v}_i} = O_p\left(\frac{1}{\sqrt{T}}\right).$$

Lemma 6. *Under Assumptions 1–4, and assuming that the rank condition (9) does not hold,*

$$\frac{\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{X}_i}{T} - \frac{\boldsymbol{X}_i'\boldsymbol{M}_q\boldsymbol{X}_i}{T} = O_p\left(\frac{1}{\sqrt{N}}\right), \quad \text{uniformly over } i, \tag{A.19}$$

$$\frac{\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{F}}{T} - \frac{\boldsymbol{X}_i'\boldsymbol{M}_q\boldsymbol{F}}{T} = O_p\left(\frac{1}{\sqrt{N}}\right), \quad \text{uniformly over } i, \tag{A.20}$$

$$\frac{\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{\varepsilon}_i}{T} - \frac{\boldsymbol{X}_i'\boldsymbol{M}_q\boldsymbol{\varepsilon}_i}{T} = O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{N}\right), \quad \text{uniformly over } i. \tag{A.21}$$

Lemma 7. *Under Assumptions 1–4, and assuming that the rank condition (9) does not hold,*

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

uniformly over i. (A.22)

Appendix B. Proofs of lemmas

Proof of Lemma 1. To prove (A.1), we first show that

$$E\left\|\bar{\mathbf{u}}_t\right\|^2 = O\left(\frac{1}{N}\right), \quad \text{and} \quad E\left\|\bar{\mathbf{u}}_t\right\| = O\left(\frac{1}{\sqrt{N}}\right). \tag{B.23}$$

We recall that

$$\bar{\mathbf{u}}_t = \left(\bar{\varepsilon}_t + \underbrace{\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\beta}_i'\mathbf{v}_{it}}_{\bar{\mathbf{v}}_t}\right), \tag{B.24}$$

where $\bar{\mathbf{v}}_t = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\beta}_i'\mathbf{v}_{it}$. Then, by the cross-sectional independence of $\mathbf{v}_{it}$ and $\boldsymbol{\beta}_i'$ specified in Assumptions 2 and 4, $E\left\|\bar{\mathbf{v}}_t\right\|^2 = \frac{1}{N^2}\sum_{i=1}^{N}E\left\|\boldsymbol{\beta}_i'\mathbf{v}_{it}\right\|^2$, and again by Assumptions 2 and 4, we have $E\left\|\bar{\mathbf{v}}_t\right\|^2 \le \frac{K}{N} = O\left(\frac{1}{N}\right)$. Similarly, $E\left(\bar{\varepsilon}_t^2\right) = O\left(\frac{1}{N}\right)$. Next, note that $T^{-1}\bar{\mathbf{U}}'\bar{\mathbf{U}} = T^{-1}\left(\sum_{t=1}^{T}\bar{\mathbf{u}}_t\bar{\mathbf{u}}_t'\right)$, where the cross-product terms in $\bar{\mathbf{u}}_t\bar{\mathbf{u}}_t'$, being functions of covariance stationary processes with finite fourth-order cumulants, are themselves stationary with finite means and variances. Also, $E\left\|T^{-1}\bar{\mathbf{U}}'\bar{\mathbf{U}}\right\| \le T^{-1}\sum_{t=1}^{T}E\left\|\bar{\mathbf{u}}_t\right\|^2$, and by (B.23) $E\left\|T^{-1}\bar{\mathbf{U}}'\bar{\mathbf{U}}\right\| = O\left(N^{-1}\right)$, which establishes (A.1).

The result for $\mathbf{V}_i'\bar{\mathbf{U}}/T$ in (A.2) is established in Lemma 2 below. The result for $\boldsymbol{\varepsilon}_i'\bar{\mathbf{U}}/T$ in (A.2) is established similarly to that for $\mathbf{V}_i'\bar{\mathbf{U}}/T$.

To establish (A.3), first we examine $T^{-1}\left(\mathbf{F}'\bar{\mathbf{U}}\right)$. Consider the ℓth row of $T^{-1}\left(\mathbf{F}'\bar{\mathbf{U}}\right)$ and note that it can be written as $T^{-1}\left(\sum_{t=1}^{T}f_{\ell t}\bar{\mathbf{u}}_t'\right)$. Since by assumption $f_{\ell t}$ and $\bar{\mathbf{u}}_t$ are independently distributed processes, it easily follows that

$$Var\left(\frac{\sum_{t=1}^{T}f_{\ell t}\bar{\mathbf{u}}_t}{T}\right) = O\left(\frac{1}{N}\right)\left\{\frac{\sum_{t=1}^{T}\sum_{t'=1}^{T}E\left(f_{\ell t}f_{\ell t'}\right)}{T^2}\right\}.$$

But, by standard unit root asymptotic analysis, we know that $\sum_{t=1}^{T}\sum_{t'=1}^{T}E\left(f_{\ell t}f_{\ell t'}\right) = O\left(T^2\right)$, and therefore

$$Var\left(\frac{\sum_{t=1}^{T}f_{\ell t}\bar{\mathbf{u}}_t}{T}\right) = O\left(\frac{1}{N}\right), \tag{B.25}$$

which establishes that $T^{-1}\sum_{t=1}^{T}f_{\ell t}\bar{\mathbf{u}}_{wt}$ converges to its limit at the desired rate of $O_p\left(1/\sqrt{N}\right)$. The result for $T^{-1}\left(\mathbf{D}'\bar{\mathbf{U}}\right)$ is obtained using the same line of arguments.

To establish (A.4), first note that

$$\frac{\mathbf{X}_i'\bar{\mathbf{U}}}{T} = \boldsymbol{\Pi}_i'\frac{(\mathbf{D},\mathbf{F})'\bar{\mathbf{U}}}{T} + \frac{\mathbf{V}_i'\bar{\mathbf{U}}}{T} = O_p\left(\frac{1}{\sqrt{N}}\right), \quad \text{uniformly over } i$$

using (A.2) and (A.3), since the elements of $\boldsymbol{\Pi}_i$ are assumed to be bounded uniformly over i.

To establish (A.5), recalling that $\mathbf{Q} = \mathbf{G}\bar{\mathbf{P}}$, and using (A.3), $\frac{\mathbf{Q}'\bar{\mathbf{U}}}{T} = \bar{\mathbf{P}}'\frac{(\mathbf{D},\mathbf{F})'\bar{\mathbf{U}}}{T} = O_p\left(\frac{1}{\sqrt{N}}\right)$, since the elements of $\bar{\mathbf{P}}$ are assumed to be bounded. (A.6) is established by $\frac{\mathbf{Q}'\mathbf{Q}}{T^2} = \bar{\mathbf{P}}'\frac{\mathbf{G}'\mathbf{G}}{T^2}\bar{\mathbf{P}} = O_p(1)$, since $\mathbf{G}'\mathbf{G}/T^2 = O_p(1)$.

To establish (A.7), first note that

$$\frac{\mathbf{Q}'\mathbf{X}_i}{T^2} = \bar{\mathbf{P}}'\left(\frac{\mathbf{G}'\mathbf{G}}{T^2}\right)\boldsymbol{\Pi}_i + \bar{\mathbf{P}}'\frac{\mathbf{G}'\mathbf{V}_i}{T^2}. \tag{B.26}$$

The first term is $O_p(1)$ uniformly over i, since the elements of $\bar{\mathbf{P}}$ and $\boldsymbol{\Pi}_i$ are assumed to be bounded in probability uniformly over i. For the second term, under Assumptions 1–2, denoting $g_{\ell t}$ as the ℓth element of $\mathbf{g}_t$ and noting that $\sup_i E\left(\mathbf{v}_{it}\mathbf{v}_{it'}'\right) = O(1)$, we have that

$$\sup_i Var\left(\frac{\sum_{t=1}^{T}g_{\ell t}\mathbf{v}_{it}'}{T}\right) = O(1)\left\{\frac{\sum_{t=1}^{T}\sum_{t'=1}^{T}E\left(g_{\ell t}g_{\ell t'}\right)}{T^2}\right\}.$$

But, by standard unit root asymptotic analysis, we know that $\sum_{t=1}^{T}\sum_{t'=1}^{T}E\left(g_{\ell t}g_{\ell t'}\right) = O\left(T^2\right)$, and therefore $\sup_i Var\left(\frac{\sum_{t=1}^{T}g_{\ell t}\mathbf{v}_{it}'}{T}\right) = O(1)$. Hence, $\mathbf{G}'\mathbf{V}_i/T = O_p(1)$ uniformly over i for sufficiently large T. Therefore, as the elements of $\bar{\mathbf{P}}$ are assumed to be bounded in probability, the second term is $O_p(1)$ uniformly over i, which establishes (A.7). (A.8) is straightforwardly proven, using (A.6).

To prove (A.9), recalling $\bar{\mathbf{H}} = \mathbf{Q} + \bar{\mathbf{U}}^*$, where $\bar{\mathbf{U}}^* = \left(\mathbf{0}, \bar{\mathbf{U}}\right)$,

$$\frac{\bar{\mathbf{H}}'\bar{\mathbf{H}}}{T^2} = \frac{\mathbf{Q}'\mathbf{Q}}{T^2} + \frac{\bar{\mathbf{U}}^{*\prime}\bar{\mathbf{U}}^*}{T^2} + \frac{\mathbf{Q}'\bar{\mathbf{U}}^*}{T^2} + \frac{\bar{\mathbf{U}}^{*\prime}\mathbf{Q}}{T^2} = O_p(1) \tag{B.27}$$

by (A.1), (A.5) and (A.6). To establish (A.10), $\frac{\bar{\mathbf{H}}'\mathbf{F}}{T^2} = \bar{\mathbf{P}}'\frac{\mathbf{G}'\mathbf{F}}{T^2} + \frac{\bar{\mathbf{U}}^{*\prime}\mathbf{F}}{T^2} = O_p(1)$, since $\mathbf{G}'\mathbf{F}/T^2$ is $O_p(1)$. (A.11) is established because

$$\frac{\bar{\mathbf{H}}'\boldsymbol{\varepsilon}_i}{T} = \bar{\mathbf{P}}'\frac{\mathbf{G}'\boldsymbol{\varepsilon}_i}{T} + \frac{\bar{\mathbf{U}}^{*\prime}\boldsymbol{\varepsilon}_i}{T} = O_p(1) \quad \text{uniformly over } i, \tag{B.28}$$

since $\mathbf{G}'\boldsymbol{\varepsilon}_i/T = O_p(1)$ uniformly over i, using the same line of the argument as in the proof of (A.7). (A.12) can be proven similarly to (A.11).

Next,

$$\frac{\bar{\mathbf{H}}'\mathbf{X}_i}{T^2} = \frac{\mathbf{Q}'\mathbf{X}_i}{T^2} + \frac{\bar{\mathbf{U}}^{*\prime}\mathbf{X}_i}{T^2} = O_p(1) \quad \text{uniformly over } i$$

by (A.4) and (A.7), which establishes (A.13). Finally, (A.14) follows by the boundedness in probability of $\bar{\mathbf{P}}$ and (A.3). □

Proof of Lemma 2. In order to prove (A.15), we need to examine more closely Lemma A.2. of Pesaran (2006). So, we have

$$\frac{\mathbf{V}_i'\bar{\mathbf{U}}}{T} = \left(T^{-1}\mathbf{V}_i'\bar{\boldsymbol{\varepsilon}} + (NT)^{-1}\mathbf{V}_i'\sum_{j=1}^{N}\mathbf{V}_j\boldsymbol{\beta}_j, T^{-1}\mathbf{V}_i'\bar{\mathbf{V}}\right), \tag{B.29}$$

where $\bar{\boldsymbol{\varepsilon}} = N^{-1}\sum_{j=1}^{N}\boldsymbol{\varepsilon}_j$ and $\bar{\mathbf{V}} = N^{-1}\sum_{j=1}^{N}\mathbf{V}_j$. Denote the tth element of $\bar{\boldsymbol{\varepsilon}}$ by $\bar{\varepsilon}_t = N^{-1}\sum_{j=1}^{N}\varepsilon_{jt}$, and consider the first term on the right-hand side (RHS) of (B.29). Since, by assumption, $\mathbf{v}_{it}$ and $\bar{\varepsilon}_t$ are independently distributed covariance stationary processes with zero means, it follows that

$$\sup_i Var\left(\frac{\sum_{t=1}^{T}v_{i\ell t}\bar{\boldsymbol{\varepsilon}}_t}{T}\right) = O\left(\frac{1}{N}\right)\sup_i\left\{\frac{\sum_{t=1}^{T}\sum_{t'=1}^{T}\Gamma_{iv\ell}\left(\left|t-t'\right|\right)}{T^2}\right\},$$

where $\Gamma_{iv\ell}\left(\left|t-t'\right|\right)$ is the autocovariance function of the stationary process, $v_{i\ell t}$. But, by Assumption 2, $\sup_i\sum_{s=1}^{\infty}\Gamma_{iv\ell}(|s|) < \infty$.

Therefore,

$$\sup_i Var\left(\frac{\sum_{t=1}^{T} v_{i\ell t}\bar{\boldsymbol{\varepsilon}}_t}{T}\right) = O\left(\frac{1}{NT}\right), \tag{B.30}$$

which establishes that

$$T^{-1}\boldsymbol{V}_i'\bar{\boldsymbol{\varepsilon}} = O_p\left(\frac{1}{\sqrt{TN}}\right), \quad \text{uniformly over } i. \tag{B.31}$$

To see how (B.31) follows from (B.30), we note that, by the Markov inequality,

$$\Pr\left(T^{-1}\boldsymbol{V}_i'\bar{\boldsymbol{\varepsilon}} \geq \epsilon\right) \leq \frac{Var\left(\frac{\sum_{t=1}^{T} v_{i\ell t}\bar{\boldsymbol{\varepsilon}}_t}{T}\right)}{\epsilon^2}, \quad \text{for all } i.$$

But, since for any two continuous functions f, g, if $f(x) \leq g(x)$ for all x, $\sup_x f(x) \leq \sup_x g(x)$, it follows that

$$\sup_i \Pr\left(T^{-1}\boldsymbol{V}_i'\bar{\boldsymbol{\varepsilon}} \geq \epsilon\right) \leq \frac{\sup_i Var\left(\frac{\sum_{t=1}^{T} v_{i\ell t}\bar{\boldsymbol{\varepsilon}}_t}{T}\right)}{\epsilon^2},$$

proving that (B.31) follows from (B.30).

Consider the second term in (B.29), and note that

$$(NT)^{-1}\boldsymbol{V}_i'\sum_{j=1}^{N}\boldsymbol{V}_j\boldsymbol{\beta}_j = N^{-1}\left(\frac{\boldsymbol{V}_i'\boldsymbol{V}_i}{T}\right)\boldsymbol{\beta}_i + \left(\frac{\boldsymbol{V}_i'\bar{\boldsymbol{V}}_{-i}^*}{T}\right), \tag{B.32}$$

where $\bar{\boldsymbol{V}}_{-i}^* = N^{-1}\sum_{j=1,j\neq i}^{N}\boldsymbol{V}_j\boldsymbol{\beta}_j$. Since $\boldsymbol{\beta}_i$ is bounded, and, by Assumption 2, $\text{plim}_{T\to\infty}\left(T^{-1}\boldsymbol{V}_i'\boldsymbol{V}_i\right) = \boldsymbol{\Sigma}_{vi}$ uniformly over i, it follows that

$$N^{-1}\left(\frac{\boldsymbol{V}_i'\boldsymbol{V}_i}{T}\right)\boldsymbol{\beta}_i = O_p\left(\frac{1}{N}\right), \quad \text{uniformly over } i. \tag{B.33}$$

Also, since the elements of $\boldsymbol{V}_i$ and $\bar{\boldsymbol{V}}_{-i}^*$ are independently distributed and covariance stationary, following the same line of analysis leading to (B.31), we have

$$\frac{\boldsymbol{V}_i'\bar{\boldsymbol{V}}_{-i}^*}{T} = O_p\left(\frac{1}{\sqrt{NT}}\right), \quad \text{uniformly over } i. \tag{B.34}$$

Using (B.33) and (B.34) in (B.32) now yields

$$(NT)^{-1}\boldsymbol{V}_i'\sum_{j=1}^{N}\boldsymbol{V}_j\boldsymbol{\beta}_j = O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{N}\right),$$

$$\text{uniformly over } i. \tag{B.35}$$

Finally, since the last term of (B.29) can be written as $T^{-1}\boldsymbol{V}_i'\bar{\boldsymbol{V}} = N^{-1}\left(\frac{\boldsymbol{V}_i'\boldsymbol{V}_i}{T}\right) + \frac{\boldsymbol{V}_i'\bar{\boldsymbol{V}}_{-i}}{T}$, where $\bar{\boldsymbol{V}}_{-i} = N^{-1}\sum_{j=1,j\neq i}^{N}\boldsymbol{V}_j$, it also follows that

$$T^{-1}\boldsymbol{V}_i'\bar{\boldsymbol{V}} = O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{N}\right), \quad \text{uniformly over } i, \tag{B.36}$$

which completes the proof of Lemma 2. □

Proof of Lemma 3. We start by proving (A.16). We need to determine the order of probability of $\left\|\frac{\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{X}_i}{T} - \frac{\boldsymbol{X}_i'\boldsymbol{M}_q\boldsymbol{X}_i}{T}\right\|$. But this is equal to

$$\left\|\frac{\boldsymbol{X}_i'\bar{\boldsymbol{H}}\left(\bar{\boldsymbol{H}}'\bar{\boldsymbol{H}}\right)^{-1}\bar{\boldsymbol{H}}'\boldsymbol{X}_i}{T} - \frac{\boldsymbol{X}_i'\boldsymbol{Q}\left(\boldsymbol{Q}'\boldsymbol{Q}\right)^{-1}\boldsymbol{Q}'\boldsymbol{X}_i}{T}\right\|$$
$$\leq \left\|\frac{1}{T}\left(\boldsymbol{X}_i'\bar{\boldsymbol{H}} - \boldsymbol{X}_i'\boldsymbol{Q}\right)\left(\bar{\boldsymbol{H}}'\bar{\boldsymbol{H}}\right)^{-1}\bar{\boldsymbol{H}}'\boldsymbol{X}_i\right\|$$
$$+ \left\|\frac{1}{T}\boldsymbol{X}_i'\boldsymbol{Q}\left(\left(\bar{\boldsymbol{H}}'\bar{\boldsymbol{H}}\right)^{-1} - \left(\boldsymbol{Q}'\boldsymbol{Q}\right)^{-1}\right)\bar{\boldsymbol{H}}'\boldsymbol{X}_i\right\|$$
$$+ \left\|\frac{1}{T}\boldsymbol{X}_i'\boldsymbol{Q}\left(\boldsymbol{Q}'\boldsymbol{Q}\right)^{-1}\left(\bar{\boldsymbol{H}}'\boldsymbol{X}_i - \boldsymbol{Q}'\boldsymbol{X}_i\right)\right\|. \tag{B.37}$$

We examine each of the above terms. So, noting that $\bar{\boldsymbol{H}} = \boldsymbol{Q} + \bar{\boldsymbol{U}}^*$, with $\bar{\boldsymbol{U}}^* = \left(\boldsymbol{0}, \bar{\boldsymbol{U}}\right)$, we have

$$\left\|\frac{1}{T}\left(\boldsymbol{X}_i'\bar{\boldsymbol{H}} - \boldsymbol{X}_i'\boldsymbol{Q}\right)\left(\bar{\boldsymbol{H}}'\bar{\boldsymbol{H}}\right)^{-1}\bar{\boldsymbol{H}}'\boldsymbol{X}_i\right\| \leq \left\|\frac{\boldsymbol{X}_i'\bar{\boldsymbol{U}}^*}{T}\right\|\left\|\left(\frac{\bar{\boldsymbol{H}}'\bar{\boldsymbol{H}}}{T^2}\right)^{-1}\frac{\bar{\boldsymbol{H}}'\boldsymbol{X}_i}{T^2}\right\|$$
$$= O_p\left(\frac{1}{\sqrt{N}}\right), \quad \text{uniformly over } i, \tag{B.38}$$

by (A.4), (A.9) and (A.13). Next, we have

$$\left\|\frac{1}{T}\boldsymbol{X}_i'\boldsymbol{Q}\left(\left(\bar{\boldsymbol{H}}'\bar{\boldsymbol{H}}\right)^{-1} - \left(\boldsymbol{Q}'\boldsymbol{Q}\right)^{-1}\right)\bar{\boldsymbol{H}}'\boldsymbol{X}_i\right\|$$
$$\leq \left\|\frac{\bar{\boldsymbol{U}}^{*\prime}\bar{\boldsymbol{U}}^*}{T} + \frac{\boldsymbol{Q}'\bar{\boldsymbol{U}}^*}{T} + \frac{\bar{\boldsymbol{U}}^{*\prime}\boldsymbol{Q}}{T}\right\|\left\|\frac{\boldsymbol{X}_i'\boldsymbol{Q}}{T^2}\left(\frac{\bar{\boldsymbol{H}}'\bar{\boldsymbol{H}}}{T^2}\right)^{-1}\right\|$$
$$\times\left\|\left(\frac{\boldsymbol{Q}'\boldsymbol{Q}}{T^2}\right)^{-1}\frac{\bar{\boldsymbol{H}}'\boldsymbol{X}_i}{T^2}\right\|$$
$$= O_p\left(\frac{1}{\sqrt{N}}\right), \quad \text{uniformly over } i, \tag{B.39}$$

by (A.1), (A.5), (A.7), (A.9), (A.6) and (A.13). Finally,

$$\left\|\frac{1}{T}\boldsymbol{X}_i'\boldsymbol{Q}\left(\boldsymbol{Q}'\boldsymbol{Q}\right)^{-1}\left(\bar{\boldsymbol{H}}'\boldsymbol{X}_i - \boldsymbol{Q}'\boldsymbol{X}_i\right)\right\| \leq \left\|\frac{\boldsymbol{X}_i'\boldsymbol{Q}}{T^2}\left(\frac{\boldsymbol{Q}'\boldsymbol{Q}}{T^2}\right)^{-1}\right\|\left\|\frac{\boldsymbol{X}_i'\bar{\boldsymbol{U}}^*}{T}\right\|$$
$$= O_p\left(\frac{1}{\sqrt{N}}\right) \quad \text{uniformly over } i, \tag{B.40}$$

by (A.7), (A.6) and (A.4). Noting that $\boldsymbol{M}_g = \boldsymbol{M}_q$ when the rank condition is satisfied, substituting (B.38)–(B.40) into (B.37), we have

$$\left\|\frac{\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{X}_i}{T} - \frac{\boldsymbol{X}_i'\boldsymbol{M}_g\boldsymbol{X}_i}{T}\right\| = O_p\left(\frac{1}{\sqrt{N}}\right), \quad \text{uniformly over } i,$$

as required.

Next, we consider (A.17). In particular, by a similar analysis to that for (A.16), we have

$$\left\|\frac{\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{\varepsilon}_i}{T} - \frac{\boldsymbol{X}_i'\boldsymbol{M}_q\boldsymbol{\varepsilon}_i}{T}\right\| \leq \left\|\frac{1}{T}\left(\boldsymbol{X}_i'\bar{\boldsymbol{H}} - \boldsymbol{X}_i'\boldsymbol{Q}\right)\left(\bar{\boldsymbol{H}}'\bar{\boldsymbol{H}}\right)^{-1}\bar{\boldsymbol{H}}'\boldsymbol{\varepsilon}_i\right\|$$
$$+ \left\|\frac{1}{T}\boldsymbol{X}_i'\boldsymbol{Q}\left(\left(\bar{\boldsymbol{H}}'\bar{\boldsymbol{H}}\right)^{-1} - \left(\boldsymbol{Q}'\boldsymbol{Q}\right)^{-1}\right)\bar{\boldsymbol{H}}'\boldsymbol{\varepsilon}_i\right\|$$
$$+ \left\|\frac{1}{T}\boldsymbol{X}_i'\boldsymbol{Q}\left(\boldsymbol{Q}'\boldsymbol{Q}\right)^{-1}\left(\bar{\boldsymbol{H}}'\boldsymbol{\varepsilon}_i - \boldsymbol{Q}'\boldsymbol{\varepsilon}_i\right)\right\|. \tag{B.41}$$

We examine each of the above terms. So, we have

$$\left\|\frac{1}{T}\left(\boldsymbol{X}_i'\bar{\boldsymbol{H}} - \boldsymbol{X}_i'\boldsymbol{Q}\right)\left(\bar{\boldsymbol{H}}'\bar{\boldsymbol{H}}\right)^{-1}\bar{\boldsymbol{H}}'\boldsymbol{\varepsilon}_i\right\|$$

$$\leq \frac{1}{T}\left\|\frac{X_i'\bar{U}^*}{T}\right\|\left\|\left(\frac{\bar{H}'\bar{H}}{T^2}\right)^{-1}\frac{\bar{H}'\varepsilon_i}{T}\right\|$$

$$= O_p\left(\frac{1}{\sqrt{NT}}\right), \quad \text{uniformly over } i, \tag{B.42}$$

by (A.4), (A.9) and (A.11). Next, we have

$$\left\|\frac{1}{T}X_i'Q\left(\left(\bar{H}'\bar{H}\right)^{-1}-\left(Q'Q\right)^{-1}\right)\bar{H}'\varepsilon_i\right\|$$

$$\leq \frac{1}{T}\left\|-\frac{\bar{U}^{*\prime}\bar{U}^*}{T}-\frac{\bar{U}^{*\prime}Q}{T}-\frac{Q'\bar{U}^*}{T}\right\|\left\|\frac{X_i'Q}{T^2}\left(\frac{\bar{H}'\bar{H}}{T^2}\right)^{-1}\right\|$$

$$\times\left\|\left(\frac{Q'Q}{T^2}\right)^{-1}\frac{\bar{H}'\varepsilon_i}{T}\right\|$$

$$= O_p\left(\frac{1}{\sqrt{NT}}\right), \quad \text{uniformly over } i, \tag{B.43}$$

by (A.1), (A.5), (A.7), (A.9), (A.6) and (A.11). Finally,

$$\left\|\frac{1}{T}X_i'Q\left(Q'Q\right)^{-1}\left(\bar{H}'\varepsilon_i-Q'\varepsilon_i\right)\right\| \leq \left\|\frac{X_i'Q}{T^2}\left(\frac{Q'Q}{T^2}\right)^{-1}\right\|\left\|\frac{\bar{U}^{*\prime}\varepsilon_i}{T}\right\|$$

$$= O_p\left(\frac{1}{\sqrt{NT}}\right)+O_p\left(\frac{1}{N}\right), \quad \text{uniformly over } i, \tag{B.44}$$

by (A.7), (A.6) and (A.2). Noting that $M_g = M_q$ when the rank condition is satisfied, substituting (B.42)–(B.44) into (B.41) yields

$$\left\|\frac{X_i'\bar{M}\varepsilon_i}{T}-\frac{X_i'M_g\varepsilon_i}{T}\right\| = O_p\left(\frac{1}{\sqrt{NT}}\right)+O_p\left(\frac{1}{N}\right) \quad \text{uniformly over } i,$$

which establishes (A.17). □

Proof of Lemma 4. We start by noting that

$$\bar{M}\bar{H} = \bar{M}\left(G\bar{P}+\bar{U}^*\right).$$

But $\bar{M}\bar{H} = \mathbf{0}$ and $\bar{M}D = \mathbf{0}$, since $\bar{H} = \left(D, \bar{Z}\right)$. Then

$$\mathbf{0} = \left(\mathbf{0}, \bar{M}F\right)\begin{pmatrix} I_n & \bar{B} \\ \mathbf{0} & \bar{C}\end{pmatrix} + \left(\mathbf{0}, \bar{M}\bar{U}\right),$$

or $\bar{M}F\bar{C} = -\bar{M}\bar{U}$. Hence,

$$\left(\bar{U}'\bar{M}F\right)\bar{C} = -\bar{U}'\bar{M}\bar{U}. \tag{B.45}$$

Also, from above,

$$\left(X_i'\bar{M}F\right)C = \bar{-}X_i'\bar{M}\bar{U}. \tag{B.46}$$

Note, however, that $X_i = G\Pi_i + V_i$, and hence

$$\begin{aligned} X_i'\bar{M}\bar{U} &= \left(\Pi_i'G'+V_i'\right)\bar{M}\bar{U} \\ &= \Pi_i'\left(G'\bar{M}\bar{U}\right)+V_i'\bar{M}\bar{U} \\ &= \left(A_i', \Gamma_i'\right)\begin{pmatrix} D' \\ F'\end{pmatrix}\bar{M}\bar{U}+V_i'\bar{\bar{M}}\bar{U} \\ &= \left(A_i', \Gamma_i'\right)\begin{pmatrix}\mathbf{0} \\ F'\bar{M}\bar{U}\end{pmatrix}+V_i'\bar{M}\bar{U} \\ &= \Gamma_i'F'\bar{M}\bar{U}+V_i'\bar{M}\bar{U}. \end{aligned} \tag{B.47}$$

By the full rank assumption for $\bar{C}$ and substituting (B.47) in (B.46), we obtain

$$\left(X_i'\bar{M}F\right) = -\Gamma_i'F'\bar{M}\bar{U}\bar{C}'\left(\bar{C}\bar{C}'\right)^{-1}-V_i'\bar{M}\bar{U}\bar{C}'\left(\bar{C}\bar{C}'\right)^{-1}. \tag{B.48}$$

Also, from (B.45),

$$\left(F'\bar{M}\bar{U}\right) = -\left(\bar{C}\bar{C}'\right)^{-1}\bar{C}\bar{U}'\bar{M}\bar{U}. \tag{B.49}$$

Then, using this result in (B.48), we have

$$\left\|\frac{X_i'\bar{M}F}{T}\right\| \leq \left\|\Gamma_i'\right\|\left\|\left(\bar{C}\bar{C}'\right)^{-1}\bar{C}\right\|^2\left\|\frac{\bar{U}'\bar{M}\bar{U}}{T}\right\|$$

$$+\left\|\frac{V_i'\bar{M}\bar{U}}{T}\right\|\left\|\bar{C}'\left(\bar{C}\bar{C}'\right)^{-1}\right\|. \tag{B.50}$$

Since the norms of $\left(\bar{C}\bar{C}'\right)^{-1}\bar{C}$ and Γ_i' are bounded, we need to establish the probability orders of $\left\|\bar{U}'\bar{M}\bar{U}/T\right\|$ and $\left\|V_i'\bar{M}\bar{U}/T\right\|$. For $\bar{U}'\bar{M}\bar{U}/T$, using (A.1), (A.9) and (A.14), we have

$$\frac{\bar{U}'\bar{M}\bar{U}}{T} = O_p\left(\frac{1}{N}\right). \tag{B.51}$$

Similarly, by (A.2) and (A.12),

$$\frac{V_i'\bar{M}\bar{U}}{T} = O_p\left(\frac{1}{N}\right)+O_p\left(\frac{1}{\sqrt{NT}}\right), \quad \text{uniformly over } i, \tag{B.52}$$

and substituting (B.51) and (B.52) into (B.50) establishes the result. □

Proof of Lemma 5. Recall that

$$X_i = G\Pi_i + V_i, \tag{B.53}$$

where $G = (D, F)$ is the $T \times m + n$ matrix of $I(1)$ factors, and V_i is a stationary error matrix. Denote the OLS residuals of the multiple regression (B.53) as $\hat{V}_i = X_i - G\hat{\Pi}_i$, where $\hat{\Pi}_i = \left(G'G\right)^{-1}G'X_i$. Observe that $\hat{V}_i = M_gX_i$. Then, we can write

$$\begin{aligned} \hat{V}_i'\hat{V}_i/T - V_i'V_i/T &= \hat{V}_i'\left(\hat{V}_i-V_i\right)/T+\left(\hat{V}_i-V_i\right)'V_i/T \\ &= -X_i'M_gG\left(\hat{\Pi}_i-\Pi_i\right)/T-\left(\hat{\Pi}_i-\Pi_i\right)'\left(G'V_i/T\right) \\ &= -\left(\hat{\Pi}_i-\Pi_i\right)'\left(G'V_i/T\right), \end{aligned}$$

because $M_gG = \mathbf{0}$. But, since $\left(G'V_i/T\right) = O_p(1)$ and $\left(\hat{\Pi}_i-\Pi_i\right) = O_p(T^{-1})$, it follows that $\hat{V}_i'\hat{V}_i/T - V_i'V_i/T = O_p(T^{-1})$. The required result now follows since, under Assumption 2, $V_i'V_i/T - \Sigma_{vi} = O_p(T^{-1/2})$, where Σ_{vi} is a non-singular matrix. □

Proof of Lemma 6. The procedure in Lemma 3 can be used to prove (A.19) and (A.21), but replacing all inverses with generalized inverses. This is required since $Q'Q$ has reduced rank when the rank condition does not hold. We need to show that

$$\left\|\frac{1}{T}X_i'Q\left[\left(\bar{H}'\bar{H}\right)^{+}-\left(Q'Q\right)^{+}\right]\bar{H}'X_i\right\| = O_p\left(\frac{1}{\sqrt{N}}\right)$$

$$\text{uniformly over } i, \tag{B.54}$$

where + denotes the Moore–Penrose inverse. To establish (B.54), we need to show that

$$\left(\frac{Q'Q}{T^2}\right)^{+}-\left(\frac{\bar{H}'\bar{H}}{T^2}\right)^{+} = O_p\left(\frac{1}{T\sqrt{N}}\right). \tag{B.55}$$

However, because the Moore–Penrose inverse is not a continuous function, it is not sufficient that

$$\left(\frac{Q'Q}{T^2}\right)-\left(\frac{\bar{H}'\bar{H}}{T^2}\right) = O_p\left(\frac{1}{T\sqrt{N}}\right), \tag{B.56}$$

for (B.55) to hold. But, by Theorem 2 of Andrews (1987), (B.56) is sufficient for (B.55), if additionally, as $(N,T)\xrightarrow{j}\infty$,

$$\lim_{N,T\xrightarrow{j}\infty}\Pr\left(rk\left(\frac{\bar{\boldsymbol{H}}'\bar{\boldsymbol{H}}}{T^2}\right)=rk\left(\frac{\boldsymbol{Q}'\boldsymbol{Q}}{T^2}\right)\right)=1, \tag{B.57}$$

where $rk(\boldsymbol{A})$ denotes the rank of $\boldsymbol{A}$. But,

$$\frac{\bar{\boldsymbol{H}}'\bar{\boldsymbol{H}}}{T^2}=\frac{\boldsymbol{Q}'\boldsymbol{Q}}{T^2}+\frac{\bar{\boldsymbol{U}}^{*\prime}\bar{\boldsymbol{U}}^*}{T^2}+\frac{\boldsymbol{Q}'\bar{\boldsymbol{U}}^*}{T^2}+\frac{\bar{\boldsymbol{U}}^{*\prime}\boldsymbol{Q}}{T^2},$$

with

$$\lim_{N,T\xrightarrow{j}\infty}\Pr\left(\left\|\frac{\bar{\boldsymbol{U}}^{*\prime}\bar{\boldsymbol{U}}^*}{T^2}+\frac{\boldsymbol{Q}'\bar{\boldsymbol{U}}^*}{T^2}+\frac{\bar{\boldsymbol{U}}^{*\prime}\boldsymbol{Q}}{T^2}\right\|>\epsilon\right)=0$$

for all $\epsilon>0$. Also, $rk(T^{-2}\boldsymbol{Q}'\boldsymbol{Q})=n+rk(\bar{\boldsymbol{C}})$, for all N and T, with $rk(T^{-2}\boldsymbol{Q}'\boldsymbol{Q})\to n+rk(\boldsymbol{C})<n+m$ as $(N,T)\xrightarrow{j}\infty$. Using these results, it is now easily seen that condition (B.57) in fact holds. Hence, the desired result follows.

Consider now (A.20). Following a similar line of analysis used to establish (A.19), we have

$$\begin{aligned}\left\|\frac{\boldsymbol{X}_i'\bar{\boldsymbol{M}}F}{T}-\frac{\boldsymbol{X}_i'\boldsymbol{M}_q\boldsymbol{F}}{T}\right\| &\le \left\|\frac{1}{T}\left(\boldsymbol{X}_i'\bar{\boldsymbol{H}}-\boldsymbol{X}_i'\boldsymbol{Q}\right)\left(\bar{\boldsymbol{H}}'\bar{\boldsymbol{H}}\right)^+\bar{\boldsymbol{H}}'\boldsymbol{F}\right\|\\ &+\left\|\frac{1}{T}\boldsymbol{X}_i'\boldsymbol{Q}\left(\left(\bar{\boldsymbol{H}}'\bar{\boldsymbol{H}}\right)^+-\left(\boldsymbol{Q}'\boldsymbol{Q}\right)^+\right)\bar{\boldsymbol{H}}'\boldsymbol{F}\right\|\\ &+\left\|\frac{1}{T}\boldsymbol{X}_i'\boldsymbol{Q}\left(\boldsymbol{Q}'\boldsymbol{Q}\right)^+\left(\bar{\boldsymbol{H}}'\boldsymbol{F}-\boldsymbol{Q}'\boldsymbol{F}\right)\right\|.\end{aligned} \tag{B.58}$$

Consider each of the above terms in turn. First,

$$\left\|\frac{1}{T}\left(\boldsymbol{X}_i'\bar{\boldsymbol{H}}-\boldsymbol{X}_i'\boldsymbol{Q}\right)\left(\bar{\boldsymbol{H}}'\bar{\boldsymbol{H}}\right)^+\bar{\boldsymbol{H}}'\boldsymbol{F}\right\|=O_p\left(\frac{1}{\sqrt{N}}\right),\quad\text{uniformly over } i, \tag{B.59}$$

by (A.4), (A.9) and (A.10). Second, by (B.55) and (B.56),

$$\left\|\frac{1}{T}\boldsymbol{X}_i'\boldsymbol{Q}\left[\left(\bar{\boldsymbol{H}}'\bar{\boldsymbol{H}}\right)^+-\left(\boldsymbol{Q}'\boldsymbol{Q}\right)^+\right]\bar{\boldsymbol{H}}'\boldsymbol{F}\right\|=O_p\left(\frac{1}{\sqrt{N}}\right),\quad\text{uniformly over } i,$$

if $\left\|\frac{\boldsymbol{Q}'\boldsymbol{Q}}{T}-\frac{\bar{\boldsymbol{H}}'\bar{\boldsymbol{H}}}{T}\right\|=O_p\left(\frac{1}{\sqrt{N}}\right)$. We have

$$\left\|\frac{\boldsymbol{Q}'\boldsymbol{Q}}{T}-\frac{\bar{\boldsymbol{H}}'\bar{\boldsymbol{H}}}{T}\right\|=O_p\left(\frac{1}{\sqrt{N}}\right)\quad\text{uniformly over } i \tag{B.60}$$

by (A.1), (A.5), (A.7), (A.9), (A.6) and (A.10). Finally,

$$\left\|\frac{1}{T}\boldsymbol{X}_i'\boldsymbol{Q}\left(\boldsymbol{Q}'\boldsymbol{Q}\right)^+\left(\bar{\boldsymbol{H}}'\boldsymbol{F}-\boldsymbol{Q}'\boldsymbol{F}\right)\right\|=O_p\left(\frac{1}{\sqrt{N}}\right),\quad\text{uniformly over } i, \tag{B.61}$$

by (A.7), (A.6) and (A.3). Substituting (B.59)–(B.61) into (B.58) yields the required result. □

Proof of Lemma 7. The result immediately follows from (B.48), (B.49), (B.51) and (B.52). □

Appendix C. Proofs of theorems for pooled estimators

Proof of Theorem 1. We know that

$$\bar{\boldsymbol{C}}=\left(\bar{\boldsymbol{\gamma}}+\bar{\boldsymbol{\Gamma}}\boldsymbol{\beta}+\frac{1}{N}\sum_{i=1}^N\boldsymbol{\Gamma}_i\boldsymbol{\varkappa}_i,\bar{\boldsymbol{\Gamma}}\right),$$

where $\bar{\boldsymbol{\Gamma}}=\frac{1}{N}\sum_{i=1}^N\boldsymbol{\Gamma}_i$ and $\bar{\boldsymbol{\gamma}}=\frac{1}{N}\sum_{i=1}^N\boldsymbol{\gamma}_i$. Substituting this result in (A.22) now yields

$$\left(\frac{\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{F}}{T}\right)\left(\bar{\boldsymbol{\gamma}}+\bar{\boldsymbol{\Gamma}}\boldsymbol{\beta}+\frac{1}{N}\sum_{i=1}^N\boldsymbol{\Gamma}_i\boldsymbol{\varkappa}_i\right)=O_p\left(\frac{1}{N}\right)+O_p\left(\frac{1}{\sqrt{NT}}\right),\quad\text{uniformly over } i,$$

$$\left(\frac{\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{F}}{T}\right)\bar{\boldsymbol{\Gamma}}=O_p\left(\frac{1}{N}\right)+O_p\left(\frac{1}{\sqrt{NT}}\right),\quad\text{uniformly over } i,$$

which in turn yields

$$\frac{\sqrt{N}\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{F}}{T}\left(\bar{\boldsymbol{\gamma}}+\frac{1}{N}\sum_{i=1}^N\boldsymbol{\Gamma}_i\boldsymbol{\varkappa}_i\right)=O_p\left(\frac{1}{\sqrt{N}}\right)+O_p\left(\frac{1}{\sqrt{T}}\right),\quad\text{uniformly over } i.$$

But, under Assumption 4, $\frac{1}{N}\sum_{i=1}^N\boldsymbol{\Gamma}_i\boldsymbol{\varkappa}_i=O_p(N^{-1/2})$, and therefore

$$\frac{\sqrt{N}(\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{F})\bar{\boldsymbol{\gamma}}}{T}=O_p\left(\frac{1}{\sqrt{N}}\right)+O_p\left(\frac{1}{\sqrt{T}}\right),\quad\text{uniformly over } i. \tag{C.62}$$

We next reconsider the second term on the RHS of (35), which is the only term affected by the fact that the rank condition does not hold. The second term on the RHS in (35) can be written as

$$\chi_{NT}\equiv\frac{1}{N}\sum_{i=1}^N\left(\frac{\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{X}_i}{T^2}\right)^+\left(\frac{\sqrt{N}\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{F}}{T^2}\right)\left(\bar{\boldsymbol{\gamma}}+\boldsymbol{\eta}_i-\bar{\boldsymbol{\eta}}\right), \tag{C.63}$$

where $\bar{\boldsymbol{\eta}}=\frac{1}{N}\sum_{i=1}^N\boldsymbol{\eta}_i$. By (A.19) and (A.20), it follows that

$$\chi_{NT}\equiv\frac{1}{N}\sum_{i=1}^N\left(\frac{\boldsymbol{X}_i'\boldsymbol{M}_q\boldsymbol{X}_i}{T^2}\right)^+\left(\frac{\sqrt{N}\boldsymbol{X}_i'\boldsymbol{M}_q\boldsymbol{F}}{T^2}\right)\times\left(\bar{\boldsymbol{\gamma}}+\boldsymbol{\eta}_i-\bar{\boldsymbol{\eta}}\right)+O_p\left(\frac{1}{\sqrt{N}}\right). \tag{C.64}$$

Note that, for the above two expressions, we have changed the normalization from T to T^2. This is because, in the case where the rank condition does not hold, the use of cross-sectional averages is not sufficient to remove the effect of the $I(1)$ unobserved factors, and so $\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{X}_i$, $\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{F}$, $\boldsymbol{X}_i'\boldsymbol{M}_q\boldsymbol{X}_i$ and $\boldsymbol{X}_i'\boldsymbol{M}_q\boldsymbol{F}$ would involve non-stationary components. Then, since, by (C.62), $\frac{\sqrt{N}(\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{F})\bar{\boldsymbol{\gamma}}}{T^2}=O_p\left(\frac{1}{T\sqrt{N}}\right)+O_p\left(\frac{1}{T^{3/2}}\right)$, uniformly over i, it is the case that, for N and T large,

$$\sqrt{N}(\hat{\boldsymbol{b}}_{MG}-\boldsymbol{\beta})\overset{d}{\sim}\frac{1}{\sqrt{N}}\sum_{i=1}^N\boldsymbol{\varkappa}_i+\frac{1}{\sqrt{N}}\sum_{i=1}^N\left(\frac{\boldsymbol{X}_i'\boldsymbol{M}_q\boldsymbol{X}_i}{T^2}\right)^+\times\left(\frac{\boldsymbol{X}_i'\boldsymbol{M}_q\boldsymbol{F}}{T^2}\right)\left(\boldsymbol{\eta}_i-\bar{\boldsymbol{\eta}}\right). \tag{C.65}$$

We next focus on analysing the RHS of (C.65). The first term on the RHS of (C.65) tends to a Normal density with mean zero and finite variance. The second term needs further analysis. Letting

$$\boldsymbol{Q}_{1iT}=\left(\frac{\boldsymbol{X}_i'\boldsymbol{M}_q\boldsymbol{X}_i}{T^2}\right)^+\left(\frac{\boldsymbol{X}_i'\boldsymbol{M}_q\boldsymbol{F}}{T^2}\right)$$

and $\bar{\boldsymbol{Q}}_{1T}=\frac{1}{N}\sum_{i=1}^N\boldsymbol{Q}_{1iT}$, we have that

$$\frac{1}{\sqrt{N}}\sum_{i=1}^N\boldsymbol{Q}_{1iT}\left(\boldsymbol{\eta}_i-\bar{\boldsymbol{\eta}}\right)=\frac{1}{\sqrt{N}}\sum_{i=1}^N\left(\boldsymbol{Q}_{1iT}-\bar{\boldsymbol{Q}}_{1T}\right)\boldsymbol{\eta}_i. \tag{C.66}$$

We note that $\boldsymbol{\eta}_i$ is i.i.d. with zero mean and finite variance and independent of all other stochastic quantities in the second term of the RHS on (C.66). We define

$$\mathbf{Q}_{1iT,-i} = \left(\frac{\mathbf{X}_i'\mathbf{M}_{q,-i}\mathbf{X}_i}{T^2}\right)^+ \left(\frac{\mathbf{X}_i'\mathbf{M}_{q,-i}\mathbf{F}}{T^2}\right)$$

and $\bar{\mathbf{Q}}_{1T,-i} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{Q}_{1iT,-i}$, where $\mathbf{M}_{q,-i} = \mathbf{I}_T - \mathbf{Q}_{-i}(\mathbf{Q}'_{-i}\mathbf{Q}_{-i})^+\mathbf{Q}'_{-i}$, $\mathbf{Q}_{-i} = \mathbf{G}\bar{\mathbf{P}}_{-i}$, $\bar{\mathbf{P}}_{-i} = \begin{pmatrix}\mathbf{I}_n & \bar{\mathbf{B}}_{-i}\\ \mathbf{0} & \bar{\mathbf{C}}_{-i}\end{pmatrix}$ $\bar{\mathbf{B}}_{-i} = \frac{1}{N}\sum_{j=1,j\neq i}^{N}\mathbf{B}_j$ and $\bar{\mathbf{C}}_{-i} = \frac{1}{N}\sum_{j=1,j\neq i}^{N}\mathbf{C}_j$. Then, it is straightforward that

$$\left(\mathbf{Q}_{1iT} - \bar{\mathbf{Q}}_{1T}\right) - \left(\mathbf{Q}_{1iT,-i} - \bar{\mathbf{Q}}_{1T,-i}\right) = O_p\left(\frac{1}{N}\right), \quad \text{uniformly over } i,$$

and

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\mathbf{Q}_{1iT} - \bar{\mathbf{Q}}_{1T}\right)\boldsymbol{\eta}_i - \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\mathbf{Q}_{1iT,-i} - \bar{\mathbf{Q}}_{1T,-i}\right)\boldsymbol{\eta}_i = O_p\left(\frac{1}{N^{1/2}}\right).$$

Then, it is easy to show that, if $z_{Ti} = x_i y_{Ti}$, x_i is an i.i.d. sequence with zero mean and finite variance and y_{Ti} is a triangular array of random variables with finite variance, then z_{Ti} is a martingale difference triangular array for which a central limit theorem holds (see, e.g., Theorem 24.3 of Davidson (1994)). But this is the case here, for any ordering over i, setting $y_{Ti} = \left(\mathbf{Q}_{1iT,-i} - \bar{\mathbf{Q}}_{1T,-i}\right)$ and $x_i = \boldsymbol{\eta}_i$. Using this result, it follows that the second term on the RHS of (C.65) tends to a Normal density if $\left(\mathbf{Q}_{1iT} - \bar{\mathbf{Q}}_{1T}\right)\boldsymbol{\eta}_i$ has variance with finite norm, uniformly over i, denoted by $\boldsymbol{\Sigma}_{iqT}$; i.e.,

$$\boldsymbol{\Sigma}_{iqT} = \mathrm{Var}[(\mathbf{Q}_{1iT} - \bar{\mathbf{Q}}_{1T})\boldsymbol{\eta}_i]. \tag{C.67}$$

In order to establish the existence of second moments, it is sufficient to prove that $\|(\mathbf{Q}_{1iT} - \bar{\mathbf{Q}}_{1T})\|$, or equivalently $\left\|\left(\mathbf{Q}_{1iT,-i} - \bar{\mathbf{Q}}_{1T,-i}\right)\right\|$, has finite second moments. We carry out the analysis for $\left\|\left(\mathbf{Q}_{1iT} - \bar{\mathbf{Q}}_{1T}\right)\right\|$. For this, we need to provide further analysis of $\frac{\mathbf{X}_i'\mathbf{M}_q\mathbf{X}_i}{T^2}$ and $\frac{\mathbf{X}_i'\mathbf{M}_q\mathbf{F}}{T^2}$. First, note that $\mathbf{X}_i$ can be written as

$$\mathbf{X}_i = \mathbf{Q}\mathbf{B}_{i1} + \mathbf{S}\mathbf{B}_{i2} + \mathbf{V}_i, \tag{C.68}$$

where $\mathbf{S}$ is the $T \times m - k - 1$-dimensional complement of $\mathbf{Q}$, i.e., $\mathbf{Q}$ and $\mathbf{S}$ are orthogonal, and

$$\mathbf{F} = \mathbf{Q}\mathbf{K}_1 + \mathbf{S}\mathbf{K}_2, \tag{C.69}$$

where $\mathbf{K}_1$ and $\mathbf{K}_2$ are full row rank matrices of constants with bounded norm. Note that, if $m < 2k+1$, we assume, without loss of generality, that $\mathbf{B}_{i2}$ has full row rank, whereas, if $m \geq 2k+1$, $\mathbf{B}_{i2}$ has full column rank. Then,

$$\begin{aligned}\mathbf{X}_i'\mathbf{M}_q\mathbf{X}_i &= \mathbf{X}_i'\mathbf{M}_q\left(\mathbf{Q}\mathbf{B}_{i1} + \mathbf{S}\mathbf{B}_{i2} + \mathbf{V}_i\right) = \mathbf{X}_i'\mathbf{M}_q\mathbf{S}\mathbf{B}_{i2} + \mathbf{X}_i'\mathbf{M}_q\mathbf{V}_i\\ &= \mathbf{B}_{i2}'\mathbf{S}'\mathbf{M}_q\mathbf{S}\mathbf{B}_{i2} + \mathbf{V}_i'\mathbf{M}_q\mathbf{V}_i + \mathbf{B}_{i2}'\mathbf{S}'\mathbf{M}_q\mathbf{V}_i + \mathbf{V}_i'\mathbf{M}_q\mathbf{S}\mathbf{B}_{i2}.\end{aligned}$$

But it easily follows that

$$\frac{\mathbf{V}_i'\mathbf{M}_q\mathbf{V}_i}{T^2} = O_p\left(\frac{1}{T}\right), \quad \text{uniformly over } i,$$

and

$$\frac{\mathbf{B}_{i2}'\mathbf{S}'\mathbf{M}_q\mathbf{V}_i}{T^2} = O_p\left(\frac{1}{T}\right), \quad \text{uniformly over } i.$$

Then,

$$\frac{\mathbf{X}_i'\mathbf{M}_q\mathbf{X}_i}{T^2} = \mathbf{B}_{i2}'\frac{\mathbf{S}'\mathbf{S}}{T^2}\mathbf{B}_{i2} + O_p\left(\frac{1}{T}\right), \quad \text{uniformly over } i. \tag{C.70}$$

Similarly, using (C.69),

$$\frac{\mathbf{X}_i'\mathbf{M}_q\mathbf{F}}{T^2} = \mathbf{B}_{i2}'\frac{\mathbf{S}'\mathbf{S}}{T^2}\mathbf{K}_2 + O_p\left(\frac{1}{T}\right), \quad \text{uniformly over } i.$$

Thus,

$$\left(\frac{\mathbf{X}_i'\mathbf{M}_q\mathbf{X}_i}{T^2}\right)^+\left(\frac{\mathbf{X}_i'\mathbf{M}_q\mathbf{F}}{T^2}\right) = \left(\mathbf{B}_{i2}'\frac{\mathbf{S}'\mathbf{S}}{T^2}\mathbf{B}_{i2}\right)^+\left(\mathbf{B}_{i2}'\frac{\mathbf{S}'\mathbf{S}}{T^2}\mathbf{K}_2\right) + O_p\left(\frac{1}{T}\right), \quad \text{uniformly over } i.$$

We need to distinguish between two cases. In the first case, $m \geq 2k+1$. Then, it is easy to see that $\frac{\mathbf{X}_i'\mathbf{M}_q\mathbf{X}_i}{T^2}$ and $\mathbf{B}_{i2}'\frac{\mathbf{S}'\mathbf{S}}{T^2}\mathbf{B}_{i2}$ have an inverse. Then, by Assumption 7(ii), $\left\|\left(\mathbf{Q}_{1iT} - \bar{\mathbf{Q}}_{1T}\right)\right\|$ has finite second moments. The case where $m < 2k+1$ is more complicated. Denoting $\Delta = T^{-2}\mathbf{S}'\mathbf{S}$ and $\tilde{\mathbf{B}}_{i2} = \Delta^{1/2}\mathbf{B}_{i2}$, we have

$$\mathbf{B}_{i2}'\frac{\mathbf{S}'\mathbf{S}}{T^2}\mathbf{B}_{i2} = \tilde{\mathbf{B}}_{i2}'\tilde{\mathbf{B}}_{i2}.$$

Then, noting that $\left(\tilde{\mathbf{B}}_{i2}'\tilde{\mathbf{B}}_{i2}\right)^+ = \tilde{\mathbf{B}}_{i2}^+\tilde{\mathbf{B}}_{i2}'^+$, and since in this case $\mathbf{B}_{i2}$ has full row rank,

$$\tilde{\mathbf{B}}_{i2}^+ = \mathbf{B}_{i2}'\left(\mathbf{B}_{i2}\mathbf{B}_{i2}'\right)^{-1}\Delta^{-1/2},$$

and we obtain

$$\left(\mathbf{B}_{i2}'\frac{\mathbf{S}'\mathbf{S}}{T^2}\mathbf{B}_{i2}\right)^+ = \mathbf{B}_{i2}'\left(\mathbf{B}_{i2}\mathbf{B}_{i2}'\right)^{-1}\left(\frac{\mathbf{S}'\mathbf{S}}{T^2}\right)^{-1}\left(\mathbf{B}_{i2}\mathbf{B}_{i2}'\right)^{-1}\mathbf{B}_{i2}. \tag{C.71}$$

Hence,

$$\left(\frac{\mathbf{X}_i'\mathbf{M}_q\mathbf{X}_i}{T^2}\right)^+\left(\frac{\mathbf{X}_i'\mathbf{M}_q\mathbf{F}}{T^2}\right) = \mathbf{B}_{i2}'\left(\mathbf{B}_{i2}\mathbf{B}_{i2}'\right)^{-1}\mathbf{K}_2 + O_p\left(\frac{1}{T}\right), \quad \text{uniformly over } i,$$

and the required result now follows by the boundedness assumption for $\mathbf{B}_{i2}$ and $\mathbf{K}_2$. The assumption that $\mathbf{B}_{i2}$ has full row rank if $m < 2k+1$ implies that the whole of $\mathbf{S}$ enters the equations for $\mathbf{X}_i$. If that is not the case, then the argument above has to be modified as follows. We have that

$$\mathbf{X}_i = \mathbf{Q}\mathbf{B}_{i1} + \mathbf{S}_1\mathbf{B}_{i2} + \mathbf{V}_i,$$

where $\mathbf{S}_1$ is a subset of $\mathbf{S}$. Then,

$$\frac{\mathbf{X}_i'\mathbf{M}_q\mathbf{X}_i}{T^2} = \mathbf{B}_{i2}'\frac{\mathbf{S}_1'\mathbf{S}_1}{T^2}\mathbf{B}_{i2} + O_p\left(\frac{1}{T}\right), \quad \text{uniformly over } i,$$

and the analysis proceeds as above until

$$\left(\frac{\mathbf{X}_i'\mathbf{M}_q\mathbf{X}_i}{T^2}\right)^+\left(\frac{\mathbf{X}_i'\mathbf{M}_q\mathbf{F}}{T^2}\right) = \mathbf{B}_{i2}'\left(\mathbf{B}_{i2}\mathbf{B}_{i2}'\right)^{-1}\left(\frac{\mathbf{S}_1'\mathbf{S}_1}{T^2}\right)^{-1} \times \left(\frac{\mathbf{S}_1'\mathbf{S}}{T^2}\right)\mathbf{K}_2 + O_p\left(\frac{1}{T}\right), \quad \text{uniformly over } i.$$

Then, the required result follows by the boundedness assumption for $\mathbf{B}_{i2}$ and $\mathbf{K}_2$ and by Assumption 7(iii), which implies that $E\left\|\left(\frac{\mathbf{S}_1'\mathbf{S}_1}{T^2}\right)^{-1}\right\| < \infty$ and $E\left\|\frac{\mathbf{S}_1'\mathbf{S}}{T^2}\right\| < \infty$.

Thus, in general, we have that

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

where

$$\boldsymbol{\Sigma}_{MG} = \boldsymbol{\Omega}_{\varkappa} + \boldsymbol{\Lambda}, \tag{C.72}$$

and

$$\boldsymbol{\Lambda} = \lim_{N,T\to\infty} \left[\frac{1}{N} \sum_{i=1}^{N} \boldsymbol{\Sigma}_{iqT} \right]. \tag{C.73}$$

To complete the proof, we have to consider two further issues. First, we note that, in (C.65), we disregard a term involving $\left(\frac{X_i' M_q X_i}{T^2}\right)^+ \left(\frac{X_i' M_q \varepsilon_i}{T}\right)$. In particular, we have to prove that

$$\left(\frac{1}{T}\right) \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left(\frac{X_i' M_q X_i}{T^2}\right)^+ \left(\frac{X_i' M_q \varepsilon_i}{T}\right) = O_p\left(\frac{1}{T}\right). \tag{C.74}$$

For this, it is enough to show that $\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left(\frac{X_i' M_q X_i}{T^2}\right)^+ \left(\frac{X_i' M_q \varepsilon_i}{T}\right)$ follows a central limit theorem. This holds under the following conditions: (i) for any ordering of the cross-sectional units, $\frac{X_i' M_q \varepsilon_i}{T}$ is a martingale difference; (ii) $\left(\frac{X_i' M_q X_i}{T^2}\right)^+ \left(\frac{X_i' M_q \varepsilon_i}{T}\right)$ has finite second moments. (ii) follows easily from the above argument about the existence of moments of $\left(\frac{X_i' M_q X_i}{T^2}\right)^+ \left(\frac{X_i' M_q F}{T}\right)$. Then, one has to simply prove (i). We need to show that, for any ordering,

$$E(Q_i^* | Q_{i-1}^*) = 0, \tag{C.75}$$

where $Q_i^* = \left(\frac{X_i' M_q X_i}{T^2}\right)^+ \left(\frac{X_i' M_q \varepsilon_i}{T}\right)$. Denote $Q_i^{**} = \left(\frac{X_i' M_q X_i}{T^2}\right)^+$. Then $Q_i^* = Q_i^{**} \left(\frac{X_i' M_q \varepsilon_i}{T}\right)$. Now $\frac{X_i' M_q \varepsilon_i}{T} = \frac{1}{T} \sum_{t=1}^{T} s_t \varepsilon_{it}$, where s_t is a unit root process (see the definition of $\boldsymbol{S}$ in (C.68) above). Then, for (C.75) to hold it is sufficient to note that, for all t, l, $E(Q_i^{**} s_t \varepsilon_{it} | Q_i^{**} s_l \varepsilon_{i-1l}) = 0$. This completes the proof of (C.74).

Finally, we need to show that the variance estimator given by

$$\hat{\boldsymbol{\Sigma}}_{MG} = \frac{1}{N-1} \sum_{i=1}^{N} (\hat{\boldsymbol{b}}_i - \hat{\boldsymbol{b}}_{MG})(\hat{\boldsymbol{b}}_i - \hat{\boldsymbol{b}}_{MG})', \tag{C.76}$$

is consistent. To see this, first note that

$$\hat{\boldsymbol{b}}_i - \boldsymbol{\beta} = \boldsymbol{\varkappa}_i + \boldsymbol{h}_{iT} + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right),$$
$$\text{uniformly over } i, \tag{C.77}$$

where

$$\boldsymbol{h}_{iT} = \left(\frac{X_i' \bar{M} X_i}{T^2}\right)^+ \frac{X_i' \bar{M} \left[F\left(\boldsymbol{\eta}_i - \bar{\boldsymbol{\eta}}\right) + \boldsymbol{\varepsilon}_i\right]}{T^2}, \tag{C.78}$$

and so

$$\hat{\boldsymbol{b}}_i - \hat{\boldsymbol{b}}_{MG} = (\boldsymbol{\varkappa}_i - \bar{\boldsymbol{\varkappa}}) + \left(\boldsymbol{h}_{iT} - \bar{\boldsymbol{h}}_T\right) + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right),$$
$$\text{uniformly over } i, \tag{C.79}$$

where $\bar{\boldsymbol{h}}_T = \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{h}_{iT}$. Since, by assumption, $\boldsymbol{\varkappa}_i$ and $\boldsymbol{h}_{iT}$ are independently distributed across i,

$$\frac{1}{N-1} \sum_{i=1}^{N} \left(\hat{\boldsymbol{b}}_i - \hat{\boldsymbol{b}}_{MG}\right) \left(\hat{\boldsymbol{b}}_i - \hat{\boldsymbol{b}}_{MG}\right)' = \boldsymbol{\Sigma}_{MG} + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right),$$

and the desired result follows. □

Proof of Theorem 2. As before, the pooled estimator, $\hat{\boldsymbol{b}}_P$, defined by (20), can be written as

$$\sqrt{N}(\hat{\boldsymbol{b}}_P - \boldsymbol{\beta}) = \left(\frac{1}{N} \sum_{i=1}^{N} \frac{X_i' \bar{M} X_i}{T^2}\right)^{-1} \times \left[\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{X_i' \bar{M} (X_i \boldsymbol{\varkappa}_i + \boldsymbol{\varepsilon}_i)}{T^2} + \boldsymbol{q}_{NT}\right], \tag{C.80}$$

where

$$\boldsymbol{q}_{NT} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{\left(X_i' \bar{M} F\right) \boldsymbol{\gamma}_i}{T^2}. \tag{C.81}$$

Assuming random coefficients, we note that $\boldsymbol{\gamma}_i = \bar{\boldsymbol{\gamma}} + \boldsymbol{\eta}_i - \bar{\boldsymbol{\eta}}$, where $\bar{\boldsymbol{\eta}} = \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{\eta}_i$. Hence,

$$\boldsymbol{q}_{NT} = \frac{1}{N} \sum_{i=1}^{N} \left(\frac{\sqrt{N} X_i' \bar{M} F}{T^2}\right) \bar{\boldsymbol{\gamma}} + \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left(\frac{X_i' \bar{M} F}{T^2}\right) (\boldsymbol{\eta}_i - \bar{\boldsymbol{\eta}}).$$

But, by (C.62), the first component of $\boldsymbol{q}_{NT}$ is $O_p\left(\frac{1}{T\sqrt{N}}\right) + O_p\left(\frac{1}{T^{3/2}}\right)$. Substituting this result in (C.80), and making use of (33) and (34), we have

$$\sqrt{N}(\hat{\boldsymbol{b}}_P - \boldsymbol{\beta}) = \left(\frac{1}{N} \sum_{i=1}^{N} \frac{X_i' M_q X_i}{T^2}\right)^{-1} \times \left[\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{X_i' M_q \left(X_i \boldsymbol{\varkappa}_i + \boldsymbol{\varepsilon}_i + F\left(\boldsymbol{\eta}_i - \bar{\boldsymbol{\eta}}\right)\right)}{T^2}\right] + O_p\left(\frac{1}{T\sqrt{N}}\right) + O_p\left(\frac{1}{T^{3/2}}\right). \tag{C.82}$$

Also, by Assumption 7, when the rank condition is not satisfied, $\frac{1}{N} \sum_{i=1}^{N} \frac{X_i' M_q X_i}{T^2}$ is non-singular. Further, by (C.70),

$$\frac{1}{N} \sum_{i=1}^{N} \frac{X_i' M_q X_i}{T^2} = \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{B}_{i2}' \frac{\boldsymbol{S}'\boldsymbol{S}}{T^2} \boldsymbol{B}_{i2} + O_p\left(\frac{1}{T}\right).$$

We note that, by Assumption 3, $\boldsymbol{B}_{i2}$ is an i.i.d. sequence with finite second moments. Further, by Assumption 7, it follows that $E\left\|\frac{\boldsymbol{S}'\boldsymbol{S}}{T^2}\right\|^2 < \infty$. Hence, $T^{-2}\boldsymbol{B}_{i2}'\boldsymbol{S}'\boldsymbol{S}\boldsymbol{B}_{i2}$ forms asymptotically a martingale difference triangular array with finite mean and variance and, as a result, $T^{-2}\boldsymbol{B}_{i2}'\boldsymbol{S}'\boldsymbol{S}\boldsymbol{B}_{i2}$ obeys the martingale difference triangular array law of large numbers across i (see, e.g., Theorem 19.7 of Davidson (1994)) and, therefore, its mean tends to a non-stochastic limit which we denote by $\boldsymbol{\Theta}$; i.e.,

$$\boldsymbol{\Theta} = \lim_{N,T\to\infty} \left(\frac{1}{N} \sum_{i=1}^{N} \boldsymbol{\Theta}_{iT}\right), \tag{C.83}$$

where

$$\boldsymbol{\Theta}_{iT} = E\left(T^{-2}\boldsymbol{B}_{i2}'\boldsymbol{S}'\boldsymbol{S}\boldsymbol{B}_{i2}\right). \tag{C.84}$$

But, by similar arguments to those used for the mean group estimator in the case when the rank condition does not hold, we can show that

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{X_i' M_q X_i}{T^2} \boldsymbol{\varkappa}_i \xrightarrow{d} N\left(\boldsymbol{0}, \boldsymbol{\Xi}\right),$$

where

$$\boldsymbol{\Xi} = \lim_{N,T\to\infty} \left(\frac{1}{N} \sum_{i=1}^{N} \boldsymbol{\Xi}_{Ti}\right), \tag{C.85}$$

and $\boldsymbol{\Xi}_{Ti} = \text{Var}\left[T^{-2}\boldsymbol{X}_i'\boldsymbol{M}_q\boldsymbol{X}_i\boldsymbol{\varkappa}_i\right]$. Further, by independence of $\boldsymbol{\varepsilon}_i$ across i,

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{\boldsymbol{X}_i'\boldsymbol{M}_q\boldsymbol{\varepsilon}_i}{T^2} = O_p\left(\frac{1}{T}\right).$$

Further, letting $\boldsymbol{Q}_{2iT} = T^{-2}\boldsymbol{X}_i'\boldsymbol{M}_q\boldsymbol{F}$ and $\bar{\boldsymbol{Q}}_{2T} = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{Q}_{2iT}$, we have

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\frac{\boldsymbol{X}_i'\boldsymbol{M}_q\boldsymbol{F}}{T^2}\right)(\boldsymbol{\eta}_i - \bar{\boldsymbol{\eta}}) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\boldsymbol{Q}_{2iT} - \bar{\boldsymbol{Q}}_{2T}\right)\boldsymbol{\eta}_i.$$

Then, similarly to the analysis used above for $T^{-2}\boldsymbol{X}_i'\boldsymbol{M}_q\boldsymbol{X}_i$, we have

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\boldsymbol{Q}_{2iT} - \bar{\boldsymbol{Q}}_{2T}\right)\boldsymbol{\eta}_i \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{\Phi}),$$

where

$$\boldsymbol{\Phi} = \lim_{N,T\to\infty}\left(\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Phi}_{Ti}\right) \tag{C.86}$$

and

$$\boldsymbol{\Phi}_{Ti} = \text{Var}[(\boldsymbol{Q}_{2iT} - \bar{\boldsymbol{Q}}_{2T})\boldsymbol{\eta}_i]. \tag{C.87}$$

Thus, overall, by the independence of $\boldsymbol{\varkappa}_i$ and $\boldsymbol{\eta}_i$, it follows that

$$\sqrt{N}(\hat{\boldsymbol{b}}_P - \boldsymbol{\beta}) \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{\Sigma}_P^*), \quad \text{as } (N,T) \xrightarrow{j} \infty, \tag{C.88}$$

where

$$\boldsymbol{\Sigma}_P^* = \boldsymbol{\Theta}^{-1}(\boldsymbol{\Xi} + \boldsymbol{\Phi})\boldsymbol{\Theta}^{-1}, \tag{C.89}$$

proving the result for the pooled estimator. The result for the consistency of the variance estimator follows along similar lines to that for the mean group estimator. □

Appendix D

Proof of Corollary 1. Using (E.106), we have

$$\sqrt{N}\left(\hat{\boldsymbol{b}}_{MG} - \boldsymbol{\beta}\right) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\boldsymbol{\varkappa}_i + \frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{\Psi}}_{iT}^{-1}\left(\frac{\sqrt{N}\boldsymbol{X}_i'\bar{\boldsymbol{M}}F}{T}\right)\boldsymbol{\gamma}_i + \frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{\Psi}}_{iT}^{-1}\left(\frac{\sqrt{N}\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{\varepsilon}_i}{T}\right), \tag{D.90}$$

where $\hat{\boldsymbol{\Psi}}_{iT} = T^{-1}\boldsymbol{X}_i'\bar{\boldsymbol{M}}X_i$. As we assume that the rank condition (9) is satisfied, we have, by Lemma 4, that

$$\frac{\sqrt{N}\left(\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{F}\right)}{T} = O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\frac{1}{\sqrt{N}}\right), \quad \text{uniformly over } i, \tag{D.91}$$

and so, by the uniform boundedness assumption on $\boldsymbol{\gamma}_i$, and by (A.16), we have that

$$\hat{\boldsymbol{\Psi}}_{iT}^{-1}\left(\frac{\sqrt{N}\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{F}}{T}\right)\boldsymbol{\gamma}_i = O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\frac{1}{\sqrt{N}}\right), \quad \text{uniformly over } i,$$

and so

$$\frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{\Psi}}_{iT}^{-1}\left(\frac{\sqrt{N}\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{F}}{T}\right)\boldsymbol{\gamma}_i = O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\frac{1}{\sqrt{N}}\right).$$

By Lemma 3, we have that

$$\hat{\boldsymbol{\Psi}}_{iT}^{-1}\left(\frac{\sqrt{N}\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{\varepsilon}_i}{T}\right) = O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\frac{1}{\sqrt{N}}\right), \quad \text{uniformly over } i,$$

which implies that

$$\frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{\Psi}}_{iT}^{-1}\left(\frac{\sqrt{N}\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{\varepsilon}_i}{T}\right) = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\boldsymbol{X}_i'\boldsymbol{M}_g\boldsymbol{X}_i}{T}\right)^{-1}\frac{\sqrt{N}\boldsymbol{X}_i'\boldsymbol{M}_g\boldsymbol{\varepsilon}_i}{T} + O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\frac{1}{\sqrt{N}}\right). \tag{D.92}$$

We examine the behaviour of the first term on the RHS of (D.92). We wish to show that

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\left(\frac{\boldsymbol{X}_i'\boldsymbol{M}_g\boldsymbol{X}_i}{T}\right)^{-1}\frac{\boldsymbol{X}_i'\boldsymbol{M}_g\boldsymbol{\varepsilon}_i}{\sqrt{T}} = O_p\left(\frac{1}{\sqrt{T}}\right). \tag{D.93}$$

For this, we have to show that $\left(\frac{\boldsymbol{X}_i'\boldsymbol{M}_g\boldsymbol{X}_i}{T}\right)^{-1}\frac{\boldsymbol{X}_i'\boldsymbol{M}_g\boldsymbol{\varepsilon}_i}{\sqrt{T}}$ has mean zero with bounded variance uniformly over i. We analyze $\frac{\boldsymbol{X}_i'\boldsymbol{M}_g\boldsymbol{\varepsilon}_i}{\sqrt{T}}$. We have $\frac{\boldsymbol{X}_i'\boldsymbol{M}_g\boldsymbol{\varepsilon}_i}{\sqrt{T}} = \frac{\boldsymbol{X}_i'\boldsymbol{M}_g\boldsymbol{M}_g\boldsymbol{\varepsilon}_i}{\sqrt{T}}$. This can then be written as $\frac{\left(\boldsymbol{V}_i'+(\hat{\boldsymbol{\theta}}_{1,i}-\boldsymbol{\theta}_{1,i})\boldsymbol{G}'\right)\left(\boldsymbol{\varepsilon}_i+(\hat{\boldsymbol{\theta}}_{2,i}-\boldsymbol{\theta}_{2,i})\boldsymbol{G}\right)}{\sqrt{T}}$, where $\hat{\boldsymbol{\theta}}_{1,i}$ is the estimated regression coefficient of $\boldsymbol{X}_i'$ on $\boldsymbol{G}$ and $\hat{\boldsymbol{\theta}}_{2,i}$ is the estimated regression coefficient of $\boldsymbol{\varepsilon}_i$ on $\boldsymbol{G}$. But $(\hat{\boldsymbol{\theta}}_{1,i} - \boldsymbol{\theta}_{1,i}) = O_p(T^{-1})$ and $(\hat{\boldsymbol{\theta}}_{2,i} - \boldsymbol{\theta}_{2,i}) = O_p(T^{-1})$. So

$$\frac{\left(\boldsymbol{V}_i' + (\hat{\boldsymbol{\theta}}_{1,i} - \boldsymbol{\theta}_{1,i})\boldsymbol{G}'\right)\left(\boldsymbol{\varepsilon}_i + (\hat{\boldsymbol{\theta}}_{2,i} - \boldsymbol{\theta}_{2,i})\boldsymbol{G}\right)}{\sqrt{T}} = \frac{\boldsymbol{V}_i'\boldsymbol{\varepsilon}_i}{\sqrt{T}} + \frac{(\hat{\boldsymbol{\theta}}_{1,i} - \boldsymbol{\theta}_{1,i})\boldsymbol{G}'\boldsymbol{\varepsilon}_i}{\sqrt{T}} + \frac{\boldsymbol{V}_i'(\hat{\boldsymbol{\theta}}_{2,i} - \boldsymbol{\theta}_{2,i})\boldsymbol{G}}{\sqrt{T}} + \frac{(\hat{\boldsymbol{\theta}}_{1,i} - \boldsymbol{\theta}_{1,i})\boldsymbol{G}'\boldsymbol{G}(\hat{\boldsymbol{\theta}}_{2,i} - \boldsymbol{\theta}_{2,i})}{\sqrt{T}}.$$

But the last three terms are $O_p\left(\frac{1}{\sqrt{T}}\right)$. So it suffices to show that $\frac{\boldsymbol{V}_i'\boldsymbol{\varepsilon}_i}{\sqrt{T}}$ has mean zero with bounded variance uniformly over i, which follows from our assumptions. Thus, from (D.92) and (D.93), we have

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

Hence,

$$\sqrt{N}\left(\hat{\boldsymbol{b}}_{MG} - \boldsymbol{\beta}\right) \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{\Omega}_{\varkappa}), \quad \text{as } (N,T) \xrightarrow{j} \infty. \tag{D.94}$$

$\boldsymbol{\Omega}_{\varkappa}$ can be consistently estimated by

$$\hat{\boldsymbol{\Sigma}}_{MG} = \frac{1}{N-1}\sum_{i=1}^{N}\left(\hat{\boldsymbol{b}}_i - \hat{\boldsymbol{b}}_{MG}\right)\left(\hat{\boldsymbol{b}}_i - \hat{\boldsymbol{b}}_{MG}\right)'. \tag{D.95}$$

To show this, from the proof of Theorem 3, we first note that

$$\left(\hat{\boldsymbol{b}}_i - \hat{\boldsymbol{b}}_{MG}\right) = (\boldsymbol{\beta}_i - \boldsymbol{\beta}) + O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\frac{1}{\sqrt{N}}\right), \quad \text{uniformly over } i,$$

which yields (noting that $\boldsymbol{\beta}_i - \boldsymbol{\beta} = \boldsymbol{\varkappa}_i$)

$$\frac{1}{N-1}\sum_{i=1}^{N}\left(\hat{\boldsymbol{b}}_i - \hat{\boldsymbol{b}}_{MG}\right)\left(\hat{\boldsymbol{b}}_i - \hat{\boldsymbol{b}}_{MG}\right)'$$

$$= \frac{1}{N-1}\sum_{i=1}^{N} \varkappa_i \varkappa_i' + O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\frac{1}{\sqrt{N}}\right).$$

But by the assumption that $\varkappa_i$ has finite fourth moments, and using the law of large numbers for i.i.d. processes, it readily follows that $\hat{\Sigma}_{MG} \to \Omega_{\varkappa}$, as $(N, T) \xrightarrow{j} \infty$. □

Proof of Corollary 2. Assuming that the rank condition is satisfied, $\hat{b}_P$, defined by (20), can be written as

$$\sqrt{N}\left(\hat{b}_P - \beta\right) = \left(\frac{1}{N}\sum_{i=1}^{N}\frac{X_i'\bar{M}X_i}{T}\right)^{-1} \times \left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{X_i'\bar{M}(X_i\varkappa_i + \varepsilon_i)}{T} + q_{NT}\right], \quad (D.96)$$

where

$$q_{NT} = \frac{1}{N}\sum_{i=1}^{N}\frac{\sqrt{N}\left(X_i'\bar{M}F\right)\gamma_i}{T}. \quad (D.97)$$

By (D.91), $q_{NT} = O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\frac{1}{\sqrt{N}}\right)$. Thus,

$$\sqrt{N}\left(\hat{b}_P - \beta\right) = \left(\frac{1}{N}\sum_{i=1}^{N}\frac{X_i'\bar{M}X_i}{T}\right)^{-1} \times \left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{X_i'\bar{M}(X_i\varkappa_i + \varepsilon_i)}{T}\right] + O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\frac{1}{\sqrt{N}}\right). \quad (D.98)$$

Further, by (A.17),

$$\sqrt{N}\left(\hat{b}_P - \beta\right) = \left(\frac{1}{N}\sum_{i=1}^{N}\frac{X_i'\bar{M}X_i}{T}\right)^{-1} \times \left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{X_i'\bar{M}X_i\varkappa_i + X_i'M_g\varepsilon_i}{T}\right] + O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\frac{1}{\sqrt{N}}\right). \quad (D.99)$$

By (A.16) and, since by Assumption 6 $N^{-1}\sum_{i=1}^{N} T^{-1}X_i'\bar{M}X_i$ is nonsingular, we have

$$\left(\frac{1}{N}\sum_{i=1}^{N}\frac{X_i'\bar{M}X_i}{T}\right)^{-1} \xrightarrow{p} \Psi^{*-1},$$

$$\text{where } \Psi^* = \lim_{N\to\infty}\left(N^{-1}\sum_{i=1}^{N}\Sigma_{v_i}\right). \quad (D.100)$$

Next, we examine the second component of the first term of the RHS on (D.98). We first consider $\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{X_i'\bar{M}X_i}{T}\varkappa_i$. We define $\bar{M}_{-i}$ as $\bar{M}_{-i} = I_T - \bar{H}_{-i}\left(\bar{H}_{-i}'\bar{H}_{-i}\right)^{-1}\bar{H}_{-i}'$, where $\bar{H}_{-i} = (D, \bar{Z}_{-i})$, $\bar{Z}_{-i}$ is a $T \times (k+1)$ matrix of observations on d_t and $\bar{z}_{t,-i}$ and $\bar{z}_{t,-i} = \frac{1}{N}\sum_{j=1, j\neq i}^{N} z_{jt}$. Then, it is straightforward to see that $\frac{X_i'\bar{M}X_i}{T} - \frac{X_i'\bar{M}_{-i}X_i}{T} = O_p\left(\frac{1}{N}\right)$, uniformly over i, and so

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{X_i'\bar{M}X_i}{T}\varkappa_i - \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{X_i'\bar{M}_{-i}X_i}{T}\varkappa_i = O_p\left(\frac{1}{N^{1/2}}\right), \quad (D.101)$$

where the uniformity follows by the assumption that $\varkappa_i$ has uniformly finite fourth moments. Since $\varkappa_i$ is i.i.d. and independent of all other stochastic quantities in the model, it follows that $\tilde{\varkappa}_{Ti} = T^{-1}X_i'\bar{M}_{-i}X_i\varkappa_i$ is a martingale difference triangular array, since, for any ordering of the cross-sectional units, $E\left(T^{-1}X_i'\bar{M}_{-i}X_i\varkappa_i | i-1, \ldots, 1\right) = 0$. Then, as long as $E\|T^{-1}X_i'\bar{M}_{-i}X_i\|^2 < \infty$, which is satisfied by Assumption 6, a central limit theorem holds for $\tilde{\varkappa}_{Ti}$, by Theorem 24.3 of Davidson (1994). Also, by Assumption 2(ii) of this paper, Theorem 1 of De Jong (1997) and Example 17.17 of Davidson (1994), it follows that

$$\frac{1}{T}\sum_{t=1}^{T} v_{it}\varepsilon_{it} = O_p\left(\frac{1}{\sqrt{T}}\right), \quad \text{uniformly over } i, \quad (D.102)$$

which implies that $\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{X_i'M_g\varepsilon_i}{T} = O_p\left(\frac{1}{\sqrt{T}}\right)$. Hence, as $(N, T) \xrightarrow{j} \infty$, $\sqrt{N}\left(b \hat{-} \beta\right) \xrightarrow{d} N(\mathbf{0}, \Sigma_P^*)$, where

$$\Sigma_P^* = \Psi^{*-1}R^*\Psi^{*-1}, \qquad R^* = \lim_{N,T\to\infty}\left[N^{-1}\sum_{i=1}^{N}\Sigma_{v\Omega iT}\right], \quad (D.103)$$

where $\Sigma_{v\Omega iT}$ denotes the variance of $\frac{X_i'M_gX_i}{T}\varkappa_i$. The variance estimator for Σ_P^* suggested by Pesaran (2006) is given by

$$\hat{\Sigma}_P^* = \hat{\Psi}^{*-1}\hat{R}^*\hat{\Psi}^{*-1}, \quad (D.104)$$

where

$$\hat{\Psi}^* = N^{-1}\sum_{i=1}^{N}\left(\frac{X_i'\bar{M}X_i}{T}\right),$$

$$\hat{R}^* = \frac{1}{(N-1)}\sum_{i=1}^{N}\left(\frac{X_i'\bar{M}X_i}{T}\right)\left(\hat{b}_i - \hat{b}_{MG}\right) \times \left(\hat{b}_i - \hat{b}_{MG}\right)'\left(\frac{X_i'\bar{M}X_i}{T}\right). \quad (D.105)$$

By a similar argument to that used to show the consistency of the variance estimator in the MG estimator case, it is easy to show that this variance estimator is consistent. □

Appendix E

Proof of Theorem 3. Using (25) in (15), we have

$$\hat{b}_i - \beta_i = \left(\frac{X_i'\bar{M}X_i}{T}\right)^{-1}\left(\frac{X_i'\bar{M}F}{T}\right)\gamma_i + \left(\frac{X_i'\bar{M}X_i}{T}\right)^{-1}\left(\frac{X_i'\bar{M}\varepsilon_i}{T}\right). \quad (E.106)$$

Using (A.17) and (A.18), and assuming that the rank condition (9) is satisfied, we have

$$\hat{b}_i - \beta_i = \left(\frac{X_i'\bar{M}X_i}{T}\right)^{-1}\left(\frac{X_i'M_g\varepsilon_i}{T}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{N}\right). \quad (E.107)$$

For N and T sufficiently large, the distribution of $\sqrt{T}\left(\hat{b}_i - \beta_i\right)$ will be asymptotically normal if the rank condition (9) is satisfied and if $\sqrt{T}/N \to 0$ as N and $T \to \infty$. To see why this additional condition is needed, using (E.107), note that

$$\sqrt{T}\left(\hat{\boldsymbol{b}}_i - \boldsymbol{\beta}_i\right) = \left(\frac{\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{X}_i}{T}\right)^{-1} \frac{\boldsymbol{X}_i'\boldsymbol{M}_g\boldsymbol{\varepsilon}_i}{\sqrt{T}} + O_p\left(\frac{\sqrt{T}}{N}\right) + O_p\left(\frac{1}{\sqrt{N}}\right), \quad \text{(E.108)}$$

and the asymptotic distribution of $\sqrt{T}\left(\hat{\boldsymbol{b}}_i - \boldsymbol{\beta}_i\right)$ will be free of nuisance parameters only if $\sqrt{T}/N \to 0$, as $(N, T) \xrightarrow{j} \infty$. We now give the necessary arguments for showing that the first term on the RHS of (E.108) is asymptotically normally distributed. We note that

$$\frac{\boldsymbol{X}_i'\boldsymbol{M}_g\boldsymbol{\varepsilon}_i}{\sqrt{T}} = -\left(\hat{\boldsymbol{\Pi}}_i - \boldsymbol{\Pi}_i\right)' \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\boldsymbol{g}_t\varepsilon_{it} + \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\boldsymbol{v}_{it}\varepsilon_{it}. \quad \text{(E.109)}$$

But, it is straightforward to show that the first term of (E.109) is $O_p(T^{-1/2})$ when $\boldsymbol{g}_t$ is $I(1)$. Then, we need to obtain a central limit theorem for the second term of (E.109). But, by the martingale difference assumption on ε_{it}, it follows that $\boldsymbol{v}_{it}\varepsilon_{it}$ is also a martingale difference sequence with finite variance given by $\sigma_i^2\boldsymbol{\Sigma}_{\boldsymbol{v}_i}$. Then, by Theorem 24.3 of Davidson (1994), it follows that

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\boldsymbol{v}_{it}\varepsilon_{it} \xrightarrow{d} N(\boldsymbol{0}, \sigma_i^2\boldsymbol{\Sigma}_{\boldsymbol{v}_i}). \quad \text{(E.110)}$$

Further, by (A.16), and noting that, by Assumptions 5 and 6, $\boldsymbol{X}_i'\bar{\boldsymbol{M}}X_i/T$ and $\boldsymbol{X}_i'\boldsymbol{M}_g\boldsymbol{X}_i/T$ are non-singular, we also have

$$\left(\frac{\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{X}_i}{T}\right)^{-1} - \left(\frac{\boldsymbol{X}_i'\boldsymbol{M}_g\boldsymbol{X}_i}{T}\right)^{-1} = O_p\left(\frac{1}{\sqrt{N}}\right),$$

and, by Lemma 5, it follows that $\left(\frac{\boldsymbol{X}_i'\boldsymbol{M}_g\boldsymbol{X}_i}{T}\right)^{-1} - \boldsymbol{\Sigma}_{\boldsymbol{v}_i}^{-1} = O_p\left(\frac{1}{\sqrt{T}}\right)$, finally implying that

$$\sqrt{T}\left(\hat{\boldsymbol{b}}_i - \boldsymbol{\beta}_i\right) \xrightarrow{d} N(0, \sigma_i^2\boldsymbol{\Sigma}_{\boldsymbol{v}_i}^{-1}), \quad \text{(E.111)}$$

and that a consistent estimator of the asymptotic variance can be obtained by

$$\mathring{\sigma}_i^2\left(\frac{\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{X}_i}{T}\right)^{-1},$$

$$\text{where } \mathring{\sigma}_i^2 = \frac{\left(\boldsymbol{y}_i - \boldsymbol{X}_i\hat{\boldsymbol{b}}_i\right)'\bar{\boldsymbol{M}}\left(\boldsymbol{y}_i - \boldsymbol{X}_i\hat{\boldsymbol{b}}_i\right)}{T-(n+2k+1)}. \quad \square \quad \text{(E.112)}$$

References

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

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

Bai, J., 2003. Inferential theory for factor models of large dimensions. Econometrica 71, 135–173.

Bai, J., Kao, C., Ng, S., 2009. Panel cointegration with global stochastic trends. Journal of Econometrics 149, 82–99.

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

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

Chudik, A., Pesaran, M.H., Tosetti, E., 2010. Weak and strong cross section dependence and estimation of large panels. Working Paper Series 1100. European Central Bank. The Econometrics Journal (forthcoming).

Coakley, J., Fuertes, A., Smith, R.P., 2002. A principal components approach to cross-section dependence in panels. Mimeo. Birkbeck College, University of London.

Conley, T.G., Dupor, B., 2003. A spatial analysis of sectoral complementarity. Journal of Political Economy 111, 311–352.

Conley, T.G., Topa, G., 2002. Socio-economic distance and spatial patterns in unemployment. Journal of Applied Econometrics 17, 303–327.

Connor, G., Korajzcyk, R., 1986. Performance measurement with the arbitrage pricing theory: a new framework for analysis. Journal of Financial Economics 15, 373–394.

Connor, G., Korajzcyk, R., 1988. Risk and return in an equilibrium APT: application to a new test methodology. Journal of Financial Economics 21, 255–289.

Davidson, J., 1994. Stochastic Limit Theory. Oxford University Press.

De Jong, R.M., 1997. Central limit theorems for dependent heterogeneous random variables. Econometric Theory 13, 353–367.

Forni, M., Hallin, M., Lippi, M., Reichlin, L., 2000. The generalised factor model: identification and estimation. Review of Economics and Statistics 82, 540–554.

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

Johansen, S., 1995. Likelihood-Based Inference in Cointegrated Vector Autoregressive Models. Oxford University Press, Oxford.

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

Lee, K.C., Pesaran, M.H., 1993. The role of sectoral interactions in wage determination in the UK economy. The Economic Journal 103, 21–55.

Newey, W.K., West, K.D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703–708.

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

Pesaran, M.H., Schuermann, T., Weiner, S.M., 2004. Modeling regional interdependencies using a global error-correcting macroeconomic model. Journal of Business & Economic Statistics 22, 129–181 (with Discussions and a Rejoinder).

Pesaran, M.H., Smith, R.P., 1995. Estimating long-run relationships from dynamic heterogeneous panels. Journal of Econometrics 68, 79–113.

Pesaran, M.H., Smith, R., Im, K.S., 1996. Dynamic linear models for heterogeneous panels. In: Matyas, L., Sevestre, P. (Eds.), The Econometrics of Panel Data. Kluwer, pp. 145–195.

Phillips, P., Moon, H.R., 1999. Linear regression limit theory for nonstationary panel data. Econometrica 67, 1057–1111.

Stock, J.H., Watson, M.W., 2002. Macroeconomic forecasting using diffusion indices. Journal of Business and Economic Statistics 20, 147–162.

Stock, J.H., Watson, M.W., 2008. Forecasting in dynamic factor models subject to structural instability. In: Castle, J., Shephard, N. (Eds.), The Methodology and Practice of Econometrics, A Festschrift in Honour of Professor David F. Hendry. Oxford University Press, Oxford.

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