



## Panels with nonstationary multifactor error structures

G. Kapetanios, M. Hashem Pesaran, T. Yamagata

### ► To cite this version:

G. Kapetanios, M. Hashem Pesaran, T. Yamagata. Panels with nonstationary multifactor error structures. *Econometrics*, 2010, 160 (2), pp.326. 10.1016/j.jeconom.2010.10.001 . hal-00768190

**HAL Id: hal-00768190**

**<https://hal.science/hal-00768190>**

Submitted on 21 Dec 2012

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Accepted Manuscript

Panels with nonstationary multifactor error structures

G. Kapetanios, M. Hashem Pesaran, T. Yamagata

PII: S0304-4076(10)00202-2

DOI: 10.1016/j.jeconom.2010.10.001

Reference: ECONOM 3409

To appear in: *Journal of Econometrics*

Received date: 15 June 2009

Revised date: 14 April 2010

Accepted date: 5 October 2010



Please cite this article as: Kapetanios, G., Hashem Pesaran, M., Yamagata, T., Panels with nonstationary multifactor error structures. *Journal of Econometrics* (2010), doi:10.1016/j.jeconom.2010.10.001

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

# Panels with Nonstationary Multifactor Error Structures\*

G. Kapetanios

Queen Mary, University of London

M. Hashem Pesaran

Cambridge University and USC

T. Yamagata

University of York

September 27, 2010

## Abstract

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. Firstly, 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. Secondly, 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.

Keywords: Cross Section Dependence, Large Panels, Unit Roots, Principal Components, Common Correlated Effects.

JEL-Classification: C12, C13, C33.

---

\*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).

# 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, Schuermann, and Weiner (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 the 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 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 components (PC) analysis of Hotelling 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 per cent 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 Korajczyk (1986) and Connor and Korajczyk (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. A related literature on dynamic factor models has also been put forward by Forni and Reichlin (1998) and Forni, Hallin, Lippi, and Reichlin (2000).

Recent uses of factor models in forecasting focus on consistent estimation of unobserved factors and their loadings. Related theoretical advances by Bai and Ng (2002) and Bai (2003) are also concerned with 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 dependencies 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, Fuertes, and Smith (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, Fuertes, and Smith (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 or multiple unobserved factors and 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 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 distribu-

tional 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 nonstationarity. 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 Appendix contains proofs of the theoretical results. Some more technical results and proofs of Lemmas are relegated to a supplementary appendix that is provided separately from the main paper.

Notations:  $K$  stands for a finite positive constant,  $\|\mathbf{A}\| = [Tr(\mathbf{A}\mathbf{A}')]^{1/2}$  is the Frobenius norm of the  $m \times n$  matrix  $\mathbf{A}$ , and  $\mathbf{A}^+$  denotes the Moore-Penrose inverse of  $\mathbf{A}$ .  $rk(\mathbf{A})$  denotes the rank of  $\mathbf{A}$ .  $\sup_i W_i$  is the supremum of  $W_i$  over  $i$ .  $a_n = O(b_n)$  states the deterministic sequence  $\{a_n\}$  is at most of order  $b_n$ ,  $\mathbf{x}_n = O_p(\mathbf{y}_n)$  states the vector of random variables,  $\mathbf{x}_n$ , is at most of order  $\mathbf{y}_n$  in probability, and  $\mathbf{x}_n = o_p(\mathbf{y}_n)$  is of smaller order in probability than  $\mathbf{y}_n$ ,  $\xrightarrow{q.m.}$  denotes convergence in quadratic mean (or mean square error),  $\xrightarrow{p}$  convergence in probability,  $\xrightarrow{d}$  convergence in distribution, and  $\stackrel{d}{\sim}$  asymptotic equivalence of probability distributions. All asymptotics are carried out under  $N \rightarrow \infty$ , either with a fixed  $T$ , or *jointly* with  $T \rightarrow \infty$ . Joint convergence of  $N$  and  $T$  will be denoted by  $(N, T) \xrightarrow{j} \infty$ . 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  $i^{th}$  cross section unit at time  $t$  for  $i = 1, 2, \dots, N$ ;  $t = 1, 2, \dots, T$ , and suppose that it is generated according to the following linear heterogeneous panel data model

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

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

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

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

Combining (1) and (2) we now have

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

where

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

$$\mathbf{B}_i = \begin{pmatrix} \alpha_i & \mathbf{A}_i \end{pmatrix} \begin{pmatrix} 1 & \mathbf{0} \\ \beta_i & \mathbf{I}_k \end{pmatrix}, \quad \mathbf{C}_i = \begin{pmatrix} \gamma_i & \mathbf{\Gamma}_i \end{pmatrix} \begin{pmatrix} 1 & \mathbf{0} \\ \beta_i & \mathbf{I}_k \end{pmatrix}, \quad (5)$$

<sup>1</sup>However, as will be shown later, our results on the estimators of  $\beta$  hold even if the factor loadings  $\gamma_i$  and/or  $\mathbf{\Gamma}_i$  are zero (or weak in the sense of Chudik, Pesaran, and Tosetti (2010)), and it is not necessary that  $\mathbf{x}_{it}$  and  $\mathbf{f}_t$  are cointegrated. What is required for our results is that conditional on  $\mathbf{d}_t$  and  $\mathbf{f}_t$ , the idiosyncratic errors  $\varepsilon_{it}$  and  $\mathbf{v}_{it}$  are stationary.

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

$$\tilde{\mathbf{\Gamma}}_i = \begin{pmatrix} \gamma_i & \mathbf{\Gamma}_i \end{pmatrix}. \quad (6)$$

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,  $\beta_i$ ,  $i = 1, 2, \dots, N$ . The common factor loadings,  $\alpha_i$  and  $\gamma_i$ , 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  $\beta_i$  and its cross sectional mean cannot be undertaken using standard methods. Pesaran (2006) has suggested using cross section averages of  $y_{it}$  and  $\mathbf{x}_{it}$  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)<sup>2</sup>

$$\bar{\mathbf{z}}_t = \bar{\mathbf{B}}' \mathbf{d}_t + \bar{\mathbf{C}}' \mathbf{f}_t + \bar{\mathbf{u}}_t, \quad (7)$$

where

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

and

$$\bar{\mathbf{B}} = \frac{1}{N} \sum_{i=1}^N \mathbf{B}_i, \quad \bar{\mathbf{C}} = \frac{1}{N} \sum_{i=1}^N \mathbf{C}_i. \quad (8)$$

We distinguish between two important cases: when the rank condition

$$rk(\bar{\mathbf{C}}) = m \leq k + 1, \text{ for all } N, \text{ and as } N \rightarrow \infty, \quad (9)$$

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,  $\bar{\mathbf{z}}_t$  and the observed common components,  $\mathbf{d}_t$ . 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,  $\beta$ , by the CCE procedure.

In the case where the rank condition is met we have

$$\mathbf{f}_t = \left( \bar{\mathbf{C}} \bar{\mathbf{C}}' \right)^{-1} \bar{\mathbf{C}} \left( \bar{\mathbf{z}}_t - \bar{\mathbf{B}}' \mathbf{d}_t - \bar{\mathbf{u}}_t \right). \quad (10)$$

---

<sup>2</sup>Pesaran (2006) considers cross section weighted averages that are more general. But to simplify the exposition we confine our discussion to simple averages throughout.

But since

$$\bar{\mathbf{u}}_t \xrightarrow{q.m.} \mathbf{0}, \text{ as } N \rightarrow \infty, \text{ for each } t, \quad (11)$$

and

$$\bar{\mathbf{C}} \xrightarrow{p} \mathbf{C} = \tilde{\mathbf{\Gamma}} \begin{pmatrix} 1 & \mathbf{0} \\ \boldsymbol{\beta} & \mathbf{I}_k \end{pmatrix}, \text{ as } N \rightarrow \infty, \quad (12)$$

where

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

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

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

This suggests that for sufficiently large  $N$ , it is valid to use  $\bar{\mathbf{h}}_t = (\mathbf{d}_t', \bar{\mathbf{z}}_t')'$  as observable proxies for  $\mathbf{f}_t$ . 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 do 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,  $\hat{\mathbf{b}}_i$  of  $\boldsymbol{\beta}_i$ ,

$$\hat{\mathbf{b}}_{MG} = N^{-1} \sum_{i=1}^N \hat{\mathbf{b}}_i, \quad (14)$$

where

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

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

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

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

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

and

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

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

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

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

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

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

### 3 Theoretical Properties of CCE Estimators in Nonstationary 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,  $\mathbf{g}_t = (\mathbf{d}_{2t}', \mathbf{f}_t')'$ , follows the multivariate unit root process

$$\mathbf{g}_t = \mathbf{g}_{t-1} + \boldsymbol{\zeta}_{gt}$$

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

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

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

$$\text{vec}(\mathbf{B}_i) = \text{vec}(\mathbf{B}) + \boldsymbol{\eta}_{B,i}, \quad \boldsymbol{\eta}_{B,i} \sim \text{IID}(\mathbf{0}, \boldsymbol{\Omega}_{B\eta}), \quad \text{for } i = 1, 2, \dots, N, \quad (21)$$

and

$$\text{vec}(\mathbf{C}_i) = \text{vec}(\mathbf{C}) + \boldsymbol{\eta}_{C,i}, \quad \boldsymbol{\eta}_{C,i} \sim \text{IID}(\mathbf{0}, \boldsymbol{\Omega}_{C\eta}), \quad \text{for } i = 1, 2, \dots, N, \quad (22)$$

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,  $\|\mathbf{B}\| < K$ ,  $\|\mathbf{C}\| < K$ ,  $\|\boldsymbol{\Omega}_{B\eta}\| < K$  and  $\|\boldsymbol{\Omega}_{C\eta}\| < K$ , for some constant  $K$ .

**Assumption 4** (random slope coefficients): The slope coefficients,  $\beta_i$ , follow the random coefficient model

$$\beta_i = \beta + \varkappa_i, \quad \varkappa_i \sim \text{IID}(\mathbf{0}, \boldsymbol{\Omega}_{\varkappa}), \quad \text{for } i = 1, 2, \dots, N, \quad (23)$$

where  $\|\beta\| < K$ ,  $\|\boldsymbol{\Omega}_{\varkappa}\| < K$ , for some constant  $K$ ,  $\boldsymbol{\Omega}_{\varkappa}$  is a  $k \times k$  symmetric non-negative definite matrix, and the random deviations,  $\varkappa_i$ , are distributed independently of  $\gamma_j, \Gamma_j, \varepsilon_{jt}, \mathbf{v}_{jt}$ , and  $\zeta_{gt}$  for all  $i, j$  and  $t$ .  $\varkappa_i$  has finite fourth moments uniformly over  $i$ .

**Assumption 5** (identification of  $\beta_i$  and  $\beta$ ):  $\left(\frac{\mathbf{X}_i' \bar{\mathbf{M}} \mathbf{X}_i}{T}\right)^{-1}$  exists for all  $i$  and  $T$ , and  $\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \Sigma_{\mathbf{v}_i}$  is nonsingular.

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

**Assumption 7** When the rank condition (9) is not satisfied, (i)  $\frac{1}{N} \sum_{i=1}^N \frac{\mathbf{X}_i' \mathbf{M}_q \mathbf{X}_i}{T^2}$  and  $\boldsymbol{\Theta} = \lim_{N, T \rightarrow \infty} \left(\frac{1}{N} \sum_{i=1}^N \boldsymbol{\Theta}_{iT}\right)$ , where  $\boldsymbol{\Theta}_{iT} = E(T^{-2} \mathbf{X}_i' \mathbf{M}_q \mathbf{X}_i)$ , are nonsingular. (ii) If  $m \geq 2k+1$ ,

then  $\left(\frac{\mathbf{X}_i' \mathbf{M}_q \mathbf{X}_i}{T^2}\right)^{-1}$  exists for all  $i$  and  $T$  and  $\sup_i E \left\| \left(\frac{\mathbf{X}_i' \mathbf{M}_q \mathbf{X}_i}{T^2}\right)^{-1} \left(\frac{\mathbf{X}_i' \mathbf{M}_q \mathbf{F}}{T^2}\right) \right\|^2 < \infty$ . (iii) If  $m < 2k + 1$ , then  $E \left\| \frac{\mathbf{F}' \mathbf{F}}{T^2} \right\|^2 < \infty$  and  $E \left\| \left(\frac{\mathbf{F}' \mathbf{F}}{T^2}\right)^{-1} \right\|^2 < \infty$ .

**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  $\varepsilon_{it}$  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.<sup>3</sup> 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  $\gamma_i$  are independently and identically distributed across  $i$ , and

$$\gamma_i = \gamma + \eta_i, \quad \eta_i \sim IID(\mathbf{0}, \mathbf{\Omega}_\eta), \text{ for } i = 1, 2, \dots, N, \quad (24)$$

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

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

$$\mathbf{y}_i = \mathbf{D}\alpha_i + \mathbf{X}_i\beta_i + \mathbf{F}\gamma_i + \varepsilon_i, \quad (25)$$

where  $\varepsilon_i = (\varepsilon_{i1}, \varepsilon_{i2}, \dots, \varepsilon_{iT})'$ . Using (25) in (15) we have

$$\hat{\mathbf{b}}_i - \beta_i = \left(\frac{\mathbf{X}_i' \bar{\mathbf{M}} \mathbf{X}_i}{T}\right)^{-1} \left(\frac{\mathbf{X}_i' \bar{\mathbf{M}} \mathbf{F}}{T}\right) \gamma_i + \left(\frac{\mathbf{X}_i' \bar{\mathbf{M}} \mathbf{X}_i}{T}\right)^{-1} \left(\frac{\mathbf{X}_i' \bar{\mathbf{M}} \varepsilon_i}{T}\right), \quad (26)$$

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

$$\mathbf{X}_i = \mathbf{G}\mathbf{\Pi}_i + \mathbf{V}_i, \quad (27)$$

and

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

---

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

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

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

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

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

and

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

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

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

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

and

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

In the next 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

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

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

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

Using this, we can formally show that

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

Hence

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

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

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

which can be used here as well. The following theorem summarises the results for the mean group estimator. The result is proven in Appendix B.

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

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

where

$$\boldsymbol{\Sigma}_{MG} = \boldsymbol{\Omega}_{\varkappa} + \boldsymbol{\Lambda}, \quad (39)$$

$$\boldsymbol{\Lambda} = \lim_{N, T \rightarrow \infty} \left[ \frac{1}{N} \sum_{i=1}^N \boldsymbol{\Sigma}_{iqT} \right]. \quad (40)$$

and  $\boldsymbol{\Sigma}_{iqT}$  is defined in (A28).  $\boldsymbol{\Sigma}_{MG}$  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,  $\hat{\mathbf{b}}_P$ . The proof is provided in Appendix B.

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

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

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

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

where

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

$\Xi_{Ti} = \text{Var} [T^{-2} \mathbf{X}_i' \mathbf{M}_q \mathbf{X}_i \varepsilon_i]$ , and  $\Phi_{Ti}$  and  $\Theta_{Ti}$  are given by (A48) and (A45), respectively.  $\Sigma_P^*$  can be estimated consistently by

$$\hat{\Sigma}_P^* = \hat{\Psi}^{*-1} \hat{\mathbf{R}}^* \hat{\Psi}^{*-1}, \quad (42)$$

where

$$\hat{\Psi}^* = N^{-1} \sum_{i=1}^N \frac{\mathbf{X}_i' \bar{\mathbf{M}} \mathbf{X}_i}{T}, \quad (43)$$

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

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  $\mathbf{F}\gamma_i = \mathbf{F}_1\delta_{1i} + \mathbf{F}_2\delta_{2i}$  where  $\mathbf{F}_1$  is an  $m-l$ -dimensional  $I(1)$  process with no cointegration whereas  $\mathbf{F}_2$  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, Smith, and Im (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  $\beta_i = \beta$  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,  $\gamma_i$ , 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 Supplementary Appendix B.

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

$$\sqrt{N} \left( \hat{\mathbf{b}}_{MG} - \beta \right) \xrightarrow{d} N(\mathbf{0}, \Sigma_{MG}),$$

where  $\Sigma_{MG}$  is given by  $\Omega_{\varkappa}$ .  $\Sigma_{MG}$  can be consistently estimated by (38).

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

$$\sqrt{N} \left( \hat{\mathbf{b}}_P - \beta \right) \xrightarrow{d} N(\mathbf{0}, \Sigma_P^*),$$

where

$$\Sigma_P^* = \Psi^{*-1} \mathbf{R}^* \Psi^{*-1}, \quad (45)$$

$$\mathbf{R}^* = \lim_{N, T \rightarrow \infty} \left[ N^{-1} \sum_{i=1}^N \Sigma_{v_i \Omega_i T} \right], \quad (46)$$

$$\Psi^* = \lim_{N \rightarrow \infty} \left( N^{-1} \sum_{i=1}^N \Sigma_{v_i} \right), \quad (47)$$

and  $\Sigma_{v_i \Omega_i T}$  denotes the variance of  $\frac{\mathbf{X}_i' \mathbf{M}_g \mathbf{X}_i}{T} \varkappa_i$ .  $\Sigma_P^*$  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  $i$ ,  $\varepsilon_{it}$  is a martingale difference sequence. For each  $i$ ,  $\mathbf{v}_{it}$  is an  $k \times 1$  vector of  $L_{2+\delta}$ ,  $\delta > 0$ , stationary near epoch dependent (NED) process of size  $1/2$ , on some  $\alpha$ -mixing process of size  $-(2 + \delta)/\delta$ .*

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

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

$$\sqrt{T} \left( \hat{\mathbf{b}}_i - \beta_i \right) \xrightarrow{d} N(\mathbf{0}, \Sigma_{b_i}). \quad (48)$$

A consistent estimator of  $\Sigma_{b_i}$  is given by

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

where

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

**Remark 5** Parts of the above result hold under weaker versions of Assumption 8. In particular we note that the central limit theorem in (B72), in Supplementary Appendix C, holds if Assumption 2(ii) holds. However, in this case the asymptotic variance has a different form as autocovariances of  $\varepsilon_{it} \mathbf{v}_{it}$  enter the asymptotic variance expression. If, then, a consistent estimate of the asymptotic variance is required a Newey and West (1987) type correction 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  $(\varepsilon_{it}, \mathbf{v}_{it}')'$  is a strongly mixing process for this consistency to hold.

**Remark 6** It is worth noting that despite the fact that under our Assumptions  $\mathbf{f}_t$ ,  $y_{it}$  and  $\mathbf{x}_{it}$  are  $I(1)$  and cointegrated, implying that  $\varepsilon_{it}$  is an  $I(0)$  process, in the results of Theorem 3, the rate of convergence of  $\hat{\mathbf{b}}_i$  to  $\beta_i$  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  $\mathbf{f}_t$  can be well approximated by the cross section averages, for pedagogic purposes we might as well consider the case where  $\mathbf{f}_t$  is observed. Without loss of generality we also abstract from  $\mathbf{d}_t$ , and substitute (2) in (1) to obtain

$$y_{it} = \beta_i' (\Gamma_i' \mathbf{f}_t + \mathbf{v}_{it}) + \gamma_i' \mathbf{f}_t + \varepsilon_{it} = \boldsymbol{\vartheta}_i' \mathbf{f}_t + \zeta_{it}, \quad (51)$$

where  $\boldsymbol{\vartheta}_i = \Gamma_i \beta_i + \gamma_i$  and  $\zeta_{it} = \varepsilon_{it} + \beta_i' \mathbf{v}_{it}$ . First, it is clear that under our assumptions and for all values of  $\beta_i$ ,  $\zeta_{it}$  is  $I(0)$  irrespective of whether  $\mathbf{f}_t$  is  $I(0)$  or  $I(1)$ . But if  $\mathbf{f}_t$  is  $I(1)$ , since  $\zeta_{it} \sim I(0)$ , then  $y_{it}$  will also be  $I(1)$  and cointegrated with  $\mathbf{f}_t$ . Hence, it follows that  $\boldsymbol{\vartheta}_i$  can be estimated superconsistently. However, the OLS estimator of  $\beta_i$  need not be superconsistent. To see this note that  $\beta_i$  can be estimated equivalently by regressing the residuals from the regressions of  $y_{it}$  on  $\mathbf{f}_t$  on the residuals from the regressions of  $\mathbf{x}_{it}$  on  $\mathbf{f}_t$ . Both these sets of residuals are stationary processes and the resulting estimator of  $\beta_i$  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  $\mathbf{x}_{it}$  in a regression of  $y_{it}$  on  $\mathbf{x}_{it}$  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 can not be derived without further assumptions. However, there is a special case which illustrates the utility of our method. Abstracting from  $\mathbf{d}_t$ , assuming that  $k = m$  and that  $\mathbf{\Gamma}_i$  is invertible, and similarly to (51) write the model for  $y_{it}$  as*

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

where  $\boldsymbol{\varrho}_i' = \beta_i' + \gamma_i' \mathbf{\Gamma}_i^{-1}$  and  $\varsigma_{it} = \varepsilon_{it} - \gamma_i' \mathbf{\Gamma}_i^{-1} \mathbf{v}_{it}$ . Note that  $\varsigma_{it}$  is, by construction, correlated with  $\mathbf{v}_{it}$ . 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  $\varsigma_{it}$  and  $\mathbf{v}_{it}$ . However, in the case of nonstationary 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  $\beta_i$ . To consistently estimate the latter we need to augment the regressions of  $y_{it}$  on  $\mathbf{x}_{it}$  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, Kao, and Ng (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 analysed by Bai, Kao, and Ng (2009), assumes the number of unobserved factors is known and only considers the case where the slopes are homogeneous.<sup>4</sup> In addition, we consider two alternative principal component augmentation approaches discussed in Kapetanios and Pesaran (2007). The first PC approach applies the Bai and Ng (2002) procedure to  $\mathbf{z}_{it} = (y_{it}, \mathbf{x}_{it}')'$  to obtain consistent estimates of the unobserved factors, and then uses the estimated factors to augment the regression (1), and thus produces consistent estimates of  $\beta$ . 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  $\mathbf{x}_{it}$  separately. In the second step  $y_{it}$  and  $\mathbf{x}_{it}$  are regressed

<sup>4</sup>See Bai, Kao, and Ng (2009), for more details.

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 the factors are observable (i.e.  $\bar{\mathbf{z}}_t$  for the CCE estimators is replaced by true factor  $\mathbf{f}_t$ ). 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 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 is computed following Bai, Kao, and Ng (2009).

#### 4.1 Baseline Design

The experimental design of the Monte Carlo study closely follows the one used in Pesaran (2006). Consider the 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

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

for  $i = 1, 2, \dots, N$ , and  $t = 1, 2, \dots, T$ . 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

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

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

$$d_{1t} = 1, \quad d_{2t} = \rho_d d_{2,t-1} + v_{dt}, \quad t = -49, \dots, 1, \dots, T, \\ v_{dt} \sim IIDN(0, 1 - \rho_d^2), \quad \rho_d = 0.5, \quad d_{2,-50} = 0,$$

$$\begin{aligned} v_{ijt} &= \rho_{vij} v_{ijt-1} + \varkappa_{ijt}, \quad t = -49, \dots, 1, \dots, T, \\ \varkappa_{ijt} &\sim IIDN(0, 1 - \rho_{vij}^2), \quad v_{ji, -50} = 0, \end{aligned}$$

and

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

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

$$\begin{aligned} f_{jt} &= f_{jt-1} + v_{fj,t}, \quad \text{for } j = 1, 2, 3, \quad t = -49, \dots, 0, \dots, T, \\ v_{fj,t} &\sim IIDN(0, 1), \quad f_{j, -50} = 0, \quad \text{for } j = 1, 2, 3. \end{aligned} \tag{55}$$

The first 50 observations are discarded.

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,

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

and

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

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

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

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

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

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

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

and

$$E(\tilde{\mathbf{\Gamma}}_{i\mathcal{A}}) = (E(\gamma_{i\mathcal{A}}), E(\mathbf{\Gamma}_i)) = \begin{pmatrix} 1 & 0.5 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0.5 \end{pmatrix}.$$

Under set  $\mathcal{B}$

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

so that

$$E(\tilde{\Gamma}_{i\mathcal{B}}) = (E(\gamma_{i\mathcal{B}}), E(\Gamma_i)) = \begin{pmatrix} 1 & 0.5 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0.5 \end{pmatrix},$$

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  $\eta_{ij} \sim IIDN(0, 0.04)$ , across replications.
- **Experiment 2** considers the case of homogeneous slopes with  $\beta_i = \beta = (1, 1)'$ .

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 the four sets of experiments, it need not be efficient since in these experiments the slope coefficients,  $\beta_i$ , and/or error variances,  $\sigma_i^2$ , differ across  $i$ . As a result the CCE or PC augmented estimators may in fact dominate the infeasible estimator in terms of 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  $\beta_1$  (the cross section mean of  $\beta_{i1}$ ) and the results for  $\beta_2$ , which are very similar to those for  $\beta_1$ , will not be reported. The results for all the estimators considered are reported in Tables 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 of the  $y$  equation (53), where  $\gamma_{i4} \sim IIDN(0.5, 0.2)$ ,  $f_{4t} = f_{4t-1} + v_{f4,t}$ ,  $v_{f4,t} \sim IIDN(0, 1)$ ,  $f_{4,-50} = 0$ . In this case, observe that

$$E(\gamma_i, \Gamma_i)' = \begin{pmatrix} 1 & 1 & 0 & 0.5 \\ 0.5 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0 \end{pmatrix}$$

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 estimated, using the information criterion ' $PC_{p2}$ ' which is proposed by (Bai and Ng, 2002, pp. 201).<sup>5</sup> 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 does not make use of the number of the factors and is 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

$$\begin{aligned} f_{jt}^t &= f_{jt-1}^t + v_{fj,t}^t, \text{ for } j = 1, 2, t = -49, \dots, 0, \dots, T, \\ v_{fj,t}^t &\sim IIDN(0, 1), f_{j,-50}^t = 0, \text{ for } j = 1, 2. \end{aligned} \quad (56)$$

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

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

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

<sup>5</sup>  $PC_{p2}$  is one of the information criteria which performed well in the finite sample investigations reported in Bai and Ng (2002).

#### 4.2.3 Semi-Strong Factor Structure

Chudik, Pesaran, and Tosetti (2010) 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 semi-strong factor structure when the factors are  $I(1)$ . For this purpose, the same DGP of the experiment 1A 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 can not be consistently estimated by the principal components and this could adversely impact the estimators of  $\beta$  that rely on the PC's 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 1A, 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  $\zeta_{jt} \sim IIDN(0, 1)$ , for  $j = 1, 2, 3$ . Results are reported in Table 8.

### 4.3 Results

Results of experiments 1A, 2A, 1B, 2B are summarized in Tables 1 to 4, respectively. We also provide results for the naive estimator (that excludes the unobserved factors or their estimates) and the infeasible estimator (that 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 1A.

As can be seen from Table 1 the naive estimator is substantially biased, performs very poorly and is subject to large size distortions; an outcome that continues to apply in the case of other experiments (not reported here). In contrast, the feasible CCE estimators perform well, have bias 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 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 pre-testing 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 slightly 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 power 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 does not seem to be much affected by the residual serial correlation of the idiosyncratic errors,  $\varepsilon_{it}$ . 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  $\mathbf{f}_t$  and  $\varepsilon_{it}$ 's.

Consider now the PC augmented estimators and recall that they are computed assuming 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 bias and RMSEs of the PC1POOL and PC1MG estimators improve as both  $N$  and  $T$  increase, but the t-tests based on these estimators 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 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 CCEP estimator are smaller than those of 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 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 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 CCE estimators. However, note that in these experiments the number of factors are treated as known, which is rarely expected in practical situation. We return to this issue below.

Tables 5-8 report the results of the experiments carried out as robustness checks.<sup>6</sup> 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.<sup>7</sup> Firstly, despite the number of unobserved factors,  $m = 4$ , exceeding the number of regressors and regressand ( $k + 1 = 3$ ), the RMSEs of 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 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 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, 7 and 8 report the results of experiments with the same DGP as

<sup>6</sup>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.

<sup>7</sup>For small  $N$  and  $T$  the information criterion tends to over-estimate the number of the factors in the first-differenced data  $(y_{it}, x_{1it}, x_{2it})$ , and the estimates tend to four as  $N$  and  $T$  get larger.

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, Pesaran, and Tosetti (2010) and Stock and Watson (2008).

## 5 Conclusions

Recently, there has been increased interest in analysis of panel data models where the standard assumption that the errors of the panel regressions are cross-sectionally uncorrelated 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 parameterisation 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 focusses 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, Kao, and Ng (2009)) show substantial size distortions even in the case of relatively large samples.

## References

- ANDREWS, D. W. (1987): “Asymptotic Results for Generalized Wald Tests,” *Econometric Theory*, 3, 348–358.
- ANSELIN, L. (2001): “Spatial Econometrics,” in *A Companion to Theoretical Econometrics*, ed. by B. Baltagi. Blackwell, Oxford.
- BAI, J. (2003): “Inferential Theory for Factor Models of Large Dimensions,” *Econometrica*, 71, 135–173.
- BAI, J., C. KAO, AND S. NG (2009): “Panel cointegration with global stochastic trends,” *Journal of Econometrics*, 149, 82–99.
- BAI, J., AND S. NG (2002): “Determining the Number of Factors in Approximate Factor Models,” *Econometrica*, 70, 191–221.
- CHAMBERLAIN, G., AND M. ROTHSCILD (1983): “Arbitrage, Factor Structure and Mean-Variance Analysis in Large Asset Markets,” *Econometrica*, 51, 1305–1324.
- CHUDIK, A., M. H. PESARAN, AND E. TOSETTI (2010): “Weak and Strong Cross Section Dependence and Estimation of Large Panels,” *Working Paper Series 1100, European Central Bank, forthcoming in The Econometrics Journal*.
- COAKLEY, J., A. FUERTES, AND R. P. SMITH (2002): “A Principal Components Approach to Cross-Section Dependence in Panels,” *Mimeo, Birkbeck College, University of London*.
- CONLEY, T. G., AND B. DUPOR (2003): “A Spatial Analysis of Sectoral Complementarity,” *Journal of Political Economy*, 111, 311–352.
- CONLEY, T. G., AND G. TOPA (2002): “Socio-economic Distance and Spatial Patterns in Unemployment,” *Journal of Applied Econometrics*, 17, 303–327.
- CONNOR, G., AND R. KORAJCZYK (1986): “Performance Measurement with the Arbitrage Pricing Theory: A New Framework for Analysis,” *Journal of Financial Economics*, 15, 373–394.
- (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., M. HALLIN, M. LIPPI, AND L. REICHLIN (2000): “The Generalised Factor Model: Identification and Estimation,” *Review of Economics and Statistics*, 82, 540–554.
- FORNI, M., AND L. REICHLIN (1998): “Let’s Get Real: A Factor Analytical Approach to Disaggregated Business Cycle Dynamics,” *Review of Economic Studies*, 65, 453–473.
- JOHANSEN, S. (1995): *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press, Oxford.

- KAPETANIOS, G., AND M. H. PESARAN (2007): “Alternative Approaches to Estimation and Inference in Large Multifactor Panels: Small Sample Results with an Application to Modelling of Asset Returns,” in *The Refinement of Econometric Estimation and Test Procedures: Finite Sample and Asymptotic Analysis*, ed. by G. D. A. Phillips, and E. Tzavalis. Cambridge University Press, Cambridge.
- LEE, K. C., AND M. H. PESARAN (1993): “The Role of Sectoral Interactions in Wage Determination in the UK Economy,” *The Economic Journal*, 103, 21–55.
- NEWBY, W. K., AND K. D. WEST (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., T. SCHUERMANN, AND S. M. WEINER (2004): “Modeling Regional Interdependencies using a Global Error-Correcting Macroeconomic Model,” *Journal of Business Economics and Statistics*, (with Discussions and a Rejoinder), 22, 129–181.
- PESARAN, M. H., R. SMITH, AND K. S. IM (1996): “Dynamic Linear Models for Heterogeneous Panels,” in *The Econometrics of Panel Data*, ed. by L. Matyas, and P. Sevestre, pp. 145–195. Kluwer.
- PESARAN, M. H., AND R. P. SMITH (1995): “Estimating Long-Run Relationships from Dynamic Heterogeneous Panels,” *Journal of Econometrics*, 68, 79–113.
- PHILLIPS, P., AND H. R. MOON (1999): “Linear Regression Limit Theory for Nonstationary Panel Data,” *Econometrica*, 67, 1057–1111.
- STOCK, J. H., AND M. W. WATSON (2002): “Macroeconomic Forecasting Using Diffusion Indices,” *Journal of Business and Economic Statistics*, 20, 147–162.
- STOCK, J. H., AND M. W. WATSON (2008): “Forecasting in Dynamic Factor Models Subject to Structural Instability,” in *The Methodology and Practice of Econometrics, A Festschrift in Honour of Professor David F. Hendry*, ed. by J. Castle, and N. Shephard. Oxford University Press, Oxford.
- STONE, R. (1947): “On the Interdependence of Blocks of Transactions,” *Supplement of the Journal of the Royal Statistical Society*, 9, 1–45.

## Appendix A

### Lemmas

Proofs of Lemmas are provided in Supplementary Appendix A.

**Lemma 1** *Under Assumptions 1-4,*

$$\frac{\bar{\mathbf{U}}' \bar{\mathbf{U}}}{T} = O_p\left(\frac{1}{N}\right) \quad (\text{A1})$$

$$\frac{\mathbf{V}_i' \bar{\mathbf{U}}}{T} = O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right), \quad \frac{\boldsymbol{\varepsilon}_i' \bar{\mathbf{U}}}{T} = O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right), \quad \text{uniformly over } i \quad (\text{A2})$$

$$\frac{\mathbf{F}' \bar{\mathbf{U}}}{T} = O_p\left(\frac{1}{\sqrt{N}}\right), \quad \frac{\mathbf{D}' \bar{\mathbf{U}}}{T} = O_p\left(\frac{1}{\sqrt{N}}\right) \quad (\text{A3})$$

$$\frac{\mathbf{X}_i' \bar{\mathbf{U}}}{T} = O_p\left(\frac{1}{\sqrt{N}}\right), \quad \text{uniformly over } i \quad (\text{A4})$$

$$\frac{\mathbf{Q}' \bar{\mathbf{U}}}{T} = O_p\left(\frac{1}{\sqrt{N}}\right) \quad (\text{A5})$$

$$\frac{\mathbf{Q}' \mathbf{Q}}{T^2} = O_p(1) \quad (\text{A6})$$

$$\frac{\mathbf{Q}' \mathbf{X}_i}{T^2} = O_p(1), \quad \text{uniformly over } i \quad (\text{A7})$$

$$\frac{\mathbf{Q}' \mathbf{G}}{T^2} = O_p(1) \quad (\text{A8})$$

$$\frac{\bar{\mathbf{H}}' \bar{\mathbf{H}}}{T^2} = O_p(1) \quad (\text{A9})$$

$$\frac{\bar{\mathbf{H}}' \mathbf{G}}{T^2} = O_p(1) \quad (\text{A10})$$

$$\frac{\bar{\mathbf{H}}' \boldsymbol{\varepsilon}_i}{T} = O_p(1), \quad \text{uniformly over } i \quad (\text{A11})$$

$$\frac{\bar{\mathbf{H}}' \mathbf{V}_i}{T} = O_p(1), \quad \text{uniformly over } i \quad (\text{A12})$$

$$\frac{\bar{\mathbf{H}}' \mathbf{X}_i}{T^2} = O_p(1), \quad \text{uniformly over } i \quad (\text{A13})$$

$$\frac{\bar{\mathbf{H}}' \bar{\mathbf{U}}}{T} = O_p\left(\frac{1}{\sqrt{N}}\right). \quad (\text{A14})$$

**Lemma 2** *Under assumptions 1-4,*

$$\frac{\mathbf{V}_i' \bar{\mathbf{U}}}{T} = O_p\left(\frac{1}{\sqrt{TN}}\right) + O_p\left(\frac{1}{N}\right) \quad \text{uniformly over } i. \quad (\text{A15})$$

**Lemma 3** *Under Assumptions 1-4 and assuming that the rank condition (9) holds, then*

$$\frac{\mathbf{X}_i' \bar{\mathbf{M}} \mathbf{X}_i}{T} - \frac{\mathbf{X}_i' \mathbf{M}_g \mathbf{X}_i}{T} = O_p\left(\frac{1}{\sqrt{N}}\right), \quad \text{uniformly over } i, \quad (\text{A16})$$

$$\frac{\mathbf{X}_i' \bar{\mathbf{M}} \boldsymbol{\varepsilon}_i}{T} - \frac{\mathbf{X}_i' \mathbf{M}_g \boldsymbol{\varepsilon}_i}{T} = O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{N}\right), \quad \text{uniformly over } i. \quad (\text{A17})$$

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

$$\frac{\mathbf{X}_i' \bar{\mathbf{M}} \mathbf{F}}{T} = O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right), \text{ uniformly over } i. \quad (\text{A18})$$

**Lemma 5** Under Assumptions 1-4,

$$\frac{\mathbf{X}_i' \mathbf{M}_g \mathbf{X}_i}{T} - \Sigma_{v_i} = O_p\left(\frac{1}{\sqrt{T}}\right).$$

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

$$\frac{\mathbf{X}_i' \bar{\mathbf{M}} \mathbf{X}_i}{T} - \frac{\mathbf{X}_i' \mathbf{M}_q \mathbf{X}_i}{T} = O_p\left(\frac{1}{\sqrt{N}}\right), \text{ uniformly over } i, \quad (\text{A19})$$

$$\frac{\mathbf{X}_i' \bar{\mathbf{M}} \mathbf{F}}{T} - \frac{\mathbf{X}_i' \mathbf{M}_q \mathbf{F}}{T} = O_p\left(\frac{1}{\sqrt{N}}\right), \text{ uniformly over } i, \quad (\text{A20})$$

$$\frac{\mathbf{X}_i' \bar{\mathbf{M}} \varepsilon_i}{T} - \frac{\mathbf{X}_i' \mathbf{M}_q \varepsilon_i}{T} = O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{N}\right), \text{ uniformly over } i. \quad (\text{A21})$$

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

$$\left(\frac{\mathbf{X}_i' \bar{\mathbf{M}} \mathbf{F}}{T}\right) \bar{\mathbf{C}} = O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right), \text{ uniformly over } i. \quad (\text{A22})$$

## Appendix B: Proofs of theorems for pooled estimators

### Proof of Theorem 1

We know that

$$\bar{\mathbf{C}} = \left( \bar{\gamma} + \bar{\Gamma} \beta + \frac{1}{N} \sum_{i=1}^N \Gamma_i \varkappa_i, \bar{\Gamma} \right),$$

where  $\bar{\Gamma} = \frac{1}{N} \sum_{i=1}^N \Gamma_i$  and  $\bar{\gamma} = \frac{1}{N} \sum_{i=1}^N \gamma_i$ . Substituting this result in (A22) now yields

$$\begin{aligned} \left(\frac{\mathbf{X}_i' \bar{\mathbf{M}} \mathbf{F}}{T}\right) \left(\bar{\gamma} + \bar{\Gamma} \beta + \frac{1}{N} \sum_{i=1}^N \Gamma_i \varkappa_i\right) &= O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right), \text{ uniformly over } i, \\ \left(\frac{\mathbf{X}_i' \bar{\mathbf{M}} \mathbf{F}}{T}\right) \bar{\Gamma} &= O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right), \text{ uniformly over } i, \end{aligned}$$

which in turn yields

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

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

$$\frac{\sqrt{N} (\mathbf{X}_i' \bar{\mathbf{M}} \mathbf{F}) \bar{\gamma}}{T} = O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right), \text{ uniformly over } i. \quad (\text{A23})$$

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

$$\chi_{NT} \equiv \frac{1}{N} \sum_{i=1}^N \left(\frac{\mathbf{X}_i' \bar{\mathbf{M}} \mathbf{X}_i}{T^2}\right)^+ \left(\frac{\sqrt{N} \mathbf{X}_i' \bar{\mathbf{M}} \mathbf{F}}{T^2}\right) (\bar{\gamma} + \eta_i - \bar{\eta}), \quad (\text{A24})$$

where  $\bar{\eta} = \frac{1}{N} \sum_{i=1}^N \eta_i$ . By (A19) and (A20) it follows that

$$\chi_{NT} \equiv \frac{1}{N} \sum_{i=1}^N \left( \frac{\mathbf{X}_i' \mathbf{M}_q \mathbf{X}_i}{T^2} \right)^+ \left( \frac{\sqrt{N} \mathbf{X}_i' \mathbf{M}_q \mathbf{F}}{T^2} \right) (\bar{\gamma} + \eta_i - \bar{\eta}) + O_p \left( \frac{1}{\sqrt{N}} \right). \quad (\text{A25})$$

Note that for the above two expressions, we have changed the normalisation 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  $\mathbf{X}_i' \bar{\mathbf{M}} \mathbf{X}_i$ ,  $\mathbf{X}_i' \bar{\mathbf{M}} \mathbf{F}$ ,  $\mathbf{X}_i' \mathbf{M}_q \mathbf{X}_i$  and  $\mathbf{X}_i' \mathbf{M}_q \mathbf{F}$  would involve nonstationary components. Then, since by (A23),  $\frac{\sqrt{N}(\mathbf{X}_i' \bar{\mathbf{M}} \mathbf{F}) \bar{\gamma}}{T^2} = O_p \left( \frac{1}{T\sqrt{N}} \right) + O_p \left( \frac{1}{T^{3/2}} \right)$ , uniformly over  $i$ , it is the case that for  $N$  and  $T$  large

$$\sqrt{N} (\hat{\mathbf{b}}_{MG} - \beta) \stackrel{d}{\sim} \frac{1}{\sqrt{N}} \sum_{i=1}^N \varkappa_i + \frac{1}{\sqrt{N}} \sum_{i=1}^N \left( \frac{\mathbf{X}_i' \mathbf{M}_q \mathbf{X}_i}{T^2} \right)^+ \left( \frac{\mathbf{X}_i' \mathbf{M}_q \mathbf{F}}{T^2} \right) (\eta_i - \bar{\eta}). \quad (\text{A26})$$

We next focus on analysing the RHS of (A26). The first term on the RHS of (A26) tends to a Normal density with mean zero and finite variance. The second term needs further analysis. Letting

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

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

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N \mathbf{Q}_{1iT} (\eta_i - \bar{\eta}) = \frac{1}{\sqrt{N}} \sum_{i=1}^N (\mathbf{Q}_{1iT} - \bar{\mathbf{Q}}_{1T}) \eta_i. \quad (\text{A27})$$

We note that  $\eta_i$  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 (A27). 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

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

and

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

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  $y_{Ti} = (\mathbf{Q}_{1iT,-i} - \bar{\mathbf{Q}}_{1T,-i})$  and  $x_i = \eta_i$ . Using this result, it follows that the second term on the RHS of (A26) tends to a Normal density if  $(\mathbf{Q}_{1iT} - \bar{\mathbf{Q}}_{1T}) \eta_i$  has variance with finite norm, uniformly over  $i$ , denoted by  $\Sigma_{iqT}$ , i.e.

$$\Sigma_{iqT} = \text{Var} [(\mathbf{Q}_{1iT} - \bar{\mathbf{Q}}_{1T}) \eta_i]. \quad (\text{A28})$$

In order to establish the existence of second moments, it is sufficient to prove that  $\|(\mathbf{Q}_{1iT} - \bar{\mathbf{Q}}_{1T})\|$ , or equivalently  $\|(\mathbf{Q}_{1iT,-i} - \bar{\mathbf{Q}}_{1T,-i})\|$ , has finite second moments. We carry out the analysis for  $\|(\mathbf{Q}_{1iT} - \bar{\mathbf{Q}}_{1T})\|$ . 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

$$\mathbf{X}_i = \mathbf{Q} \mathbf{B}_{i1} + \mathbf{S} \mathbf{B}_{i2} + \mathbf{V}_i, \quad (\text{A29})$$

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

$$\mathbf{F} = \mathbf{Q}\mathbf{K}_1 + \mathbf{S}\mathbf{K}_2, \quad (\text{A30})$$

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,

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

But, it easily follows that

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

and

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

Then,

$$\frac{\mathbf{X}'_i \mathbf{M}_q \mathbf{X}_i}{T^2} = \mathbf{B}'_{i2} \frac{\mathbf{S}' \mathbf{S}}{T^2} \mathbf{B}_{i2} + O_p\left(\frac{1}{T}\right), \text{ uniformly over } i. \quad (\text{A31})$$

Similarly, using (A30),

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

Thus

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

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)  $\|(\mathbf{Q}_{1iT} - \bar{\mathbf{Q}}_{1T})\|$  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  $(\tilde{\mathbf{B}}_{i2}' \tilde{\mathbf{B}}_{i2})^+ = \tilde{\mathbf{B}}_{i2}^+ \tilde{\mathbf{B}}_{i2}^+$  and since in this case  $\mathbf{B}_{i2}$  has full row rank then

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

and we obtain

$$\left(\mathbf{B}'_{i2} \frac{\mathbf{S}' \mathbf{S}}{T^2} \mathbf{B}_{i2}\right)^+ = \mathbf{B}'_{i2} (\mathbf{B}_{i2} \mathbf{B}'_{i2})^{-1} \left(\frac{\mathbf{S}' \mathbf{S}}{T^2}\right)^{-1} (\mathbf{B}_{i2} \mathbf{B}'_{i2})^{-1} \mathbf{B}_{i2}. \quad (\text{A32})$$

Hence

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

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,

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

and the analysis proceeds as above until

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

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

Thus, in general we have that

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

where

$$\boldsymbol{\Sigma}_{MG} = \boldsymbol{\Omega}_\kappa + \boldsymbol{\Lambda}, \quad (\text{A33})$$

and

$$\boldsymbol{\Lambda} = \lim_{N, T \rightarrow \infty} \left[ \frac{1}{N} \sum_{i=1}^N \boldsymbol{\Sigma}_{iqT} \right]. \quad (\text{A34})$$

To complete the proof we have to consider two further issues. First we note that in (A26), we disregard a term involving  $\left(\frac{\mathbf{X}_i' \mathbf{M}_q \mathbf{X}_i}{T^2}\right)^+ \left(\frac{\mathbf{X}_i' \mathbf{M}_q \boldsymbol{\varepsilon}_i}{T}\right)$ . In particular we have to prove that

$$\left(\frac{1}{T}\right) \frac{1}{\sqrt{N}} \sum_{i=1}^N \left(\frac{\mathbf{X}_i' \mathbf{M}_q \mathbf{X}_i}{T^2}\right)^+ \left(\frac{\mathbf{X}_i' \mathbf{M}_q \boldsymbol{\varepsilon}_i}{T}\right) = O_p\left(\frac{1}{T}\right). \quad (\text{A35})$$

For this, it is enough to show that  $\frac{1}{\sqrt{N}} \sum_{i=1}^N \left(\frac{\mathbf{X}_i' \mathbf{M}_q \mathbf{X}_i}{T^2}\right)^+ \left(\frac{\mathbf{X}_i' \mathbf{M}_q \boldsymbol{\varepsilon}_i}{T}\right)$  follows a CLT. This follows if (i) for any ordering of the cross-sectional units,  $\frac{\mathbf{X}_i' \mathbf{M}_q \boldsymbol{\varepsilon}_i}{T}$  is a martingale difference and (ii)  $\left(\frac{\mathbf{X}_i' \mathbf{M}_q \mathbf{X}_i}{T^2}\right)^+ \left(\frac{\mathbf{X}_i' \mathbf{M}_q \boldsymbol{\varepsilon}_i}{T}\right)$  has finite second moments. (ii) follows easily from the argument made in other parts of the appendix about the existence of moments of  $\left(\frac{\mathbf{X}_i' \mathbf{M}_q \mathbf{X}_i}{T^2}\right)^+ \left(\frac{\mathbf{X}_i' \mathbf{M}_q \mathbf{F}}{T}\right)$ . Then, one has to simply prove (i). We need to show that for any ordering

$$E(Q_i^* | Q_{i-1}^*) = 0, \quad (\text{A36})$$

where  $Q_i^* = \left(\frac{\mathbf{X}_i' \mathbf{M}_q \mathbf{X}_i}{T^2}\right)^+ \left(\frac{\mathbf{X}_i' \mathbf{M}_q \boldsymbol{\varepsilon}_i}{T}\right)$ . Denote  $Q_i^{**} = \left(\frac{\mathbf{X}_i' \mathbf{M}_q \mathbf{X}_i}{T^2}\right)^+$ . Then  $Q_i^* = Q_i^{**} \left(\frac{\mathbf{X}_i' \mathbf{M}_q \boldsymbol{\varepsilon}_i}{T}\right)$ . Now  $\frac{\mathbf{X}_i' \mathbf{M}_q \boldsymbol{\varepsilon}_i}{T} = \frac{1}{T} \sum_{t=1}^T s_t \varepsilon_{it}$  where  $s_t$  is a unit root process (cf. the definition of  $\mathbf{S}$  in (A29) above). Then, for (A36) to hold it is sufficient to note that for all  $t, l$ ,  $E(Q_i^{**} s_t \varepsilon_{it} | Q_i^{**} s_l \varepsilon_{i-l}) = 0$ . This completes the proof of (A35).

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

$$\hat{\boldsymbol{\Sigma}}_{MG} = \frac{1}{N-1} \sum_{i=1}^N (\hat{\mathbf{b}}_i - \hat{\mathbf{b}}_{MG}) (\hat{\mathbf{b}}_i - \hat{\mathbf{b}}_{MG})', \quad (\text{A37})$$

which is consistent. To see this first note that

$$\hat{\mathbf{b}}_i - \boldsymbol{\beta} = \boldsymbol{\kappa}_i + \mathbf{h}_{iT} + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right), \text{ uniformly over } i, \quad (\text{A38})$$

where

$$\mathbf{h}_{iT} = \left(\frac{\mathbf{X}_i' \bar{\mathbf{M}} \mathbf{X}_i}{T^2}\right)^+ \frac{\mathbf{X}_i' \bar{\mathbf{M}} [\mathbf{F}(\boldsymbol{\eta}_i - \bar{\boldsymbol{\eta}}) + \boldsymbol{\varepsilon}_i]}{T^2}, \quad (\text{A39})$$

and so

$$\hat{\mathbf{b}}_i - \hat{\mathbf{b}}_{MG} = (\varkappa_i - \bar{\varkappa}) + (\mathbf{h}_{iT} - \bar{\mathbf{h}}_T) + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right), \text{ uniformly over } i, \quad (\text{A40})$$

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

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

and the desired result follows.

## Proof of Theorem 2

As before the pooled estimator,  $\hat{\mathbf{b}}_P$ , defined by (20), can be written as

$$\sqrt{N}(\hat{\mathbf{b}}_P - \boldsymbol{\beta}) = \left( \frac{1}{N} \sum_{i=1}^N \frac{\mathbf{X}_i' \bar{\mathbf{M}} \mathbf{X}_i}{T^2} \right)^{-1} \left[ \frac{1}{\sqrt{N}} \sum_{i=1}^N \frac{\mathbf{X}_i' \bar{\mathbf{M}} (\mathbf{X}_i \varkappa_i + \boldsymbol{\varepsilon}_i)}{T^2} + \mathbf{q}_{NT} \right], \quad (\text{A41})$$

where

$$\mathbf{q}_{NT} = \frac{1}{\sqrt{N}} \sum_{i=1}^N \frac{(\mathbf{X}_i' \bar{\mathbf{M}} \mathbf{F}) \boldsymbol{\gamma}_i}{T^2}. \quad (\text{A42})$$

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

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

But by (A23), the first component of  $\mathbf{q}_{NT}$  is  $O_p\left(\frac{1}{T\sqrt{N}}\right) + O_p\left(\frac{1}{T^{3/2}}\right)$ . Substituting this result in (A41), and making use of (33) and (34) we have

$$\begin{aligned} \sqrt{N}(\hat{\mathbf{b}}_P - \boldsymbol{\beta}) &= \left( \frac{1}{N} \sum_{i=1}^N \frac{\mathbf{X}_i' \mathbf{M}_q \mathbf{X}_i}{T^2} \right)^{-1} \left[ \frac{1}{\sqrt{N}} \sum_{i=1}^N \frac{\mathbf{X}_i' \mathbf{M}_q (\mathbf{X}_i \varkappa_i + \boldsymbol{\varepsilon}_i + \mathbf{F}(\boldsymbol{\eta}_i - \bar{\boldsymbol{\eta}}))}{T^2} \right] + \\ &O_p\left(\frac{1}{T\sqrt{N}}\right) + O_p\left(\frac{1}{T^{3/2}}\right). \end{aligned} \quad (\text{A43})$$

Also by Assumption 7, when the rank condition is not satisfied,  $\frac{1}{N} \sum_{i=1}^N \frac{\mathbf{X}_i' \mathbf{M}_q \mathbf{X}_i}{T^2}$  is nonsingular. Further, by (A31),

$$\frac{1}{N} \sum_{i=1}^N \frac{\mathbf{X}_i' \mathbf{M}_q \mathbf{X}_i}{T^2} = \frac{1}{N} \sum_{i=1}^N \mathbf{B}_{i2}' \frac{\mathbf{S}' \mathbf{S}}{T^2} \mathbf{B}_{i2} + O_p\left(\frac{1}{T}\right).$$

We note that, by assumption 3,  $\mathbf{B}_{i2}$  is an i.i.d. sequence with finite second moments. Further, by Assumption 7, it follows that  $E \left\| \frac{\mathbf{S}' \mathbf{S}}{T^2} \right\|^2 < \infty$ . Hence,  $T^{-2} \mathbf{B}_{i2}' \mathbf{S}' \mathbf{S} \mathbf{B}_{i2}$  forms asymptotically a martingale difference triangular array with finite mean and variance and, as a result,  $T^{-2} \mathbf{B}_{i2}' \mathbf{S}' \mathbf{S} \mathbf{B}_{i2}$  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 nonstochastic limit which we denote by  $\boldsymbol{\Theta}$ , i.e.

$$\boldsymbol{\Theta} = \lim_{N, T \rightarrow \infty} \left( \frac{1}{N} \sum_{i=1}^N \boldsymbol{\Theta}_{iT} \right), \quad (\text{A44})$$

where

$$\boldsymbol{\Theta}_{iT} = E(T^{-2} \mathbf{B}_{i2}' \mathbf{S}' \mathbf{S} \mathbf{B}_{i2}). \quad (\text{A45})$$

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

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N \frac{\mathbf{X}_i' \mathbf{M}_q \mathbf{X}_i}{T^2} \varkappa_i \xrightarrow{d} N(\mathbf{0}, \Xi),$$

where

$$\Xi = \lim_{N, T \rightarrow \infty} \left( \frac{1}{N} \sum_{i=1}^N \Xi_{Ti} \right), \quad (\text{A46})$$

and  $\Xi_{Ti} = \text{Var} [T^{-2} \mathbf{X}_i' \mathbf{M}_q \mathbf{X}_i \varkappa_i]$ . Further, by independence of  $\varepsilon_i$  across  $i$ ,

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

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

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

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

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N (\mathbf{Q}_{2iT} - \bar{\mathbf{Q}}_{2T}) \eta_i \xrightarrow{d} N(\mathbf{0}, \Phi)$$

where

$$\Phi = \lim_{N, T \rightarrow \infty} \left( \frac{1}{N} \sum_{i=1}^N \Phi_{Ti} \right) \quad (\text{A47})$$

and

$$\Phi_{Ti} = \text{Var} [(\mathbf{Q}_{2iT} - \bar{\mathbf{Q}}_{2T}) \eta_i]. \quad (\text{A48})$$

Thus, overall by the independence of  $\varkappa_i$  and  $\eta_i$ , it follows that

$$\sqrt{N} (\hat{\mathbf{b}}_P - \beta) \xrightarrow{d} N(\mathbf{0}, \Sigma_P^*), \text{ as } (N, T) \xrightarrow{j} \infty, \quad (\text{A49})$$

where

$$\Sigma_P^* = \Theta^{-1} (\Xi + \Phi) \Theta^{-1} \quad (\text{A50})$$

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.

Table 1: Small Sample Properties of Common Correlated Effects Type Estimators  
in the Case of Experiment 1.4 (Heterogeneous Slopes + Full Rank)

(N,T)	Bias ( $\times 100$ )					Root Mean Square Errors ( $\times 100$ )					Size (5% level, $H_0: \beta_1 = 1.00$ )					Power (5% level, $H_1: \beta_1 = 0.95$ )				
CCE Type Estimators	20	30	50	100	200	20	30	50	100	200	20	30	50	100	200	20	30	50	100	200
CCCEMG																				
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	35.35	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 $f_{it}$ 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	5.20	5.70	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	5.30	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
Naïve Estimators (excluding $f_{it}$ and $f_{2t}$ )																				
Naïve MG																				
20	22.18	23.13	26.86	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
Naïve Pooled																				
20	25.25	26.60	31.27	33.59	34.84	35.30	37.01	42.66	45.42	47.67	42.15	46.65	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	67.75	64.55	64.55
50	26.54	28.75	30.39	34.01	35.88	35.61	37.39	38.05	44.04	45.93	64.75	67.15	69.25	70.35	69.35	73.55	76.25	78.25	77.45	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	86.60
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	87.40	87.40	87.20	89.95	91.90	93.55	92.20	92.20

(Table 1 Continued)

		Bias ( $\times 100$ )					Root Mean Square Errors ( $\times 100$ )					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
Principal Component Estimators, Augmented																					
PC1MG		-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
20		-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
30		-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
50		-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
100		-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
200																					
PC1POOL		-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
20		-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
30		-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
50		-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
100		-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
200																					
Principal Component Estimators, Orthogonalised																					
PC2MG		-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
20		-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
30		-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
50		-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
100		-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
200																					
PC2POOL		-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
20		-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
30		-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
50		-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
100		-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
200																					

Notes: The DGP is  $y_{it} = \alpha_1 d_{it} + \beta_1 x_{1t} + \beta_2 x_{2t} + \gamma_{11} f_{1t} + \gamma_{12} f_{2t} + \varepsilon_{it}$  with  $\varepsilon_{it} = \rho_{i\ell} \varepsilon_{i,t-1} + \sigma_i (1 - \rho_{i\ell}^2)^{1/2} \omega_{it}$ ,  $i = 1, 2, \dots, [N/2]$ , and  $\varepsilon_{it} = \sigma_i (1 + \theta_{i\ell}^2)^{-1/2} (\omega_{it} + \theta_{i\ell} \omega_{i,t-1})$ ,  $i = [N/2] + 1, \dots, N$ ,  $\omega_{it} \sim IIDN(0, 1)$ ,  $\sigma_i^2 \sim IIDU[0.5, 1.5]$ ,  $\rho_{i\ell} \sim IIDU[0.05, 0.95]$ ,  $\theta_{i\ell} \sim IIDU[0, 1]$ . Regressors are generated by  $x_{1t} = \alpha_{11} d_{it} + \alpha_{12} d_{it} + \gamma_{13} f_{1t} + \gamma_{13} f_{2t} + v_{1t}$ ,  $j = 1, 2$ , for  $i = 1, 2, \dots, N$ ,  $d_{it} = 1$ ,  $d_{it} = 0.5 d_{it-1} + v_{dt}$ ,  $v_{dt} \sim IIDN(0, 1 - 0.5^2)$ ,  $d_{2t} = 0$ ,  $f_{1t} = f_{1t-1} + v_{f1,t}$ ,  $v_{f1,t} \sim IIDN(0, 1)$ ,  $f_{2t} = 0$ , for  $j = 1, 2, 3$ ;  $v_{1t} \sim IIDN(1, 1)$ ,  $v_{1t} \sim IIDN(0.5, 0.5)$  for  $j = 1, 2$ ,  $\ell = 1, 2$ ;  $\gamma_{11}$  and  $\gamma_{12} \sim IIDU[0.05, 0.95]$  for  $j = 1, 2$ , for  $i = -49, \dots, T$  with the first 50 observations discarded;  $\alpha_{11} \sim IIDN(1, 1)$ ,  $\alpha_{12} \sim IIDN(0.5, 0.5)$  for  $j = 1, 2$ ,  $\ell = 1, 2$ ;  $\gamma_{11}$  and  $\gamma_{12} \sim IIDU[0.05, 0.95]$  for  $j = 1, 2$ , for  $i = -49, \dots, T$ ;  $\gamma_{11}$  and  $\gamma_{12} \sim IIDN(1, 0.2)$ ,  $\beta_{1j} = 1 + \eta_{1j}$  with  $\eta_{1j} \sim IIDN(0, 0.04)$  for  $j = 1, 2$ ,  $\rho_{i\ell}$ ,  $\rho_{i\ell}$ ,  $\theta_{i\ell}$ ,  $\sigma_i^2$ ,  $\alpha_{11}$ ,  $\alpha_{12}$  for  $j = 1, 2$ ,  $\ell = 1, 2$  are fixed across replications. COEMG and CCEP are defined by (14) and (20). CupBC is bias-corrected, iterated principal component estimator of Bai et al. (2009). The PC1 and PC2 estimators are of 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 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 2.4 (Homogeneous Slopes + Full Rank)

(N,T)	Bias ( $\times 100$ )					Root Mean Square Errors ( $\times 100$ )					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	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	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	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	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	5.60	4.45	6.35	5.20	5.70	44.20	67.95	91.90	99.90	100.00
CCEP	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	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	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	2.87	2.37	1.78	1.24	0.93	5.60	6.20	6.40	5.25	5.35	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	5.60	3.95	5.70	5.60	5.35	65.25	84.40	96.95	100.00	100.00
CupBC	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	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	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	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	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  $\beta_{ij} = 1$  for all  $i$  and  $j$ ,  $i = 1, 2, \dots, N$ ,  $j = 1, 2$ . See notes to Table 1.

Table 3: Small Sample Properties of Common Correlated Effects Type Estimators in the Case of Experiment 1b (Heterogeneous Slopes + Rank Deficient)

(N,T)	Bias ( $\times 100$ )					Root Mean Square Errors ( $\times 100$ )					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	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	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	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	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	19.45	23.70	29.75	37.25	43.45
CCEP	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	9.30	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	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	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	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	5.40	4.70	5.25	4.10	3.95	25.75	28.50	34.50	41.10	46.05
CupBC	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	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	4.98	4.38	4.32	64.05	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	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	61.80	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  $\gamma_{12} \sim IIDN(0, 1)$ , so that 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 2b (Homogeneous Slopes + Rank Deficient)

(N,T)	Bias ( $\times 100$ )					Root Mean Square Errors ( $\times 100$ )					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	-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
	-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
	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
	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
	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	-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
	-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
	-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
	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
	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	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
	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
	0.12	0.36	0.03	0.13	0.07	4.90	3.62	2.82	1.29	0.70	60.90	47.10	32.85	20.00	14.75	77.00	82.25	88.35	98.95	100.00
	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
	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  $\gamma_{i2} \sim IIDN(0,1)$ , so that the rank condition is not satisfied, and  $\beta_{ij} = 1$  for all  $i$  and  $j$ ,  $i = 1, 2, \dots, N$ ,  $j = 1, 2$ . See notes to Table 1.

Table 5: Small Sample Properties of Common Correlated Effects Type Estimators, The Number of Factors  $m = 4$  Exceeds  $k + 1 = 3$ , In the case of Heterogeneous Slopes

(N,T)	Bias ( $\times 100$ )					Root Mean Square Errors ( $\times 100$ )				
	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
	30	0.20	0.08	-0.07	0.14	-0.03	8.98	7.65	6.84	6.29
	50	-0.04	0.00	-0.16	-0.19	0.14	6.81	6.03	5.12	4.71
	100	0.12	-0.06	0.01	-0.01	0.12	4.81	4.25	3.69	3.46
	200	0.01	-0.04	0.03	-0.04	-0.10	3.78	3.08	2.84	2.53
CCEP	20	0.09	0.50	-0.02	-0.22	-0.11	9.57	8.94	8.07	7.70
	30	0.03	-0.05	-0.08	0.04	-0.09	7.96	7.21	6.60	6.25
	50	-0.04	-0.05	-0.13	-0.14	0.13	6.06	5.59	4.85	4.49
	100	0.06	-0.07	-0.01	0.01	0.11	4.21	3.85	3.51	3.37
	200	-0.04	-0.05	0.00	-0.03	-0.10	3.13	2.74	2.62	2.37
CupBC	20	0.49	0.32	0.06	0.11	0.11	11.56	10.26	8.94	7.09
	30	0.01	0.12	0.12	0.21	0.07	9.38	7.98	6.68	5.58
	50	-0.11	0.25	-0.08	-0.02	0.21	7.07	6.29	5.03	4.04
	100	0.06	0.04	0.10	0.04	-0.04	4.81	4.32	3.58	2.82
	200	0.05	-0.11	0.00	0.00	-0.03	3.55	3.14	2.61	2.00

Notes: The DGP is the same as that of Table 1, except an extra term  $\gamma_{14}f_{14}t$  is added to the  $y$  equation, where  $\gamma_{14} \sim IIDN(0.5, 0.2)$ ,  $f_{14}t = f_{14}t - 1 + v_{f_{14}}t \sim IIDN(0, 1)$ ,  $f_{14} - 50 = 0$ . For CupBC estimator, the number of unobserved factors is treated unknown but estimated by the information criterion  $PCP_2$ , which is proposed by Bai and NG (2002). We set the maximum number of factors to six. 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 1A

(N,T)	Bias ( $\times 100$ )					Root Mean Square Errors ( $\times 100$ )				
	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
	30	0.02	0.01	0.01	-0.11	0.10	7.74	6.21	5.14	4.43
	50	-0.12	0.16	-0.11	0.14	-0.04	5.96	4.57	3.99	3.42
	100	0.01	0.03	0.05	0.02	0.04	4.23	3.51	2.87	2.33
	200	-0.06	0.01	-0.01	0.05	0.00	3.06	2.46	2.00	1.72
CCEP	20	0.09	-0.07	-0.06	0.04	-0.12	8.64	7.49	6.34	5.65
	30	-0.19	-0.10	0.09	-0.08	0.13	7.12	5.90	5.12	4.49
	50	0.01	0.13	-0.05	0.13	-0.02	5.27	4.46	3.93	3.43
	100	0.04	0.08	0.02	0.00	0.03	3.77	3.28	2.84	2.35
	200	-0.07	-0.03	-0.04	0.05	0.00	2.68	2.30	1.96	1.70
CupBC	20	0.23	0.46	0.17	0.43	0.45	12.29	10.55	8.09	6.75
	30	-0.20	0.09	0.38	0.20	0.49	9.53	8.03	6.39	5.14
	50	0.39	0.07	0.06	0.20	0.15	7.34	6.08	5.07	3.99
	100	0.18	0.18	0.06	0.05	0.09	4.99	4.40	3.61	2.69
	200	0.00	0.03	0.03	0.09	0.01	3.77	3.03	2.55	1.98

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

Table 6: Small Sample Properties of Common Correlated Effects Type Estimators, Heterogeneous Slopes and Full Rank, Cointegrated Factors, in the Case of Experiment 1A

(N,T)	Bias ( $\times 100$ )					Root Mean Square Errors ( $\times 100$ )				
	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
	30	-0.14	0.09	0.03	-0.02	0.02	7.35	6.02	5.18	4.54
	50	-0.03	0.14	-0.05	0.11	0.11	5.85	4.70	4.06	3.49
	100	-0.05	-0.01	0.03	-0.05	0.00	4.15	3.40	2.87	2.19
	200	-0.05	0.14	0.03	0.04	-0.04	3.08	2.46	2.02	1.59
CCEP	20	-0.06	-0.01	-0.23	0.06	-0.01	8.52	7.54	6.65	5.95
	30	-0.06	-0.07	-0.07	-0.02	0.01	6.78	5.90	5.25	4.70
	50	-0.03	0.14	-0.09	0.12	0.13	5.35	4.54	4.05	3.55
	100	-0.02	0.03	0.06	-0.03	-0.02	3.77	3.18	2.84	2.50
	200	-0.04	0.10	-0.01	0.05	-0.04	2.70	2.33	1.99	1.72
CupBC	20	0.54	0.85	0.61	0.68	0.78	11.01	9.58	8.01	6.94
	30	0.69	0.52	0.54	0.50	0.68	8.65	7.48	6.26	5.39
	50	0.49	0.54	0.50	0.53	0.58	6.82	5.69	4.99	4.24
	100	0.33	0.37	0.38	0.22	0.26	4.61	3.86	3.43	2.84
	200	0.13	0.31	0.13	0.25	0.09	3.39	2.88	2.41	2.03

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

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 Errors ( $\times 100$ )				
	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
	30	0.14	-0.03	-0.02	-0.13	0.10	7.68	6.08	5.11	4.54
	50	-0.21	0.20	-0.11	0.14	-0.04	5.91	4.64	4.01	3.43
	100	0.02	0.03	0.05	0.03	0.04	4.26	3.48	2.88	2.33
	200	-0.08	-0.02	-0.02	0.06	0.00	3.08	2.49	2.01	1.51
CCEP	20	0.17	0.00	-0.05	0.00	-0.13	8.73	7.61	6.86	6.21
	30	-0.15	-0.13	0.07	-0.14	0.14	7.10	5.98	5.31	4.78
	50	-0.03	0.18	-0.06	0.11	-0.01	5.30	4.53	3.97	3.47
	100	0.05	0.09	0.04	0.01	0.02	3.80	3.26	2.85	2.34
	200	-0.06	-0.04	-0.05	0.05	0.00	2.72	2.29	1.95	1.71
CupBC	20	0.52	0.77	0.79	0.80	0.89	11.18	9.87	8.39	7.52
	30	0.32	0.58	0.77	0.58	0.84	8.91	7.80	6.55	5.68
	50	0.58	0.75	0.38	0.61	0.54	6.78	6.01	5.03	4.18
	100	0.28	0.35	0.38	0.29	0.32	4.85	4.22	3.41	2.75
	200	0.10	0.08	0.08	0.23	0.17	3.57	2.93	2.44	2.01

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