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

June 13, 2009

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 the $I(1)$ 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 for most helpful comments on a previous version of this paper. 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, assuming that the regressors and the errors are independently distributed.

In the case of panels with a large cross section dimension, SURE approach is not practical and 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. 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 Korajzcyk (1986) and Connor and Korajzcyk (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 distributional results for $I(1)$ processes are radically different to those for $I(0)$ processes and involve functionals of Brownian motion whose use requires separate tabulations of critical values.

It is very important to appreciate that our primary focus is on estimating the coefficients of the panel regression model. We do not wish to investigate the integration properties of the unobserved factors or the cointegration properties of the relationship between the dependent and explanatory variables. 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.

Notations: K stands for a finite positive constant, $\|\mathbf{A}\| = [Tr(\mathbf{AA}')]^{1/2}$ is the Frobenius norm of the $m \times n$ matrix $\mathbf{A}$, and $\mathbf{A}^+$ denotes the Moore-Penrose inverse of $\mathbf{A}$. $rk(\boldsymbol{A})$ denotes the rank of $\boldsymbol{A}$. $\sup_i W_i$ is the supremum of W_i over i. $a_n = O(b_n)$ states 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 $\mathbf{y}_n$, $\overset{q.m.}{\rightarrow}$ denotes convergence in quadratic mean (or mean square error), $\overset{p}{\rightarrow}$ convergence in probability, $\overset{d}{\rightarrow}$ convergence in distribution, and $\overset{d}{\sim}$ 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 $(N, T) \overset{j}{\rightarrow} \infty$. 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 $i = 1, 2, ..., N$; $t = 1, 2, ..., T$, and suppose that it is generated according to the following linear heterogeneous panel data model

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

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

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

where $\boldsymbol{A}_i$ and $\boldsymbol{\Gamma}_i$ are $n \times k$ and $m \times k$ factor loading matrices with fixed and bounded components, $\boldsymbol{v}_{it} = (v_{i1t}, ..., v_{ikt})'$ are the specific components of $\boldsymbol{x}_{it}$ distributed independently of the common effects and across i, but assumed to follow general covariance stationary processes.

Combining (1) and (2) we now have

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

where

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

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

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

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

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

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

where

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

and

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

We distinguish between two important cases: when the rank condition

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

holds, and when it does not. Under the former, the analysis simplifies considerably since it is possible to proxy the unobserved factors by linear combinations of cross section averages, $\bar{\boldsymbol{z}}_t$ and the observed common components, $\boldsymbol{d}_t$. But if the rank condition is not satisfied this is not possible, although as we shall see it is still possible to consistently estimate the mean of the regression coefficients, $\boldsymbol{\beta}$, by the CCE procedure. As the rank condition is very important for the analysis of this paper we will usually refer to (9) as simply the rank condition, when no confusion is likely to arise.

In the case where the rank condition is met we have

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

But since

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

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

and

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

where

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

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

$$\boldsymbol{f}_t - \left(\boldsymbol{C}\boldsymbol{C}'\right)^{-1} \boldsymbol{C} \left(\bar{\boldsymbol{z}}_t - \bar{\boldsymbol{B}}'\boldsymbol{d}_t\right) \xrightarrow{p} \mathbf{0}, \text{ as } N \to \infty.$$

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

When the rank condition is not satisfied it will not be possible to consistently estimate the individual slope coefficients, $\boldsymbol{\beta}_i$ by the CCE procedure. But consistent estimates of the mean of the slope coefficients, $\boldsymbol{\beta}$, and their asymptotic distribution can be obtained if it is further assumed that the factor loadings are distributed independently of the factors and the individual-specific error processes.

2.1 The CCE Estimators

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

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

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

where

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

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

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

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

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

and

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

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

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

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

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

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

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

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

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

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

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

$$vec(\boldsymbol{B}_i) = vec(\boldsymbol{B}) + \boldsymbol{\eta}_{B,i}, \;\; \boldsymbol{\eta}_{B,i} \backsim IID\ (\mathbf{0}, \boldsymbol{\Omega}_{B\boldsymbol{\eta}}),\ \text{for } i = 1, 2, ..., N, \tag{21}$$

and

$$vec(\boldsymbol{C}_i) = vec(\boldsymbol{C}) + \boldsymbol{\eta}_{C,i}, \;\; \boldsymbol{\eta}_{C,i} \backsim IID\ (\mathbf{0}, \boldsymbol{\Omega}_{C\boldsymbol{\eta}}),\ \text{for } i = 1, 2, ..., N, \tag{22}$$

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

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

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \varkappa_i, \;\; \varkappa_i \backsim IID\ (\mathbf{0}, \boldsymbol{\Omega}_{\varkappa}),\ \text{for } i = 1, 2, ..., N, \tag{23}$$

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

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

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

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

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

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

$$\boldsymbol{\gamma}_i = \boldsymbol{\gamma} + \boldsymbol{\eta}_i, \;\; \boldsymbol{\eta}_i \backsim IID\ (\mathbf{0}, \boldsymbol{\Omega}_{\boldsymbol{\eta}}), \text{ for } i = 1, 2, ..., N, \tag{24}$$

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

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

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

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

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

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

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

and

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

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

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

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

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

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

and

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

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

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

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

and

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

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

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

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

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

Using this, we can formally show that

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

Hence

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

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

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

which can be used here as well. The following theorem summarises the results for the mean group estimator. The result for the case where the rank condition holds is proven in Appendix B, whereas the proofs for the case where the rank condition does not hold is given in Appendix C.

Theorem 1 *Consider the panel data model (1) and (2). Distinguish between the case where the rank condition, (9), holds and when it does not. If the rank condition is met suppose that Assumptions 1-6 hold. If the rank condition does not hold suppose that Assumptions 7(ii) and 7(iii) hold as well. Then, for the Common Correlated Effects Mean Group estimator,* $\hat{\boldsymbol{b}}_{MG}$*, defined by (14), we have, as* $(N,T)\xrightarrow{j}\infty$*, that*

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

where $\boldsymbol{\Sigma}_{MG}$ *is given by* $\boldsymbol{\Omega}_{\varkappa}$ *when the rank condition holds and by (A98) when it does not. In both cases, the variance matrix can be consistently estimated by (38).*

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. But in the case where the rank condition is satisfied the technical Assumption 7 will not be needed, and Assumption 3 can be relaxed. Namely the factor loadings, $\boldsymbol{\gamma}_i$, need not follow the random coefficient model. It would be sufficient that they are bounded.

The following Theorem summarizes the results for the second pooled estimator, $\hat{\boldsymbol{b}}_P$. The proofs are provided in Appendix B when the rank condition is met, and in Appendix C when it is not.

Theorem 2 *Consider the panel data model (1) and (2), and suppose that Assumptions 1-6 hold. If the rank condition, (9), does not hold further suppose that Assumption 7(i) holds. Then, for the Common Correlated Effects Pooled estimator,* $\hat{\boldsymbol{b}}_P$*, defined by (20), as* $(N,T)\xrightarrow{j}\infty$*, we have that*

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

where $\boldsymbol{\Sigma}_P^*$ is given by (A83), if the rank condition holds and by (A110) otherwise. In either case the variance matrix can be estimated consistently by

$$\hat{\boldsymbol{\Sigma}}_P^* = \hat{\boldsymbol{\Psi}}^{*-1} \hat{\boldsymbol{R}}^* \hat{\boldsymbol{\Psi}}^{*-1}, \tag{39}$$

where

$$\hat{\boldsymbol{\Psi}}^* = N^{-1} \sum_{i=1}^{N} \frac{\boldsymbol{X}_i' \bar{\boldsymbol{M}} \boldsymbol{X}_i}{T}, \tag{40}$$

$$\hat{\boldsymbol{R}}^* = \frac{1}{(N-1)} \sum_{i=1}^{N} \left(\frac{\boldsymbol{X}_i' \bar{\boldsymbol{M}} \boldsymbol{X}_i}{T} \right) \left(\hat{\boldsymbol{b}}_i - \hat{\boldsymbol{b}}_{MG} \right) \left(\hat{\boldsymbol{b}}_i - \hat{\boldsymbol{b}}_{MG} \right)' \left(\frac{\boldsymbol{X}_i' \bar{\boldsymbol{M}} \boldsymbol{X}_i}{T} \right). \tag{41}$$

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

3.2 Estimation of Individual Slope Coefficients

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

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

Then, we have the following result.

Theorem 3 *Consider the panel data model (1) and (2) and suppose that Assumptions 1, 2(i) and 3-8 hold. Let* $\sqrt{T}/N \to 0$, *as* $(N,T) \xrightarrow{j} \infty$, *and assume that the rank condition (9) be satisfied. As* $(N,T) \xrightarrow{j} \infty$ *(in no particular order),* $\hat{\boldsymbol{b}}_i$, *defined by (15), is a consistent estimator of* $\boldsymbol{\beta}_i$. *Further*

$$\sqrt{T}\left(\hat{\boldsymbol{b}}_i - \boldsymbol{\beta}_i\right) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}_{b_i}). \tag{42}$$

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

$$\hat{\boldsymbol{\Sigma}}_{b_i} = \mathring{\sigma}_i^2 \left(\frac{\boldsymbol{X}_i' \bar{\boldsymbol{M}} \boldsymbol{X}_i}{T}\right)^{-1}, \tag{43}$$

where

$$\mathring{\sigma}_i^2 = \frac{\left(\boldsymbol{y}_i - \boldsymbol{X}_i \hat{\boldsymbol{b}}_i\right)' \bar{\boldsymbol{M}} \left(\boldsymbol{y}_i - \boldsymbol{X}_i \hat{\boldsymbol{b}}_i\right)}{T - (n + 2k + 1)}. \tag{44}$$

Remark 4 *Parts of the above result hold under weaker versions of Assumption 8. In particular we note that the central limit theorem in (A115) holds if Assumption 2(ii) holds. However, in this case the asymptotic variance has a different form as autocovariances of* $\varepsilon_{it}\boldsymbol{v}_{it}$ *enter the asymptotic variance expression. If, then, a consistent estimate of the asymptotic variance is required a Newey and West (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* $(\varepsilon_{it}, \boldsymbol{v}_{it}')'$ *is a strongly mixing process for this consistency to hold.*

Remark 5 *It is worth noting that despite the fact that under our Assumptions* $\boldsymbol{f}_t$, y_{it} *and* $\boldsymbol{x}_{it}$ *are* $I(1)$ *and cointegrated, in the results of Theorem 1 the rate of convergence of* $\hat{\boldsymbol{b}}_i$ *to* $\boldsymbol{\beta}_i$ *as* $(N,T) \xrightarrow{j} \infty$ *is* $\sqrt{T}$ *and not* T. *It is helpful to develop some intuition behind this result. Since for* N *sufficiently large* $\boldsymbol{f}_t$ *can be well approximated by the cross section averages, for pedagogic purposes we might as well consider the case where* $\boldsymbol{f}_t$ *is observed. Without loss of generality we also abstract from* $\boldsymbol{d}_t$, *and substitute (2) in (1) to obtain*

$$y_{it} = \boldsymbol{\beta}_i' \left(\boldsymbol{\Gamma}_i' \boldsymbol{f}_t + \boldsymbol{v}_{it}\right) + \boldsymbol{\gamma}_i' \boldsymbol{f}_t + \varepsilon_{it} = \boldsymbol{\vartheta}_i' \boldsymbol{f}_t + \zeta_{it}, \tag{45}$$

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

Remark 6 *An issue related to the above remark concerns the probability limit of the OLS estimator of the coefficients of* $\boldsymbol{x}_{it}$ *in a regression of* y_{it} *on* $\boldsymbol{x}_{it}$ *alone. In general, such a regression will be subject to the omitted variable problem and hence misspecified. Also the asymptotic properties of such OLS estimators can not be derived without further assumptions. However, there is a special case which illustrates the utility of our method. Abstracting from* $\boldsymbol{d}_t$*, assuming that* $k = m$ *and that* $\boldsymbol{\Gamma}_i$ *is invertible, and similarly to (45) write the model for* y_{it} *as*

$$y_{it} = \boldsymbol{\beta}_i'\boldsymbol{x}_{it} + \boldsymbol{\gamma}_i'\boldsymbol{\Gamma}_i'^{-1}(\boldsymbol{x}_{it} - \boldsymbol{v}_{it}) + \varepsilon_{it} = \boldsymbol{\varrho}_i'\boldsymbol{x}_{it} + \varsigma_{it} \tag{46}$$

where $\boldsymbol{\varrho}_i' = \boldsymbol{\beta}_i' + \boldsymbol{\gamma}_i'\boldsymbol{\Gamma}_i'^{-1}$ *and* $\varsigma_{it} = \varepsilon_{it} - \boldsymbol{\gamma}_i'\boldsymbol{\Gamma}_i'^{-1}\boldsymbol{v}_{it}$*. Note that* ς_{it} *is, by construction, correlated with* $\boldsymbol{v}_{it}$*. The question is whether estimating a regression of the form (46) provides a consistent estimate of* $\boldsymbol{\varrho}_i$*. For stationary processes this would not be case due to the correlation between* ς_{it} *and* $\boldsymbol{v}_{it}$*. However, in the case of 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 (46) can only lead to a consistent estimator of* $\boldsymbol{\varrho}_i$ *and not of* $\boldsymbol{\beta}_i$*. To consistently estimate the latter we need to augment the regressions of* y_{it} *on* $\boldsymbol{x}_{it}$ *with their cross-section averages.*

Remark 7 *When the rank condition, (9), is not satisfied consistent estimation of the individual slope coefficients by the CCE procedure is not possible.*

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. We also consider the two alternative principal component augmentation approaches discussed in Kapetanios and Pesaran (2007). The first PC approach applies the Bai and Ng (2002) procedure to $\boldsymbol{z}_{it} = (y_{it}, \boldsymbol{x}_{it}')'$ to obtain consistent estimates of the unobserved factors, and then uses the estimated factors to augment the regression (1), and thus produces consistent estimates of $\boldsymbol{\beta}$. We consider both pooled and mean group versions of this estimator which we refer to as $PC1POOL$ and $PC1MG$. The second PC approach begins with extracting the principal component estimates of the unobserved factors from y_{it} and $\boldsymbol{x}_{it}$ separately. In the second step y_{it} and $\boldsymbol{x}_{it}$ are regressed on their respective factor estimates, and in the third step the residuals from these regressions are used to compute the standard pooled and mean group estimators, with no cross-sectional dependence adjustments. We refer to the estimators based on this approach as $PC2POOL$ and $PC2MG$, respectively.

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

$$y_{it} = \alpha_{i1} d_{1t} + \beta_{i1} x_{1it} + \beta_{i2} x_{2it} + \gamma_{i1} f_{1t} + \gamma_{i2} f_{2t} + \varepsilon_{it}, \tag{47}$$

and

$$x_{ijt} = a_{ij1} d_{1t} + a_{ij2} d_{2t} + \gamma_{ij1} f_{1t} + \gamma_{ij3} f_{3t} + \mathrm{v}_{ijt},\ j = 1, 2, \tag{48}$$

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

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

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

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

$$\begin{aligned} &\mathrm{v}_{ijt} = \rho_{vij} \mathrm{v}_{ijt-1} + \varkappa_{ijt},\ t = -49, ...1, ..., T, \\ &\varkappa_{ijt} \sim IIDN\left(0, 1 - \rho_{vij}^2\right),\ \mathrm{v}_{ji,-50} = 0, \end{aligned}$$

and

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

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

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

The first 50 observations are discarded.

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

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

and

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

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

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

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

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

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

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

and

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

Under set $\mathcal{B}$

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

so that

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

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

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

The two versions of experiment 1 will be denoted by $1\mathcal{A}$ and $1\mathcal{B}$, and those of experiment 2 by $2\mathcal{A}$ and $2\mathcal{B}$. For this Monte Carlo study we also computed the CCEMG and the CCEP estimators as well as the associated "infeasible" estimators (MG and Pooled) that include f_{1t} and f_{2t} in the regressions of y_{it} on $(d_{1t}, \boldsymbol{x}_{it})$, and the "naive" estimators that exclude these factors. The naive estimators 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.

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, $\boldsymbol{\beta}_i$, and/or error variances, σ_i^2, differ across i. As a result the CCE or PC augmented estimators may in fact dominate the infeasible estimator in terms of 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 PC type estimators is the fact that the computation of the PC augmented 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 prior* estimate for m.

Each experiment was replicated 2000 times for the (N, T) pairs with $N, T = 20, 30, 50, 100, 200$. In what follows we shall focus on β_1 (the cross section mean of β_{i1}). Results for β_2 are very similar and will not be reported. Finally, for completeness we state below the exact formulae for the variance estimators used for the different estimators. The non-parametric variance estimators of the mean group estimators, $\tilde{\boldsymbol{b}}_{MG} = N^{-1}\sum_{i=1}^{N}\tilde{\boldsymbol{b}}_i$, are computed as

$$\widehat{Var}(\tilde{\boldsymbol{b}}_{MG}) = \frac{1}{N(N-1)}\sum_{i=1}^{N}\left(\tilde{\boldsymbol{b}}_i - \tilde{\boldsymbol{b}}_{MG}\right)\left(\tilde{\boldsymbol{b}}_i - \tilde{\boldsymbol{b}}_{MG}\right)', \tag{50}$$

where

$$\tilde{\boldsymbol{b}}_i = \left(\boldsymbol{X}_i'\tilde{\boldsymbol{M}}_x\boldsymbol{X}_i\right)^{-1}\boldsymbol{X}_i'\tilde{\boldsymbol{M}}_x\tilde{\boldsymbol{M}}_y\boldsymbol{y}_i,$$

$$\tilde{\boldsymbol{M}}_x = \boldsymbol{I}_T - \tilde{\boldsymbol{H}}_x\left(\tilde{\boldsymbol{H}}_x'\tilde{\boldsymbol{H}}_x\right)^{-1}\tilde{\boldsymbol{H}}_x',\ \tilde{\boldsymbol{M}}_y = \boldsymbol{I}_T - \tilde{\boldsymbol{H}}_y\left(\tilde{\boldsymbol{H}}_y'\tilde{\boldsymbol{H}}_y\right)^{-1}\tilde{\boldsymbol{H}}_y'.$$

For the CCEMG estimator, $\tilde{\boldsymbol{H}}_x = \tilde{\boldsymbol{H}}_y = \bar{\boldsymbol{H}} = (\boldsymbol{D}, \bar{\boldsymbol{Z}})$, so that $\tilde{\boldsymbol{b}}_i = \hat{\boldsymbol{b}}_i$, which is defined by (15); for the PC1MG estimator, $\tilde{\boldsymbol{H}}_x = \tilde{\boldsymbol{H}}_y = \hat{\boldsymbol{F}}_z$, where $\hat{\boldsymbol{F}}_z$ is a $T \times (n+m)$ matrix of extracted factors from $\boldsymbol{Z}_i = (\boldsymbol{y}_i, \boldsymbol{X}_i)$ for all i, together with observed common factors; for the PC2MG estimator $\tilde{\boldsymbol{H}}_x = \hat{\boldsymbol{F}}_x$ and $\tilde{\boldsymbol{H}}_y = \hat{\boldsymbol{F}}_y$, where $\hat{\boldsymbol{F}}_x$ and $\hat{\boldsymbol{F}}_y$ are $T \times (n_x + m_x)$ and $T \times (n_y + m_y)$ matrices of extracted factors from $\boldsymbol{X}_i$ and $\boldsymbol{y}_i$ respectively for all i, together with the observed common factors with n_x and n_y being the number of observed common factors in $\boldsymbol{X}_i$ and $\boldsymbol{y}_i$ respectively, and m_x and m_y defined similarly; for the infeasible mean group estimator, $\tilde{\boldsymbol{H}}_x = \tilde{\boldsymbol{H}}_y = \boldsymbol{F}_y$, which is a $T \times m_y$ matrix of unobserved factors in $\boldsymbol{y}_i$; for the naive mean group estimator, $\tilde{\boldsymbol{H}}_x = \tilde{\boldsymbol{H}}_y = \boldsymbol{D}$. Next, the non-parametric variance of the pooled estimator,

$\tilde{\boldsymbol{b}}_P$, is computed as

$$\widehat{Var}(\tilde{\boldsymbol{b}}_P) = N^{-1}\tilde{\boldsymbol{\Psi}}^{-1}\tilde{\boldsymbol{R}}\tilde{\boldsymbol{\Psi}}^{-1}, \tag{51}$$

where

$$\tilde{\boldsymbol{b}}_P = \left(\sum_{i=1}^{N}\boldsymbol{X}_i'\tilde{\boldsymbol{M}}_x\boldsymbol{X}_i\right)^{-1}\sum_{i=1}^{N}\boldsymbol{X}_i'\tilde{\boldsymbol{M}}_x\tilde{\boldsymbol{M}}_y\boldsymbol{y}_i,$$

$$\tilde{\boldsymbol{\Psi}} = N^{-1}\sum_{i=1}^{N}\left(\frac{\boldsymbol{X}_i'\tilde{\boldsymbol{M}}_x\boldsymbol{X}_i}{T}\right),$$

$$\tilde{\boldsymbol{R}} = \frac{1}{(N-1)}\sum_{i=1}^{N}\left(\frac{\boldsymbol{X}_i'\tilde{\boldsymbol{M}}_x\boldsymbol{X}_i}{T}\right)\left(\tilde{\boldsymbol{b}}_i-\tilde{\boldsymbol{b}}_{MG}\right)\left(\tilde{\boldsymbol{b}}_i-\tilde{\boldsymbol{b}}_{MG}\right)'\left(\frac{\boldsymbol{X}_i'\tilde{\boldsymbol{M}}_x\boldsymbol{X}_i}{T}\right).$$

In order to show the effect of another type of violation of the rank condition, consider the following data generating process (DGP):

$$y_{it} = \alpha_{i1}d_{1t} + \beta_{i1}x_{1it} + \beta_{i2}x_{2it} + \gamma_{i1}f_{1t} + \gamma_{i2}f_{2t} + \gamma_{i4}f_{4t} + \varepsilon_{it}, \tag{52}$$

and

$$x_{ijt} = a_{ij1}d_{1t} + a_{ij2}d_{2t} + \gamma_{ij1}f_{1t} + \gamma_{ij3}f_{3t} + \text{v}_{ijt}, \tag{53}$$

$j = 1, 2$, for $i = 1, 2, ..., N$, and $t = 1, 2, ..., T$. Sets $n = k = 2$ with $\boldsymbol{\alpha}_i' = (\alpha_{i1}, 0)$, $\boldsymbol{\beta}_i' = (\beta_{i1}, \beta_{i2})$, and $\boldsymbol{\gamma}_i' = (\gamma_{i1}, \gamma_{i2}, 0, \gamma_{i4})$ with $\gamma_{i\ell} \sim IIDN(1, 0.2)$ for $\ell = 1, 2$ and $\gamma_{i4} \sim IIDN(0.5, 0.2)$,

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

$$\boldsymbol{\Gamma}_i' \sim IID\begin{pmatrix} N(0.5, 0.5) & 0 & N(0, 0.5) & 0 \\ N(0, 0.5) & 0 & N(0.5, 0.5) & 0 \end{pmatrix}.$$

Observe that

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

whose rank is $k + 1 = 3$, which is less than the number of unobserved factors, $m = 4$. The rest of variables are generated similarly to the first DGP. We considered both cases of homogeneous and heterogeneous slopes.

Using this DGP, we implement two different set of experiments. In the first set, the number of factors are treated as known, and factors are extracted as before. In the second set, the number of factors are treated as unknown, and the number of factors are estimated, using the information criterion PC_{p2} which is proposed by Bai and Ng (2002, p.201).[3] The information

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

criterion is applied to the first differenced variables, and the maximum number of factors is set to six.

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 asymptotic 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 above, $1\mathcal{A}$, $1\mathcal{B}$, $2\mathcal{A}$ and $2\mathcal{B}$, 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 $t < \lfloor 2T/3 \rfloor$ and $f_{jt} = 1 + \varphi_{jt}$ for $t \geq \lfloor 2T/3 \rfloor$ with $\lfloor A \rfloor$ being the greatest integer less than or equal to A, where $\varphi_{jt} = \varphi_{j,t-1} + \zeta_{jt}$, and $\zeta_{jt} \sim IIDN(0, 1)$, for $j = 1, 2, 3$.

4.1 Results

Results of experiments $1\mathcal{A}$, $2\mathcal{A}$, $1\mathcal{B}$, $2\mathcal{B}$ 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 $1\mathcal{A}$.

As can be seen from Table 1 the naive estimator is substantially biased, performs very poorly and is subject to large size distortions; 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 sightly higher RMSEs than those with full rank. 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, ε_{it}. The robustness of the CCE estimator to the short run dynamics is particularly helpful in practice where typically little is known about such dynamics. In fact a comparison of the results for the CCEP estimator with the infeasible counterpart given in Table 1 shows that the former can even be more efficient (in the RMSE sense). For example the RMSE of the CCEP for $N = T = 50$ is 3.97 whilst the RMSE of the infeasible pooled estimator is 4.31. This might seem counter intuitive at first, but as indicated above the infeasible estimator does not take account of the residual serial correlation of the idiosyncratic errors, but the CCE estimator does allow for such possibilities indirectly through the use of the cross section averages that partly embody the serial correlation properties of $\boldsymbol{f}_t$ and ε_{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 summarized in Tables 1-4 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 in the case of the experiments $1\mathcal{A}$ and $2\mathcal{A}$ (in Tables 1 and 2) where the full rank conditions are met. It is also interesting that in the case of some of the experiments the bias distortions of the tests based on the PC augmented estimators do not improve even for relatively large N and T. An interesting distinction arises when comparing results for experiments $1\mathcal{A}$ and $1\mathcal{B}$. For $1\mathcal{A}$ (heterogeneous slopes and full rank) results are very poor for small values of N and T but improve considerably as N rises and less perceptibly as T rises. For experiment $1\mathcal{B}$ (heterogeneous slopes and rank deficient) results are much better for small values of N and T. Finally, it is worth noting that in both cases the performance of the PC estimators actually get 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.

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. For the sake of brevity we include the results of the CCE type estimators and the principal component estimators with augmentation (PC1 estimators) only, since the other PC estimator performed rather poorly by comparison. In this experiment, PC1 estimates are obtained after extracting factors in two cases: i) m is known, and; ii) m is unknown but the number of factors is estimated. Firstly, in spite of the number of unobserved factors, $m = 4$, exceeding the number of regressors and regressand ($k + 1 = 3$), the root mean square errors (RMSEs) of CCE estimators decrease as N and T get larger, which confirms the consistency of the estimators. Furthermore, when PC1 estimators are obtained assuming m known, the RMSEs of CCE estimators dominates those of PC1 estimators, except N and/or T are large. However, in practice the number of factors is usually not known and need to be estimated first before the PC1 estimator can be computed. In order to take this extra uncertainty into account, we also computed the PC1 estimator with the number of factors 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.[4] In this more realistic circumstances, CCE estimators outperform the PC1 estimators in all experiments, in terms of RMSEs. In addition, the size of the CCE estimators is very close to the nominal 5% level, whilst the size distortion of the PC1 estimators are serious unless both N and T is large (this is not reported for brevity). Table 6 reports the results of experiments based on the same DGP as in Table 5, except that slope coefficients are homogeneous. As to be expected the slope homogeneous results in Table 6 show smaller bias and RMSEs, but otherwise are very similar to those reported in Table 5.

Tables 7-10 provide the results of experiments where the unobserved factors are subject to mean shifts. These results are very similar to those reported in Tables 1-4, which confirm the robustness of the CCE type estimators to such structural breaks. This is consistent with the findings of 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 parametrisation 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

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

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.

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., and S. Ng (2002): "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70, 191–221.

Chamberlain, G., and M. Rothschild (1983): "Arbitrage, Factor Structure and Mean-Variance Analysis in Large Asset Markets," *Econometrica*, 51, 1305–1324.

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. Korajzcyk (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.

Newey, 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., 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.

6 Appendix A

Lemmas

Lemma 1 *Under Assumptions 1-4,*

$$\frac{\bar{\boldsymbol{U}}'\bar{\boldsymbol{U}}}{T}=O_p\left(\frac{1}{N}\right) \tag{A1}$$

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

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

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

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

$$\frac{\boldsymbol{Q}'\boldsymbol{Q}}{T^2}=O_p\left(1\right) \tag{A6}$$

$$\frac{\boldsymbol{Q}'\boldsymbol{X}_i}{T^2}=O_p\left(1\right),\ \textit{uniformly over } i \tag{A7}$$

$$\frac{\boldsymbol{Q}'\boldsymbol{G}}{T^2}=O_p\left(1\right) \tag{A8}$$

$$\frac{\bar{\boldsymbol{H}}'\bar{\boldsymbol{H}}}{T^2}=O_p(1) \tag{A9}$$

$$\frac{\bar{\boldsymbol{H}}'\boldsymbol{G}}{T^2}=O_p(1) \tag{A10}$$

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

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

$$\frac{\bar{\boldsymbol{H}}'\boldsymbol{X}_i}{T^2}=O_p(1),\ \textit{uniformly over } i \tag{A13}$$

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

Proof. To prove (A1) we first show that

$$E\left\|\bar{\mathbf{u}}_t\right\|^2=O\left(\frac{1}{N}\right),\ \text{and } E\left\|\bar{\mathbf{u}}_t\right\|=O\left(\frac{1}{\sqrt{N}}\right), \tag{A15}$$

We recall that

$$\bar{\mathbf{u}}_t=\begin{pmatrix} \bar{\varepsilon}_t+\frac{1}{N}\sum_{i=1}^N\boldsymbol{\beta}_i'\boldsymbol{v}_{it} \\ \bar{\boldsymbol{v}}_t \end{pmatrix}, \tag{A16}$$

where $\bar{\boldsymbol{v}}_t=\frac{1}{N}\sum_{i=1}^N\boldsymbol{\beta}_i'\boldsymbol{v}_{it}$. Then, by the cross-sectional independence of $\boldsymbol{v}_{it}$ and $\boldsymbol{\beta}_i'$ specified in Assumptions 2 and 4, $E\left\|\bar{\boldsymbol{v}}_t\right\|^2=\frac{1}{N^2}\sum_{i=1}^N E\left\|\boldsymbol{\beta}_i'\boldsymbol{v}_{it}\right\|^2$, and again by Assumptions 2 and 4, we have

$$E\left\|\bar{\boldsymbol{v}}_t\right\|^2\leq\frac{K}{N}=O\left(\frac{1}{N}\right). \tag{A17}$$

Similarly,

$$E\left(\bar{\varepsilon}_t^2\right) = O\left(\frac{1}{N}\right). \tag{A18}$$

Next, note that $T^{-1}\bar{\mathbf{U}}'\bar{\mathbf{U}} = T^{-1}\left(\sum_{t=1}^{T}\bar{\mathbf{u}}_t\bar{\mathbf{u}}_t'\right)$, where the cross-product terms in $\bar{\mathbf{u}}_t\bar{\mathbf{u}}_t'$, being functions of covariance stationary processes with finite fourth-order cumulants, are themselves stationary with finite means and variances. Also, $E\left\|T^{-1}\bar{\mathbf{U}}'\bar{\mathbf{U}}\right\| \le T^{-1}\sum_{t=1}^{T}E\left\|\bar{\mathbf{u}}_t\right\|^2$, and by (A15) $E\left\|T^{-1}\bar{\mathbf{U}}'\bar{\mathbf{U}}\right\| = O\left(N^{-1}\right)$, which establishes (A1).

The result for $\mathbf{V}_i'\bar{\boldsymbol{U}}/T$ in (A2) is established in Lemma 2 below. The result for $\boldsymbol{\varepsilon}_i'\bar{\boldsymbol{U}}/T$ in (A2) is established similarly to that for $\mathbf{V}_i'\bar{\boldsymbol{U}}/T$.

To establish (A3), firstly we examine $T^{-1}\left(\boldsymbol{F}'\bar{\boldsymbol{U}}\right)$. Consider the ℓ^{th} row of $T^{-1}\left(\boldsymbol{F}'\bar{\boldsymbol{U}}\right)$ and note that it can be written as $T^{-1}\left(\sum_{t=1}^{T}f_{\ell t}\bar{\boldsymbol{u}}_t'\right)$. Since by assumption $f_{\ell t}$ and $\bar{\mathbf{u}}_t$ are independently distributed processes then

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

where $E\left(\bar{\mathbf{u}}_t\bar{\mathbf{u}}_{t'}'\right) = O\left(N^{-1}\right)$. Hence,

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

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

$$Var\left(\frac{\sum_{t=1}^{T}f_{\ell t}\bar{\mathbf{u}}_t}{T}\right) = O\left(\frac{1}{N}\right), \tag{A19}$$

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

To establish (A4), first note that

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

using (A2) and (A3), since the elements of $\mathbf{\Pi}_i$ are assumed to be bounded uniformly over i.

To establish (A5), recalling that $\boldsymbol{Q} = \boldsymbol{G}\bar{\boldsymbol{P}},$ and using (A3)

$$\frac{\boldsymbol{Q}'\bar{\boldsymbol{U}}}{T} = \bar{\boldsymbol{P}}'\frac{\left(\boldsymbol{D},\boldsymbol{F}\right)'\bar{\boldsymbol{U}}}{T} = O_p\left(\frac{1}{\sqrt{N}}\right)$$

since the elements of $\bar{\boldsymbol{P}}$ are assumed to be bounded.

(A6) is established by

$$\frac{\boldsymbol{Q}'\boldsymbol{Q}}{T^2} = \bar{\boldsymbol{P}}'\frac{\boldsymbol{G}'\boldsymbol{G}}{T^2}\bar{\boldsymbol{P}} = O_p(1),$$

since $\boldsymbol{G}'\boldsymbol{G}/T^2 = O_p(1)$.

To establish (A7), first note that

$$\frac{\boldsymbol{Q}'\boldsymbol{X}_i}{T^2} = \bar{\boldsymbol{P}}'\left(\frac{\boldsymbol{G}'\boldsymbol{G}}{T^2}\right)\mathbf{\Pi}_i + \bar{\boldsymbol{P}}'\frac{\boldsymbol{G}'\mathbf{V}_i}{T^2}. \tag{A20}$$

The first term is $O_p(1)$ uniformly over i, since the elements of $\bar{\boldsymbol{P}}$ and $\mathbf{\Pi}_i$ are assumed to be bounded in probability uniformly over i. For the second term, under Assumptions 1-2, we have that

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

where $g_{\ell t}$ is the ℓ^{th} element of $\boldsymbol{g}_t$ and $\sup_i E\left(\boldsymbol{v}_t \boldsymbol{v}'_{t'}\right) = O\left(1\right)$. Hence,

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

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

$$\sup_i Var\left(\frac{\sum_{t=1}^T g_{\ell t}\boldsymbol{v}'_{it}}{T}\right) = O\left(1\right). \tag{A21}$$

Hence, $\boldsymbol{G}'\boldsymbol{V}_i/T = O_p(1)$ uniformly over i for sufficiently large T. Therefore, as the elements of $\bar{\boldsymbol{P}}$ are assumed to be bounded in probability, the second term is $O_p(1)$ uniformly over i, which establishes (A7). (A8) is straightforwardly proven, using (A6).

To prove (A9), recalling $\bar{\boldsymbol{H}} = \boldsymbol{Q} + \bar{\boldsymbol{U}}^*$, where $\bar{\boldsymbol{U}}^* = (\boldsymbol{0}, \bar{\boldsymbol{U}})$,

$$\frac{\bar{\boldsymbol{H}}'\bar{\boldsymbol{H}}}{T^2} = \frac{\boldsymbol{Q}'\boldsymbol{Q}}{T^2} + \frac{\bar{\boldsymbol{U}}^{*\prime}\bar{\boldsymbol{U}}^*}{T^2} + \frac{\boldsymbol{Q}'\bar{\boldsymbol{U}}^*}{T^2} + \frac{\bar{\boldsymbol{U}}^{*\prime}\boldsymbol{Q}}{T^2} = O_p(1) \tag{A22}$$

by (A1), (A5) and (A6).

To establish (A10),

$$\frac{\bar{\boldsymbol{H}}'\boldsymbol{F}}{T^2} = \bar{\boldsymbol{P}}'\frac{\boldsymbol{G}'\boldsymbol{F}}{T^2} + \frac{\bar{\boldsymbol{U}}^{*\prime}\boldsymbol{F}}{T^2} = O_p(1) \tag{A23}$$

since $\boldsymbol{G}'\boldsymbol{F}/T^2$ is $O_p(1)$.

(A11) is established because

$$\frac{\bar{\boldsymbol{H}}'\boldsymbol{\varepsilon}_i}{T} = \bar{\boldsymbol{P}}'\frac{\boldsymbol{G}'\boldsymbol{\varepsilon}_i}{T} + \frac{\bar{\boldsymbol{U}}^{*\prime}\boldsymbol{\varepsilon}_i}{T} = O_p(1) \text{ uniformly over } i \tag{A24}$$

since $\boldsymbol{G}'\boldsymbol{\varepsilon}_i/T = O_p(1)$ uniformly over i, using the same line of the argument as in the proof of (A7). (A12) can be proven similarly to (A11).

Next,

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

by (A4) and (A7), which establishes (A13). Finally, (A14) follows by the boundedness in probability of $\bar{\boldsymbol{P}}$ and (A3). ■

Lemma 2 *Under assumptions 1-4,*

$$\frac{\boldsymbol{V}'_i\bar{\boldsymbol{U}}}{T} = O_p\left(\frac{1}{\sqrt{TN}}\right) + O_p\left(\frac{1}{N}\right) \text{ uniformly over } i. \tag{A25}$$

Proof. In order to prove (A25) we need to examine more closely Lemma A.2. of Pesaran (2006). So, we have

$$\frac{\boldsymbol{V}'_i\bar{\boldsymbol{U}}}{T} = \left(T^{-1}\boldsymbol{V}'_i\bar{\boldsymbol{\varepsilon}} + (NT)^{-1}\boldsymbol{V}'_i\sum_{j=1}^N \boldsymbol{V}_j\boldsymbol{\beta}_j,\ T^{-1}\boldsymbol{V}'_i\bar{\boldsymbol{V}}\right), \tag{A26}$$

where $\bar{\boldsymbol{\varepsilon}} = N^{-1}\sum_{j=1}^N \boldsymbol{\varepsilon}_j$ and $\bar{\boldsymbol{V}} = N^{-1}\sum_{j=1}^N \boldsymbol{V}_j$. Denote the t^{th} element of $\bar{\boldsymbol{\varepsilon}}$ by $\bar{\varepsilon}_t = N^{-1}\sum_{j=1}^N \varepsilon_{jt}$, consider the first term on the RHS of (A26). Since by assumption, $\boldsymbol{v}_{it}$ and $\bar{\varepsilon}_t$ are independently distributed covariance stationary processes with zero means, then

$$\sup_i Var\left(\frac{\sum_{t=1}^T v_{i\ell t}\bar{\varepsilon}_t}{T}\right) = \sup_i \frac{\sum_{t=1}^T \sum_{t'=1}^T E\left(v_{i\ell t}v_{i\ell t'}\right)E\left(\bar{\varepsilon}_t\bar{\varepsilon}_{t'}\right)}{T^2},$$

where $E\left(\bar{\varepsilon}_t \bar{\varepsilon}_{t'}\right) = O\left(N^{-1}\right)$. Hence,

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

where $\Gamma_{iv\ell}\left(|t-t'|\right)$ is the autocovariance function of the stationary process, $v_{i\ell t}$. But, by Assumption 2,

$$\sup_i \sum_{s=1}^{\infty} \Gamma_{iv\ell}\left(|s|\right) < \infty.$$

Therefore,

$$\sup_i Var\left(\frac{\sum_{t=1}^{T} v_{i\ell t}\bar{\boldsymbol{\varepsilon}}_t}{T}\right) = O\left(\frac{1}{NT}\right), \tag{A27}$$

which establishes that

$$T^{-1}\boldsymbol{V}_i'\,\bar{\boldsymbol{\varepsilon}} = O_p\left(\frac{1}{\sqrt{TN}}\right), \quad \text{uniformly over } i. \tag{A28}$$

To see how (A28) follows from (A27),we note that by the Markov inequality

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

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

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

proving that (A28) follows from (A27).

Consider the second term in (A26) and note that

$$(NT)^{-1}\,\boldsymbol{V}_i'\sum_{j=1}^{N}\boldsymbol{V}_j\boldsymbol{\beta}_j = N^{-1}\left(\frac{\boldsymbol{V}_i'\boldsymbol{V}_i}{T}\right)\boldsymbol{\beta}_i + \left(\frac{\boldsymbol{V}_i'\bar{\boldsymbol{V}}_{-i}^{*}}{T}\right), \tag{A29}$$

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

$$N^{-1}\left(\frac{\boldsymbol{V}_i'\boldsymbol{V}_i}{T}\right)\boldsymbol{\beta}_i = O_p\left(\frac{1}{N}\right), \quad \text{uniformly over } i. \tag{A30}$$

Also, since the elements of $\boldsymbol{V}_i$ and $\bar{\boldsymbol{V}}_{-i}^{*}$ are independently distributed and covariance stationary, following the same line of analysis leading to (A28), we have

$$\frac{\boldsymbol{V}_i'\bar{\boldsymbol{V}}_{-i}^{*}}{T} = O_p\left(\frac{1}{\sqrt{NT}}\right), \quad \text{uniformly over } i. \tag{A31}$$

Using (A30) and (A31) in (A29) now yields

$$(NT)^{-1}\,\boldsymbol{V}_i'\sum_{j=1}^{N}\boldsymbol{V}_j\boldsymbol{\beta}_j = O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{N}\right), \quad \text{uniformly over } i. \tag{A32}$$

Finally, since the last term of (A26) can be written as

$$T^{-1}\boldsymbol{V}_i'\bar{\boldsymbol{V}} = N^{-1}\left(\frac{\boldsymbol{V}_i'\boldsymbol{V}_i}{T}\right) + \frac{\boldsymbol{V}_i'\bar{\boldsymbol{V}}_{-i}}{T},$$

where $\bar{\boldsymbol{V}}_{-i} = N^{-1}\sum_{j=1,j\neq i}^{N}\boldsymbol{V}_j$, it also follows that

$$T^{-1}\boldsymbol{V}_i'\bar{\boldsymbol{V}} = O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{N}\right), \quad \text{uniformly over } i, \tag{A33}$$

which completes the proof of Lemma 2. ■

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

$$\frac{\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{X}_i}{T} - \frac{\boldsymbol{X}_i'\boldsymbol{M}_g\boldsymbol{X}_i}{T} = O_p\left(\frac{1}{\sqrt{N}}\right), \text{ uniformly over } i, \tag{A34}$$

$$\frac{\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{\varepsilon}_i}{T} - \frac{\boldsymbol{X}_i'\boldsymbol{M}_g\boldsymbol{\varepsilon}_i}{T} = O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{N}\right), \text{ uniformly over } i. \tag{A35}$$

Proof. We start by proving (A34). We need to determine the order of probability of $\left\|\frac{\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{X}_i}{T} - \frac{\boldsymbol{X}_i'\boldsymbol{M}_g\boldsymbol{X}_i}{T}\right\|$. But this is equal to

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

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

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

$$\leq \left\|\frac{1}{T}\left(\boldsymbol{X}_i'\bar{\boldsymbol{H}} - \boldsymbol{X}_i'\boldsymbol{Q}\right)\left(\bar{\boldsymbol{H}}'\bar{\boldsymbol{H}}\right)^{-1}\bar{\boldsymbol{H}}'\boldsymbol{X}_i\right\| + \left\|\frac{1}{T}\boldsymbol{X}_i'\boldsymbol{Q}\left(\left(\bar{\boldsymbol{H}}'\bar{\boldsymbol{H}}\right)^{-1} - \left(\boldsymbol{Q}'\boldsymbol{Q}\right)^{-1}\right)\bar{\boldsymbol{H}}'\boldsymbol{X}_i\right\| +$$
$$\left\|\frac{1}{T}\boldsymbol{X}_i'\boldsymbol{Q}\left(\boldsymbol{Q}'\boldsymbol{Q}\right)^{-1}\left(\bar{\boldsymbol{H}}'\boldsymbol{X}_i - \boldsymbol{Q}'\boldsymbol{X}_i\right)\right\|. \tag{A36}$$

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

$$\begin{aligned}\left\|\frac{1}{T}\left(\boldsymbol{X}_i'\bar{\boldsymbol{H}} - \boldsymbol{X}_i'\boldsymbol{Q}\right)\left(\bar{\boldsymbol{H}}'\bar{\boldsymbol{H}}\right)^{-1}\bar{\boldsymbol{H}}'\boldsymbol{X}_i\right\| &= \left\|\frac{1}{T}\left(\boldsymbol{X}_i'\bar{\boldsymbol{H}} - \boldsymbol{X}_i'\boldsymbol{Q}\right)\left(\frac{\bar{\boldsymbol{H}}'\bar{\boldsymbol{H}}}{T^2}\right)^{-1}\frac{\bar{\boldsymbol{H}}'\boldsymbol{X}_i}{T^2}\right\| \\ &\leq \left\|\frac{\boldsymbol{X}_i'\bar{\boldsymbol{U}}^*}{T}\right\|\left\|\left(\frac{\bar{\boldsymbol{H}}'\bar{\boldsymbol{H}}}{T^2}\right)^{-1}\frac{\bar{\boldsymbol{H}}'\boldsymbol{X}_i}{T^2}\right\| \\ &= O_p\left(\frac{1}{\sqrt{N}}\right), \quad \text{uniformly over } i,\end{aligned} \tag{A37}$$

by (A4), (A9) and (A13). Next, we have

$$\begin{aligned}
&\left\|\frac{1}{T}\boldsymbol{X}_i'\boldsymbol{Q}\left(\left(\bar{\boldsymbol{H}}'\bar{\boldsymbol{H}}\right)^{-1}-\left(\boldsymbol{Q}'\boldsymbol{Q}\right)^{-1}\right)\bar{\boldsymbol{H}}'\boldsymbol{X}_i\right\| \\
&=\left\|\frac{\boldsymbol{X}_i'\boldsymbol{Q}}{T^2}\left(\frac{\bar{\boldsymbol{H}}'\bar{\boldsymbol{H}}}{T^2}\right)^{-1}\frac{1}{T}\left(\left(\boldsymbol{Q}'\boldsymbol{Q}\right)-\left(\bar{\boldsymbol{H}}'\bar{\boldsymbol{H}}\right)\right)\left(\frac{\boldsymbol{Q}'\boldsymbol{Q}}{T^2}\right)^{-1}\frac{\bar{\boldsymbol{H}}'\boldsymbol{X}_i}{T^2}\right\| \\
&\leq\left\|\frac{\bar{\boldsymbol{H}}'\bar{\boldsymbol{H}}}{T}-\frac{\boldsymbol{Q}'\boldsymbol{Q}}{T}\right\|\left\|\frac{\boldsymbol{X}_i'\boldsymbol{Q}}{T^2}\left(\frac{\bar{\boldsymbol{H}}'\bar{\boldsymbol{H}}}{T^2}\right)^{-1}\right\|\left\|\left(\frac{\boldsymbol{Q}'\boldsymbol{Q}}{T^2}\right)^{-1}\frac{\bar{\boldsymbol{H}}'\boldsymbol{X}_i}{T^2}\right\| \\
&=\left\|\frac{\bar{\boldsymbol{U}}^{*\prime}\bar{\boldsymbol{U}}^*}{T}+\frac{\boldsymbol{Q}'\bar{\boldsymbol{U}}^*}{T}+\frac{\bar{\boldsymbol{U}}^{*\prime}\boldsymbol{Q}}{T}\right\|\left\|\frac{\boldsymbol{X}_i'\boldsymbol{Q}}{T^2}\left(\frac{\bar{\boldsymbol{H}}'\bar{\boldsymbol{H}}}{T^2}\right)^{-1}\right\|\left\|\left(\frac{\boldsymbol{Q}'\boldsymbol{Q}}{T^2}\right)^{-1}\frac{\bar{\boldsymbol{H}}'\boldsymbol{X}_i}{T^2}\right\| \\
&=O_p\left(\frac{1}{\sqrt{N}}\right),\quad \text{uniformly over } i,
\end{aligned}\tag{A38}$$

by (A1), (A5), (A7), (A9), (A6) and (A13). Finally,

$$\begin{aligned}
\left\|\frac{1}{T}\boldsymbol{X}_i'\boldsymbol{Q}\left(\boldsymbol{Q}'\boldsymbol{Q}\right)^{-1}\left(\bar{\boldsymbol{H}}'\boldsymbol{X}_i-\boldsymbol{Q}'\boldsymbol{X}_i\right)\right\| &\leq\left\|\frac{\boldsymbol{X}_i'\boldsymbol{Q}}{T^2}\left(\frac{\boldsymbol{Q}'\boldsymbol{Q}}{T^2}\right)^{-1}\right\|\left\|\frac{\boldsymbol{X}_i'\bar{\boldsymbol{H}}}{T}-\frac{\boldsymbol{X}_i'\boldsymbol{Q}}{T}\right\| \\
&=\left\|\frac{\boldsymbol{X}_i'\boldsymbol{Q}}{T^2}\left(\frac{\boldsymbol{Q}'\boldsymbol{Q}}{T^2}\right)^{-1}\right\|\left\|\frac{\boldsymbol{X}_i'\bar{\boldsymbol{U}}^*}{T}\right\|;\ \text{since } \bar{\boldsymbol{H}}=\boldsymbol{Q}+\bar{\boldsymbol{U}}^* \\
&=O_p\left(\frac{1}{\sqrt{N}}\right)\ \text{uniformly over } i,
\end{aligned}\tag{A39}$$

by (A7), (A6), and (A4). Noting that $\boldsymbol{M}_g=\boldsymbol{M}_q$ when the rank condition is satisfied, substituting (A37), (A38) and (A39) into (A36), we have

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

as required.

Next, we consider (A35). In particular, by a similar analysis to that for (A34), we have

$$\begin{aligned}
\left\|\frac{\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{\varepsilon}_i}{T}-\frac{\boldsymbol{X}_i'\boldsymbol{M}_q\boldsymbol{\varepsilon}_i}{T}\right\| &=\left\|\frac{\boldsymbol{X}_i'\bar{\boldsymbol{H}}\left(\bar{\boldsymbol{H}}'\bar{\boldsymbol{H}}\right)^{-1}\bar{\boldsymbol{H}}'\boldsymbol{\varepsilon}_i}{T}-\frac{\boldsymbol{X}_i'\boldsymbol{Q}\left(\boldsymbol{Q}'\boldsymbol{Q}\right)^{-1}\boldsymbol{Q}'\boldsymbol{\varepsilon}_i}{T}\right\| \\
&\leq\left\|\frac{1}{T}\left(\boldsymbol{X}_i'\bar{\boldsymbol{H}}-\boldsymbol{X}_i'\boldsymbol{Q}\right)\left(\bar{\boldsymbol{H}}'\bar{\boldsymbol{H}}\right)^{-1}\bar{\boldsymbol{H}}'\boldsymbol{\varepsilon}_i\right\|+\left\|\frac{1}{T}\boldsymbol{X}_i'\boldsymbol{Q}\left(\left(\bar{\boldsymbol{H}}'\bar{\boldsymbol{H}}\right)^{-1}-\left(\boldsymbol{Q}'\boldsymbol{Q}\right)^{-1}\right)\bar{\boldsymbol{H}}'\boldsymbol{\varepsilon}_i\right\|+ \\
&\quad\left\|\frac{1}{T}\boldsymbol{X}_i'\boldsymbol{Q}\left(\boldsymbol{Q}'\boldsymbol{Q}\right)^{-1}\left(\bar{\boldsymbol{H}}'\boldsymbol{\varepsilon}_i-\boldsymbol{Q}'\boldsymbol{\varepsilon}_i\right)\right\|.
\end{aligned}\tag{A40}$$

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

$$\begin{aligned}
\left\|\frac{1}{T}\left(\boldsymbol{X}_i'\bar{\boldsymbol{H}}-\boldsymbol{X}_i'\boldsymbol{Q}\right)\left(\bar{\boldsymbol{H}}'\bar{\boldsymbol{H}}\right)^{-1}\bar{\boldsymbol{H}}'\boldsymbol{\varepsilon}_i\right\| &=\left\|\frac{1}{T^2}\left(\boldsymbol{X}_i'\bar{\boldsymbol{H}}-\boldsymbol{X}_i'\boldsymbol{Q}\right)\left(\frac{\bar{\boldsymbol{H}}'\bar{\boldsymbol{H}}}{T^2}\right)^{-1}\frac{\bar{\boldsymbol{H}}'\boldsymbol{\varepsilon}_i}{T}\right\| \\
&\leq\frac{1}{T}\left\|\frac{\boldsymbol{X}_i'\bar{\boldsymbol{U}}^*}{T}\right\|\left\|\left(\frac{\bar{\boldsymbol{H}}'\bar{\boldsymbol{H}}}{T^2}\right)^{-1}\frac{\bar{\boldsymbol{H}}'\boldsymbol{\varepsilon}_i}{T}\right\| \\
&=O_p\left(\frac{1}{\sqrt{N}T}\right),\ \text{uniformly over } i,
\end{aligned}\tag{A41}$$

by (A4), (A9) and (A11). Next, we have

$$
\begin{aligned}
&\left\| \frac{1}{T} \boldsymbol{X}_i' \boldsymbol{Q} \left(\left(\bar{\boldsymbol{H}}' \bar{\boldsymbol{H}} \right)^{-1} - \left(\boldsymbol{Q}' \boldsymbol{Q} \right)^{-1} \right) \bar{\boldsymbol{H}}' \boldsymbol{\varepsilon}_i \right\| \\
&= \left\| \frac{\boldsymbol{X}_i' \boldsymbol{Q}}{T^2} \left(\frac{\bar{\boldsymbol{H}}' \bar{\boldsymbol{H}}}{T^2} \right)^{-1} \frac{1}{T^2} \left(\left(\boldsymbol{Q}' \boldsymbol{Q} \right) - \left(\bar{\boldsymbol{H}}' \bar{\boldsymbol{H}} \right) \right) \left(\frac{\boldsymbol{Q}' \boldsymbol{Q}}{T^2} \right)^{-1} \frac{\bar{\boldsymbol{H}}' \boldsymbol{\varepsilon}_i}{T} \right\| \\
&\leq \frac{1}{T} \left\| - \frac{\bar{\boldsymbol{U}}^{*\prime} \bar{\boldsymbol{U}}^{*}}{T} - \frac{\bar{\boldsymbol{U}}^{*\prime} \boldsymbol{Q}}{T} - \frac{\boldsymbol{Q}' \bar{\boldsymbol{U}}^{*}}{T} \right\| \left\| \frac{\boldsymbol{X}_i' \boldsymbol{Q}}{T^2} \left(\frac{\bar{\boldsymbol{H}}' \bar{\boldsymbol{H}}}{T^2} \right)^{-1} \right\| \left\| \left(\frac{\boldsymbol{Q}' \boldsymbol{Q}}{T^2} \right)^{-1} \frac{\bar{\boldsymbol{H}}' \boldsymbol{\varepsilon}_i}{T} \right\| \\
&= O_p \left(\frac{1}{\sqrt{N} T} \right), \quad \text{uniformly over } i, \qquad \text{(A42)}
\end{aligned}
$$

by (A1), (A5), (A7), (A9), (A6) and (A11). Finally,

$$
\begin{aligned}
&\left\| \frac{1}{T} \boldsymbol{X}_i' \boldsymbol{Q} \left(\boldsymbol{Q}' \boldsymbol{Q} \right)^{-1} \left(\bar{\boldsymbol{H}}' \boldsymbol{\varepsilon}_i - \boldsymbol{Q}' \boldsymbol{\varepsilon}_i \right) \right\| \\
&\leq \left\| \frac{\boldsymbol{X}_i' \boldsymbol{Q}}{T^2} \left(\frac{\boldsymbol{Q}' \boldsymbol{Q}}{T^2} \right)^{-1} \right\| \left\| \frac{\bar{\boldsymbol{U}}^{*\prime} \boldsymbol{\varepsilon}_i}{T} \right\| \\
&= O_p \left(\frac{1}{\sqrt{NT}} \right) + O_p \left(\frac{1}{N} \right), \text{ uniformly over } i, \qquad \text{(A43)}
\end{aligned}
$$

by (A7), (A6), and (A2). Noting that $\boldsymbol{M}_g = \boldsymbol{M}_q$ when the rank condition is satisfied, substituting (A41)-(A43) into (A40) yields

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

which establishes (A35). ■

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

$$
\frac{\boldsymbol{X}_i' \bar{\boldsymbol{M}} \boldsymbol{F}}{T} = O_p \left(\frac{1}{N} \right) + O_p \left(\frac{1}{\sqrt{NT}} \right), \text{ uniformly over } i. \qquad \text{(A44)}
$$

Proof. We start by noting that

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

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

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

or

$$
\bar{\boldsymbol{M}} \boldsymbol{F} \bar{\boldsymbol{C}} = - \bar{\boldsymbol{M}} \bar{\boldsymbol{U}}. \qquad \text{(A45)}
$$

Hence

$$
\left(\bar{\boldsymbol{U}}' \bar{\boldsymbol{M}} \boldsymbol{F} \right) \bar{\boldsymbol{C}} = - \bar{\boldsymbol{U}}' \bar{\boldsymbol{M}} \bar{\boldsymbol{U}}. \qquad \text{(A46)}
$$

Also, from (A45)

$$
\left(\boldsymbol{X}_i' \bar{\boldsymbol{M}} \boldsymbol{F} \right) \bar{\boldsymbol{C}} = - \boldsymbol{X}_i' \bar{\boldsymbol{M}} \bar{\boldsymbol{U}}. \qquad \text{(A47)}
$$

Note, however, that $\boldsymbol{X}_i = \boldsymbol{G}\boldsymbol{\Pi}_i + \boldsymbol{V}_i$, and hence

$$\begin{aligned}
\boldsymbol{X}_i'\bar{\boldsymbol{M}}\bar{\boldsymbol{U}} &= \left(\boldsymbol{\Pi}_i'\boldsymbol{G}' + \boldsymbol{V}_i'\right)\bar{\boldsymbol{M}}\bar{\boldsymbol{U}} \\
&= \boldsymbol{\Pi}_i'\left(\boldsymbol{G}'\bar{\boldsymbol{M}}\bar{\boldsymbol{U}}\right) + \boldsymbol{V}_i'\bar{\boldsymbol{M}}\bar{\boldsymbol{U}} \\
&= \left(\boldsymbol{A}_i', \boldsymbol{\Gamma}_i'\right)\begin{pmatrix} \boldsymbol{D}' \\ \boldsymbol{F}' \end{pmatrix}\bar{\boldsymbol{M}}\bar{\boldsymbol{U}} + \boldsymbol{V}_i'\bar{\boldsymbol{M}}\bar{\boldsymbol{U}} \\
&= \left(\boldsymbol{A}_i', \boldsymbol{\Gamma}_i'\right)\begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{F}'\bar{\boldsymbol{M}}\bar{\boldsymbol{U}} \end{pmatrix} + \boldsymbol{V}_i'\bar{\boldsymbol{M}}\bar{\boldsymbol{U}} \\
&= \boldsymbol{\Gamma}_i'\boldsymbol{F}'\bar{\boldsymbol{M}}\bar{\boldsymbol{U}} + \boldsymbol{V}_i'\bar{\boldsymbol{M}}\bar{\boldsymbol{U}}. \qquad \text{(A48)}
\end{aligned}$$

Substituting (A48) in (A47) yields

$$\left(\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{F}\right)\bar{\boldsymbol{C}} = -\boldsymbol{\Gamma}_i'\boldsymbol{F}'\bar{\boldsymbol{M}}\bar{\boldsymbol{U}} - \boldsymbol{V}_i'\bar{\boldsymbol{M}}\bar{\boldsymbol{U}} \qquad \text{(A49)}$$

or, by the full rank assumption for $\bar{\boldsymbol{C}}$,

$$\left(\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{F}\right) = -\boldsymbol{\Gamma}_i'\,\boldsymbol{F}'\bar{\boldsymbol{M}}\bar{\boldsymbol{U}}\bar{\boldsymbol{C}}'\left(\bar{\boldsymbol{C}}\bar{\boldsymbol{C}}'\right)^{-1} - \boldsymbol{V}_i'\bar{\boldsymbol{M}}\bar{\boldsymbol{U}}\bar{\boldsymbol{C}}'\left(\bar{\boldsymbol{C}}\bar{\boldsymbol{C}}'\right)^{-1}. \qquad \text{(A50)}$$

Also, from (A46)

$$\bar{\boldsymbol{C}}'\left(\boldsymbol{F}'\bar{\boldsymbol{M}}\bar{\boldsymbol{U}}\right) = -\bar{\boldsymbol{U}}'\bar{\boldsymbol{M}}\bar{\boldsymbol{U}} \qquad \text{(A51)}$$

or

$$\left(\boldsymbol{F}'\bar{\boldsymbol{M}}\bar{\boldsymbol{U}}\right) = -\left(\bar{\boldsymbol{C}}\bar{\boldsymbol{C}}'\right)^{-1}\bar{\boldsymbol{C}}\bar{\boldsymbol{U}}'\bar{\boldsymbol{M}}\bar{\boldsymbol{U}}. \qquad \text{(A52)}$$

Then, using this result in (A50) we have

$$\left(\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{F}\right) = \boldsymbol{\Gamma}_i'\left(\bar{\boldsymbol{C}}\bar{\boldsymbol{C}}'\right)^{-1}\bar{\boldsymbol{C}}\left(\bar{\boldsymbol{U}}'\bar{\boldsymbol{M}}\bar{\boldsymbol{U}}\right)\bar{\boldsymbol{C}}'\left(\bar{\boldsymbol{C}}\bar{\boldsymbol{C}}'\right)^{-1} - \left(\boldsymbol{V}_i'\bar{\boldsymbol{M}}\bar{\boldsymbol{U}}\right)\bar{\boldsymbol{C}}'\left(\bar{\boldsymbol{C}}\bar{\boldsymbol{C}}'\right)^{-1}, \qquad \text{(A53)}$$

and hence

$$\left\|\frac{\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{F}}{T}\right\| \leq \left\|\boldsymbol{\Gamma}_i'\right\|\left\|\left(\bar{\boldsymbol{C}}\bar{\boldsymbol{C}}'\right)^{-1}\bar{\boldsymbol{C}}\right\|^2\left\|\frac{\bar{\boldsymbol{U}}'\bar{\boldsymbol{M}}\bar{\boldsymbol{U}}}{T}\right\| + \left\|\frac{\boldsymbol{V}_i'\bar{\boldsymbol{M}}\bar{\boldsymbol{U}}}{T}\right\|\left\|\bar{\boldsymbol{C}}'\left(\bar{\boldsymbol{C}}\bar{\boldsymbol{C}}'\right)^{-1}\right\|. \qquad \text{(A54)}$$

Since the norms of $\left(\bar{\boldsymbol{C}}\bar{\boldsymbol{C}}'\right)^{-1}\bar{\boldsymbol{C}}$ and $\boldsymbol{\Gamma}_i'$ are bounded, we need to establish the probability orders of $\left\|\bar{\boldsymbol{U}}'\bar{\boldsymbol{M}}\bar{\boldsymbol{U}}/T\right\|$ and $\left\|\boldsymbol{V}_i'\bar{\boldsymbol{M}}\bar{\boldsymbol{U}}/T\right\|$. For $\bar{\boldsymbol{U}}'\bar{\boldsymbol{M}}\bar{\boldsymbol{U}}/T$, we have:

$$\frac{\bar{\boldsymbol{U}}'\bar{\boldsymbol{M}}\bar{\boldsymbol{U}}}{T} = \frac{\bar{\boldsymbol{U}}'\bar{\boldsymbol{U}}}{T} - \left(\frac{\bar{\boldsymbol{U}}'\bar{\boldsymbol{H}}}{T^{3/2}}\right)\left(\frac{\bar{\boldsymbol{H}}'\bar{\boldsymbol{H}}}{T^2}\right)^{-1}\left(\frac{\bar{\boldsymbol{H}}'\bar{\boldsymbol{U}}}{T^{3/2}}\right). \qquad \text{(A55)}$$

From (A1), (A9) and (A14), we have that

$$\frac{\bar{\boldsymbol{U}}'\bar{\boldsymbol{U}}}{T} = O_p\left(\frac{1}{N}\right), \; \frac{\bar{\boldsymbol{U}}'\bar{\boldsymbol{H}}}{T^{3/2}} = O_p\left(\frac{1}{\sqrt{NT}}\right), \; \frac{\bar{\boldsymbol{H}}'\bar{\boldsymbol{H}}}{T^2} = O_p(1).$$

Hence,

$$\frac{\bar{\boldsymbol{U}}'\bar{\boldsymbol{M}}\bar{\boldsymbol{U}}}{T} = O_p\left(\frac{1}{N}\right). \qquad \text{(A56)}$$

Similarly

$$\frac{\boldsymbol{V}_i'\bar{\boldsymbol{M}}\bar{\boldsymbol{U}}}{T} = \frac{\boldsymbol{V}_i'\bar{\boldsymbol{U}}}{T} - \left(\frac{\boldsymbol{V}_i'\bar{\boldsymbol{H}}}{T^{3/2}}\right)\left(\frac{\bar{\boldsymbol{H}}'\bar{\boldsymbol{H}}}{T^2}\right)^{-1}\left(\frac{\bar{\boldsymbol{H}}'\bar{\boldsymbol{U}}}{T^{3/2}}\right). \qquad \text{(A57)}$$

By (A2) and (A12)

$$\frac{\boldsymbol{V}_i'\bar{\boldsymbol{U}}}{T} = O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right) \text{ and } \frac{\boldsymbol{V}_i'\bar{\boldsymbol{H}}}{T^{3/2}} = O_p\left(\frac{1}{\sqrt{T}}\right), \text{ uniformly over } i.$$

Hence,

$$\frac{\boldsymbol{V}_i'\bar{\boldsymbol{M}}\bar{\boldsymbol{U}}}{T} = O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right), \text{ uniformly over } i, \tag{A58}$$

and substituting (A56) and (A58) into (A54) establishes the result. ■

Lemma 5 *Under Assumptions 1-4,*

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

Proof. Recall that

$$\boldsymbol{X}_i = \boldsymbol{G}\boldsymbol{\Pi}_i + \boldsymbol{V}_i \tag{A59}$$

where $\boldsymbol{G} = (\boldsymbol{D}, \boldsymbol{F})$ is the $T \times m + n$ matrix of $I(1)$ factors, and $\boldsymbol{V}_i$ is a stationary error matrix. Denote the OLS residuals of the multiple regression (A59) as $\hat{\boldsymbol{V}}_i = \boldsymbol{X}_i - \boldsymbol{G}\hat{\boldsymbol{\Pi}}_i$, where $\hat{\boldsymbol{\Pi}}_i = \left(\boldsymbol{G}'\boldsymbol{G}\right)^{-1}\boldsymbol{G}'\boldsymbol{X}_i$. Observe that $\hat{\boldsymbol{V}}_i = \boldsymbol{M}_g\boldsymbol{X}_i$. Then, we can write

$$\begin{aligned}
\hat{\boldsymbol{V}}_i'\hat{\boldsymbol{V}}_i/T - \boldsymbol{V}_i'\boldsymbol{V}_i/T &= \hat{\boldsymbol{V}}_i'\hat{\boldsymbol{V}}_i/T - \hat{\boldsymbol{V}}_i'\boldsymbol{V}_i/T + \hat{\boldsymbol{V}}_i'\boldsymbol{V}_i/T - \boldsymbol{V}_i'\boldsymbol{V}_i/T \\
&= \hat{\boldsymbol{V}}_i'\left(\hat{\boldsymbol{V}}_i - \boldsymbol{V}_i\right)/T + \left(\hat{\boldsymbol{V}}_i - \boldsymbol{V}_i\right)'\boldsymbol{V}_i/T \\
&= -\boldsymbol{X}_i'\boldsymbol{M}_g\boldsymbol{G}\left(\hat{\boldsymbol{\Pi}}_i - \boldsymbol{\Pi}_i\right)/T - \left(\hat{\boldsymbol{\Pi}}_i - \boldsymbol{\Pi}_i\right)'\left(\boldsymbol{G}'\boldsymbol{V}_i/T\right) \\
&= -\left(\hat{\boldsymbol{\Pi}}_i - \boldsymbol{\Pi}_i\right)'\left(\boldsymbol{G}'\boldsymbol{V}_i/T\right),
\end{aligned}$$

because $\boldsymbol{M}_g\boldsymbol{G} = \boldsymbol{0}$. But, since $\left(\boldsymbol{G}'\boldsymbol{V}_i/T\right) = O_p(1)$ and $\left(\hat{\boldsymbol{\Pi}}_i - \boldsymbol{\Pi}_i\right) = O_p(T^{-1})$, it follows that

$$\hat{\boldsymbol{V}}_i'\hat{\boldsymbol{V}}_i/T - \boldsymbol{V}_i'\boldsymbol{V}_i/T = O_p(T^{-1}).$$

The required result now follows since under Assumption 2, $\boldsymbol{V}_i'\boldsymbol{V}_i/T - \boldsymbol{\Sigma}_{\boldsymbol{v}i} = O_p(T^{-1/2})$, where $\boldsymbol{\Sigma}_{\boldsymbol{v}i}$ is non-singular matrix. ■

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

$$\frac{\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{X}_i}{T} - \frac{\boldsymbol{X}_i'\boldsymbol{M}_q\boldsymbol{X}_i}{T} = O_p\left(\frac{1}{\sqrt{N}}\right), \text{ uniformly over } i, \tag{A60}$$

$$\frac{\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{F}}{T} - \frac{\boldsymbol{X}_i'\boldsymbol{M}_q\boldsymbol{F}}{T} = O_p\left(\frac{1}{\sqrt{N}}\right), \text{ uniformly over } i, \tag{A61}$$

$$\frac{\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{\varepsilon}_i}{T} - \frac{\boldsymbol{X}_i'\boldsymbol{M}_q\boldsymbol{\varepsilon}_i}{T} = O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{N}\right), \text{ uniformly over } i. \tag{A62}$$

Proof. The procedure in Lemma 3 can be used to prove (A60) and (A62), but replacing all inverses with generalised inverses. This is required since $\boldsymbol{Q}'\boldsymbol{Q}$ has reduced rank when the rank condition does not hold. We need to show that

$$\left\|\frac{1}{T}\boldsymbol{X}_i'\boldsymbol{Q}\left[\left(\bar{\boldsymbol{H}}'\bar{\boldsymbol{H}}\right)^{+} - \left(\boldsymbol{Q}'\boldsymbol{Q}\right)^{+}\right]\bar{\boldsymbol{H}}'\boldsymbol{X}_i\right\| = O_p\left(\frac{1}{\sqrt{N}}\right) \text{ uniformly over } i, \tag{A63}$$

where $^{+}$ denotes the Moore-Penrose inverse. To establish (A63) we need to show that

$$\left(\frac{\boldsymbol{Q}'\boldsymbol{Q}}{T^2}\right)^{+} - \left(\frac{\bar{\boldsymbol{H}}'\bar{\boldsymbol{H}}}{T^2}\right)^{+} = O_p\left(\frac{1}{T\sqrt{N}}\right). \tag{A64}$$

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

$$\left(\frac{\boldsymbol{Q}'\boldsymbol{Q}}{T^2}\right) - \left(\frac{\bar{\boldsymbol{H}}'\bar{\boldsymbol{H}}}{T^2}\right) = O_p\left(\frac{1}{T\sqrt{N}}\right), \tag{A65}$$

for (A64) to hold. But, by Theorem 2 of Andrews (1987), (A65) is sufficient for (A64), if additionally, as $(N,T) \xrightarrow{j} \infty$

$$\lim_{N,T \xrightarrow{j} \infty} \Pr\left(rk\left(\frac{\bar{\boldsymbol{H}}'\bar{\boldsymbol{H}}}{T^2}\right) = rk\left(\frac{\boldsymbol{Q}'\boldsymbol{Q}}{T^2}\right)\right) = 1 \tag{A66}$$

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

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

with

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

for all $\epsilon > 0$. Also

$$rk(T^{-2}\boldsymbol{Q}'\boldsymbol{Q}) = n + rk(\bar{\boldsymbol{C}}),$$

for all N and T, with $rk(T^{-2}\boldsymbol{Q}'\boldsymbol{Q}) \to n + rk(\boldsymbol{C}) < n + m$ as $(N,T) \xrightarrow{j} \infty$. Using these results it is now easily seen that condition (A66) in fact holds. Hence, the desired result follows.

Consider now (A61). Following a similar line of analysis used to establish (A60), we have

$$\left\|\frac{\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{F}}{T} - \frac{\boldsymbol{X}_i'\boldsymbol{M}_q\boldsymbol{F}}{T}\right\| = \left\|\frac{\boldsymbol{X}_i'\bar{\boldsymbol{H}}\left(\bar{\boldsymbol{H}}'\bar{\boldsymbol{H}}\right)^+\bar{\boldsymbol{H}}'\boldsymbol{F}}{T} - \frac{\boldsymbol{X}_i'\boldsymbol{Q}\left(\boldsymbol{Q}'\boldsymbol{Q}\right)^+\boldsymbol{Q}'\boldsymbol{F}}{T}\right\|$$

$$\leq \left\|\frac{1}{T}\left(\boldsymbol{X}_i'\bar{\boldsymbol{H}} - \boldsymbol{X}_i'\boldsymbol{Q}\right)\left(\bar{\boldsymbol{H}}'\bar{\boldsymbol{H}}\right)^+\bar{\boldsymbol{H}}'\boldsymbol{F}\right\| + \left\|\frac{1}{T}\boldsymbol{X}_i'\boldsymbol{Q}\left(\left(\bar{\boldsymbol{H}}'\bar{\boldsymbol{H}}\right)^+ - \left(\boldsymbol{Q}'\boldsymbol{Q}\right)^+\right)\bar{\boldsymbol{H}}'\boldsymbol{F}\right\| +$$
$$\left\|\frac{1}{T}\boldsymbol{X}_i'\boldsymbol{Q}\left(\boldsymbol{Q}'\boldsymbol{Q}\right)^+\left(\bar{\boldsymbol{H}}'\boldsymbol{F} - \boldsymbol{Q}'\boldsymbol{F}\right)\right\|. \tag{A67}$$

Consider each of the above terms in turn. First,

$$\left\|\frac{1}{T}\left(\boldsymbol{X}_i'\bar{\boldsymbol{H}} - \boldsymbol{X}_i'\boldsymbol{Q}\right)\left(\bar{\boldsymbol{H}}'\bar{\boldsymbol{H}}\right)^+\bar{\boldsymbol{H}}'\boldsymbol{F}\right\| \leq \left\|\frac{\boldsymbol{X}_i'\bar{\boldsymbol{H}}}{T} - \frac{\boldsymbol{X}_i'\boldsymbol{Q}}{T}\right\|\left\|\left(\frac{\bar{\boldsymbol{H}}'\bar{\boldsymbol{H}}}{T^2}\right)^+\frac{\bar{\boldsymbol{H}}'\boldsymbol{F}}{T^2}\right\|$$
$$= \left\|\frac{\boldsymbol{X}_i'\bar{\boldsymbol{U}}^*}{T}\right\|\left\|\left(\frac{\bar{\boldsymbol{H}}'\bar{\boldsymbol{H}}}{T^2}\right)^+\frac{\bar{\boldsymbol{H}}'\boldsymbol{F}}{T^2}\right\|$$
$$= O_p\left(\frac{1}{\sqrt{N}}\right), \text{ uniformly over } i, \tag{A68}$$

by (A4), (A9) and (A10). Second, by (A64) and (A65),

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

if

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

We have

$$\left\| \frac{\boldsymbol{Q}'\boldsymbol{Q}}{T} - \frac{\bar{\boldsymbol{H}}'\bar{\boldsymbol{H}}}{T} \right\| = \left\| \frac{\bar{\boldsymbol{U}}^{*\prime}\bar{\boldsymbol{U}}^{*}}{T} + \frac{\boldsymbol{Q}'\bar{\boldsymbol{U}}^{*}}{T} + \frac{\bar{\boldsymbol{U}}^{*\prime}\boldsymbol{Q}}{T} \right\|$$
$$= O_p\left(\frac{1}{\sqrt{N}}\right) \text{ uniformly over } i \tag{A69}$$

by (A1), (A5), (A7), (A9), (A6) and (A10). Finally

$$\left\| \frac{1}{T}\boldsymbol{X}_i'\boldsymbol{Q}\left(\boldsymbol{Q}'\boldsymbol{Q}\right)^{+}\left(\bar{\boldsymbol{H}}'\boldsymbol{F} - \boldsymbol{Q}'\boldsymbol{F}\right) \right\| \leq \left\| \frac{\boldsymbol{X}_i'\boldsymbol{Q}}{T^2}\left(\frac{\boldsymbol{Q}'\boldsymbol{Q}}{T^2}\right)^{+} \right\| \left\| \frac{\bar{\boldsymbol{U}}^{*\prime}\boldsymbol{F}}{T} \right\|$$
$$= O_p\left(\frac{1}{\sqrt{N}}\right), \quad \text{uniformly over } i, \tag{A70}$$

by (A7), (A6), and (A3). Substituting (A68)-(A70) into (A67) yields

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

as required. ■

Appendix B: Proofs of theorems for pooled estimators when the rank condition holds

Proof of Theorem 1

Using (A111) we have

$$\sqrt{N}\left(\hat{\boldsymbol{b}}_{MG} - \boldsymbol{\beta}\right) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\varkappa_i + \frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{\Psi}}_{iT}^{-1}\left(\frac{\sqrt{N}\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{F}}{T}\right)\boldsymbol{\gamma}_i +$$
$$\frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{\Psi}}_{iT}^{-1}\left(\frac{\sqrt{N}\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{\varepsilon}_i}{T}\right) \tag{A71}$$

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

$$\frac{\sqrt{N}\left(\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{F}\right)}{T} = O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\frac{1}{\sqrt{N}}\right), \text{ uniformly over } i, \tag{A72}$$

and so, by the uniform boundedness assumption on $\boldsymbol{\gamma}_i$, and by (A34), we have that

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

and so

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

By Lemma 3, we have that

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

which implies that

$$\frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{\Psi}}_{iT}^{-1}\left(\frac{\sqrt{N}\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{\varepsilon}_i}{T}\right)=O_p\left(\frac{1}{\sqrt{T}}\right)+O_p\left(\frac{1}{\sqrt{N}}\right). \tag{A73}$$

Thus

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

Hence

$$\sqrt{N}\left(\hat{\boldsymbol{b}}_{MG}-\boldsymbol{\beta}\right)\stackrel{d}{\rightarrow}N(\boldsymbol{0},\boldsymbol{\Omega}_{\varkappa}),\ \text{as}\ (N,T)\stackrel{j}{\rightarrow}\infty. \tag{A74}$$

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

$$\hat{\boldsymbol{\Sigma}}_{MG}=\frac{1}{N-1}\sum_{i=1}^{N}\left(\hat{\boldsymbol{b}}_i-\hat{\boldsymbol{b}}_{MG}\right)\left(\hat{\boldsymbol{b}}_i-\hat{\boldsymbol{b}}_{MG}\right)'. \tag{A75}$$

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

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

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

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

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

Proof of Theorem 2

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

$$\sqrt{N}\left(\hat{\boldsymbol{b}}_P-\boldsymbol{\beta}\right)=\left(\frac{1}{N}\sum_{i=1}^{N}\frac{\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{X}_i}{T}\right)^{-1}\left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{\boldsymbol{X}_i'\bar{\boldsymbol{M}}(\boldsymbol{X}_i\varkappa_i+\boldsymbol{\varepsilon}_i)}{T}+\boldsymbol{q}_{NT}\right], \tag{A76}$$

where

$$\boldsymbol{q}_{NT}=\frac{1}{N}\sum_{i=1}^{N}\frac{\sqrt{N}\left(\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{F}\right)\boldsymbol{\gamma}_i}{T}. \tag{A77}$$

By (A72), $\boldsymbol{q}_{NT}=O_p\left(\frac{1}{\sqrt{T}}\right)+O_p\left(\frac{1}{\sqrt{N}}\right)$. Thus

$$\sqrt{N}\left(\hat{\boldsymbol{b}}_P-\boldsymbol{\beta}\right)=\left(\frac{1}{N}\sum_{i=1}^{N}\frac{\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{X}_i}{T}\right)^{-1}\left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{\boldsymbol{X}_i'\bar{\boldsymbol{M}}(\boldsymbol{X}_i\varkappa_i+\boldsymbol{\varepsilon}_i)}{T}\right]+O_p\left(\frac{1}{\sqrt{T}}\right)+O_p\left(\frac{1}{\sqrt{N}}\right). \tag{A78}$$

Further, by (A35)

$$\sqrt{N}\left(\hat{\boldsymbol{b}}_P-\boldsymbol{\beta}\right)=\left(\frac{1}{N}\sum_{i=1}^{N}\frac{\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{X}_i}{T}\right)^{-1}\left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{X}_i\varkappa_i+\boldsymbol{X}_i'\boldsymbol{M}_g\boldsymbol{\varepsilon}_i}{T}\right]+O_p\left(\frac{1}{\sqrt{T}}\right)+O_p\left(\frac{1}{\sqrt{N}}\right). \tag{A79}$$

By (A34) and since by Assumption 6, $N^{-1}\sum_{i=1}^{N}T^{-1}\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{X}_i$ is nonsingular, we have

$$\left(\frac{1}{N}\sum_{i=1}^{N}\frac{\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{X}_i}{T}\right)^{-1}\stackrel{p}{\rightarrow}\boldsymbol{\Psi}^{*-1},$$

where

$$\boldsymbol{\Psi}^* = \lim_{N\to\infty}\left(N^{-1}\sum_{i=1}^{N}\boldsymbol{\Sigma}_{\boldsymbol{v}_i}\right). \tag{A80}$$

Next, we examine the second component of the first term of the RHS on (A78). We first consider $\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{X}_i}{T}\varkappa_i$. We define $\bar{\boldsymbol{M}}_{-i}$ as $\bar{\boldsymbol{M}}_{-i} = \boldsymbol{I}_T - \bar{\boldsymbol{H}}_{-i}\left(\bar{\boldsymbol{H}}'_{-i}\bar{\boldsymbol{H}}_{-i}\right)^{-1}\bar{\boldsymbol{H}}'_{-i}$ where $\bar{\boldsymbol{H}}_{-i} = (\boldsymbol{D},\ \bar{\boldsymbol{Z}}_{-i})$, $\bar{\boldsymbol{Z}}_{-i}$ is $T\times(k+1)$ matrix of observations on $\boldsymbol{d}_t$ and $\bar{\boldsymbol{z}}_{t,-i}$ and $\bar{\boldsymbol{z}}_{t,-i} = \frac{1}{N}\sum_{j=1,j\neq i}^{N}\boldsymbol{z}_{jt}$. Then, it is straightforward to see that

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

and so

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{X}_i}{T}\varkappa_i - \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{\boldsymbol{X}_i'\bar{\boldsymbol{M}}_{-i}\boldsymbol{X}_i}{T}\varkappa_i = O_p\left(\frac{1}{N^{1/2}}\right) \tag{A81}$$

where the uniformity follows by the assumption that $\varkappa_i$ has uniformly finite fourth moments. Since $\varkappa_i$ is i.i.d. and independent of all other stochastic quantities in the model, it follows that $\tilde{\varkappa}_{Ti} = T^{-1}\boldsymbol{X}_i'\bar{\boldsymbol{M}}_{-i}\boldsymbol{X}_i\varkappa_i$ is a martingale difference triangular array, since, for any ordering of the cross-sectional units,

$$E\left(T^{-1}\boldsymbol{X}_i'\bar{\boldsymbol{M}}_{-i}\boldsymbol{X}_i\varkappa_i|i-1,...,1\right) = 0.$$

Then, as long as $E\left\|T^{-1}\boldsymbol{X}_i'\bar{\boldsymbol{M}}_{-i}\boldsymbol{X}_i\right\|^2 < \infty$, which is satisfied by Assumption 6, a central limit theorem holds for $\tilde{\varkappa}_{Ti}$, by Theorem 24.3 of Davidson (1994). Also, by Assymption 2(ii) of this paper, Theorem 1 of De Jong (1997) and Example 17.17 of Davidson (1994), it follows that

$$\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{v}_{it}\varepsilon_{it} = O_p\left(\frac{1}{\sqrt{T}}\right), \text{ uniformly over } i, \tag{A82}$$

which implies that

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

Hence, as $(N,T)\overset{j}{\to}\infty$

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

where

$$\boldsymbol{\Sigma}_P^* = \boldsymbol{\Psi}^{*-1}\boldsymbol{R}^*\boldsymbol{\Psi}^{*-1}, \tag{A83}$$

$$\boldsymbol{R}^* = \lim_{N,T\to\infty}\left[N^{-1}\sum_{i=1}^{N}\boldsymbol{\Sigma}_{\boldsymbol{v\Omega}iT}\right], \tag{A84}$$

where $\boldsymbol{\Sigma}_{\boldsymbol{v\Omega}iT}$ denotes the variance of $\frac{\boldsymbol{X}_i'\boldsymbol{M}_g\boldsymbol{X}_i}{T}\varkappa_i$. The variance estimator for $\boldsymbol{\Sigma}_P^*$ suggested by Pesaran (2006) is given by

$$\hat{\boldsymbol{\Sigma}}_P^* = \hat{\boldsymbol{\Psi}}^{*-1}\hat{\boldsymbol{R}}^*\hat{\boldsymbol{\Psi}}^{*-1}, \tag{A85}$$

where

$$\hat{\boldsymbol{\Psi}}^* = N^{-1}\sum_{i=1}^{N}\left(\frac{\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{X}_i}{T}\right), \tag{A86}$$

$$\hat{\boldsymbol{R}}^* = \frac{1}{(N-1)}\sum_{i=1}^{N}\left(\frac{\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{X}_i}{T}\right)\left(\hat{\boldsymbol{b}}_i - \hat{\boldsymbol{b}}_{MG}\right)\left(\hat{\boldsymbol{b}}_i - \hat{\boldsymbol{b}}_{MG}\right)'\left(\frac{\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{X}_i}{T}\right). \tag{A87}$$

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

Appendix C: Proofs of theorems for pooled estimator when the rank condition does not hold

Proof of Theorem 1 when the rank condition does not hold

In the case where the rank condition does not hold, it is easy to infer from (A49), (A51), (A56) and (A58) that

$$\left(\frac{\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{F}}{T}\right)\bar{\boldsymbol{C}} = O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right), \text{ uniformly over } i. \tag{A88}$$

But, we know that

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

where $\bar{\boldsymbol{\Gamma}} = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Gamma}_i$ and $\bar{\boldsymbol{\gamma}} = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\gamma}_i$. Substituting this result in (A88) now yields

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

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

which in turn yields

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

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

$$\frac{\sqrt{N}\left(\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{F}\right)\bar{\boldsymbol{\gamma}}}{T} = O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right), \text{ uniformly over } i. \tag{A89}$$

We next reconsider the second term on the RHS of (A71), which is the only term affected by the fact that rank condition does not hold. The second term on the RHS in (A71) can be written as

$$\boldsymbol{\chi}_{NT} \equiv \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{X}_i}{T^2}\right)^{+}\left(\frac{\sqrt{N}\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{F}}{T^2}\right)\left(\bar{\boldsymbol{\gamma}} + \boldsymbol{\eta}_i - \bar{\boldsymbol{\eta}}\right), \tag{A90}$$

where $\bar{\boldsymbol{\eta}} = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\eta}_i$. By (A60) and (A61) it follows that

$$\boldsymbol{\chi}_{NT} \equiv \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\boldsymbol{X}_i'\boldsymbol{M}_q\boldsymbol{X}_i}{T^2}\right)^{+}\left(\frac{\sqrt{N}\boldsymbol{X}_i'\boldsymbol{M}_q\boldsymbol{F}}{T^2}\right)\left(\bar{\boldsymbol{\gamma}} + \boldsymbol{\eta}_i - \bar{\boldsymbol{\eta}}\right) + O_p\left(\frac{1}{\sqrt{N}}\right). \tag{A91}$$

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 $\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{X}_i$, $\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{F}$, $\boldsymbol{X}_i'\boldsymbol{M}_q\boldsymbol{X}_i$ and $\boldsymbol{X}_i'\boldsymbol{M}_q\boldsymbol{F}$ would involve nonstationary components. Then, since by (A89), $\frac{\sqrt{N}\left(\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{F}\right)\bar{\boldsymbol{\gamma}}}{T^2} = O_p\left(\frac{1}{T\sqrt{N}}\right) + O_p\left(\frac{1}{T^{3/2}}\right)$, uniformly over i, it is the case that for N and T large

$$\sqrt{N}\left(\hat{\boldsymbol{b}}_{MG} - \boldsymbol{\beta}\right) \overset{d}{\sim} \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\varkappa_i + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\frac{\boldsymbol{X}_i'\boldsymbol{M}_q\boldsymbol{X}_i}{T^2}\right)^{+}\left(\frac{\boldsymbol{X}_i'\boldsymbol{M}_q\boldsymbol{F}}{T^2}\right)\left(\boldsymbol{\eta}_i - \bar{\boldsymbol{\eta}}\right). \tag{A92}$$

The first term on the right hand side of (A92) tends to a Normal density with mean zero and finite variance. The second term needs further analysis. Letting

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

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

$$\frac{1}{\sqrt{N}}\sum_{i=1}^N \boldsymbol{Q}_{1iT}\left(\boldsymbol{\eta}_i - \bar{\boldsymbol{\eta}}\right) = \frac{1}{\sqrt{N}}\sum_{i=1}^N \left(\boldsymbol{Q}_{1iT} - \bar{\boldsymbol{Q}}_{1T}\right)\boldsymbol{\eta}_i. \tag{A93}$$

We note that $\boldsymbol{\eta}_i$ is i.i.d. with zero mean and finite variance and independent of all other stochastic quantities in the second term of the RHS on (A93). Next, we carry out a similar analysis to that around (A81). We define

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

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

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

and

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

Then, it is easy to show that if $z_{Ti} = x_i y_{Ti}$, x_i is an i.i.d. sequence with zero mean and finite variance and y_{Ti} is a triangular array of random variables with finite variance then z_{Ti} is a martingale difference triangular array for which a central limit theorem holds (see, e.g., Theorem 24.3 of Davidson (1994)). But this is the case here, for any ordering over i, setting $y_{Ti} = \left(\boldsymbol{Q}_{1iT,-i} - \bar{\boldsymbol{Q}}_{1T,-i}\right)$ and $x_i = \boldsymbol{\eta}_i$. Using this result, it follows that the second term on the RHS of (A71) tends to a Normal density if $\left(\boldsymbol{Q}_{1iT} - \bar{\boldsymbol{Q}}_{1T}\right)\boldsymbol{\eta}_i$ has variance with finite norm, uniformly over i, denoted by Σ_{iqT}. In order to establish the existence of second moments, it is sufficient to prove that $\left\|\left(\boldsymbol{Q}_{1iT} - \bar{\boldsymbol{Q}}_{1T}\right)\right\|$, or equivalently $\left\|\left(\boldsymbol{Q}_{1iT,-i} - \bar{\boldsymbol{Q}}_{1T,-i}\right)\right\|$, has finite second moments. We carry out the analysis for $\left\|\left(\boldsymbol{Q}_{1iT} - \bar{\boldsymbol{Q}}_{1T}\right)\right\|$. For this, we need to provide further analysis of $\frac{\boldsymbol{X}_i'\boldsymbol{M}_q\boldsymbol{X}_i}{T^2}$ and $\frac{\boldsymbol{X}_i'\boldsymbol{M}_q\boldsymbol{F}}{T^2}$. First, note that $\boldsymbol{X}_i$ can be written as

$$\boldsymbol{X}_i = \boldsymbol{Q}\boldsymbol{B}_{i1} + \boldsymbol{S}\boldsymbol{B}_{i2} + \boldsymbol{V}_i, \tag{A94}$$

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

$$\boldsymbol{F} = \boldsymbol{Q}\boldsymbol{K}_1 + \boldsymbol{S}\boldsymbol{K}_2. \tag{A95}$$

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

$$\boldsymbol{X}_i'\boldsymbol{M}_q\boldsymbol{X}_i = \boldsymbol{X}_i'\boldsymbol{M}_q\left(\boldsymbol{Q}\boldsymbol{B}_{i1} + \boldsymbol{S}\boldsymbol{B}_{i2} + \boldsymbol{V}_i\right) = \boldsymbol{X}_i'\boldsymbol{M}_q\boldsymbol{S}\boldsymbol{B}_{i2} + \boldsymbol{X}_i'\boldsymbol{M}_q\boldsymbol{V}_i =$$
$$\boldsymbol{B}_{i2}'\boldsymbol{S}'\boldsymbol{M}_q\boldsymbol{S}\boldsymbol{B}_{i2} + \boldsymbol{V}_i'\boldsymbol{M}_q\boldsymbol{V}_i + \boldsymbol{B}_{i2}'\boldsymbol{S}'\boldsymbol{M}_q\boldsymbol{V}_i + \boldsymbol{V}_i'\boldsymbol{M}_q\boldsymbol{S}\boldsymbol{B}_{i2}.$$

But, it easily follows that

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

and

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

Then,

$$\frac{\boldsymbol{X}_i'\boldsymbol{M}_q\boldsymbol{X}_i}{T^2} = \boldsymbol{B}_{i2}'\frac{\boldsymbol{S}'\boldsymbol{S}}{T^2}\boldsymbol{B}_{i2} + O_p\left(\frac{1}{T}\right), \text{ uniformly over } i. \tag{A96}$$

Similarly, using (A95),

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

Thus

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

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

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

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

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

and we obtain

$$\left(\boldsymbol{B}_{i2}'\frac{\boldsymbol{S}'\boldsymbol{S}}{T^2}\boldsymbol{B}_{i2}\right)^+ = \boldsymbol{B}_{i2}'\left(\boldsymbol{B}_{i2}\boldsymbol{B}_{i2}'\right)^{-1}\left(\frac{\boldsymbol{S}'\boldsymbol{S}}{T^2}\right)^{-1}\left(\boldsymbol{B}_{i2}\boldsymbol{B}_{i2}'\right)^{-1}\boldsymbol{B}_{i2}. \tag{A97}$$

Hence

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

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

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

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

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

and the analysis proceeds as above until

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

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

Thus, in general we have that

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

where

$$\boldsymbol{\Sigma}_{MG} = \boldsymbol{\Omega}_{\varkappa} + \boldsymbol{\Lambda}, \tag{A98}$$

and

$$\boldsymbol{\Lambda} = \lim_{N,T\to\infty}\left[\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Sigma}_{iqT}\right]. \tag{A99}$$

To complete the proof we need to show that the variance estimator given by (A75) is still consistent. To see this first note that

$$\hat{\boldsymbol{b}}_i - \boldsymbol{\beta} = \varkappa_i + \boldsymbol{h}_{iT} + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right), \text{ uniformly over } i, \tag{A100}$$

where

$$\boldsymbol{h}_{iT} = \left(\frac{\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{X}_i}{T^2}\right)^{+}\frac{\boldsymbol{X}_i'\bar{\boldsymbol{M}}\left[\boldsymbol{F}\left(\boldsymbol{\eta}_i - \bar{\boldsymbol{\eta}}\right) + \boldsymbol{\varepsilon}_i\right]}{T^2}, \tag{A101}$$

and so

$$\hat{\boldsymbol{b}}_i - \hat{\boldsymbol{b}}_{MG} = \left(\varkappa_i - \bar{\varkappa}\right) + \left(\boldsymbol{h}_{iT} - \bar{\boldsymbol{h}}_T\right) + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right), \text{ uniformly over } i, \tag{A102}$$

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

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

and the desired result follows.

Proof of Theorem 2 when the rank condition does not hold

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

$$\sqrt{N}\left(\hat{\boldsymbol{b}}_P - \boldsymbol{\beta}\right) = \left(\frac{1}{N}\sum_{i=1}^{N}\frac{\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{X}_i}{T^2}\right)^{-1}\left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{\boldsymbol{X}_i'\bar{\boldsymbol{M}}(\boldsymbol{X}_i\varkappa_i + \boldsymbol{\varepsilon}_i)}{T^2} + \boldsymbol{q}_{NT}\right], \tag{A103}$$

where

$$\boldsymbol{q}_{NT} = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{\left(\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{F}\right)\boldsymbol{\gamma}_i}{T^2}. \tag{A104}$$

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

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

But by (A89), the first component of $\boldsymbol{q}_{NT}$ is $O_p\left(\frac{1}{T\sqrt{N}}\right) + O_p\left(\frac{1}{T^{3/2}}\right)$. Substituting this result in (A103), and making use of (33) and (34) we have

$$\sqrt{N}\left(\hat{\boldsymbol{b}}_P - \boldsymbol{\beta}\right) = \left(\frac{1}{N}\sum_{i=1}^{N}\frac{\boldsymbol{X}_i'\boldsymbol{M}_q\boldsymbol{X}_i}{T^2}\right)^{-1}\left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{\boldsymbol{X}_i'\boldsymbol{M}_q(\boldsymbol{X}_i\varkappa_i + \boldsymbol{\varepsilon}_i + \boldsymbol{F}\left(\boldsymbol{\eta}_i - \bar{\boldsymbol{\eta}}\right))}{T^2}\right] + O_p\left(\frac{1}{T\sqrt{N}}\right) + O_p\left(\frac{1}{T^{3/2}}\right). \tag{A105}$$

Also by Assumption 7, when the rank condition is not satisfied, $\frac{1}{N}\sum_{i=1}^{N}\frac{\boldsymbol{X}_i'\boldsymbol{M}_q\boldsymbol{X}_i}{T^2}$ is nonsingular. Further, by (A96),

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

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

$$\boldsymbol{\Theta}=\lim_{N,T\to\infty}\left(\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Theta}_{iT}\right), \tag{A106}$$

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

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

where

$$\boldsymbol{\Xi}=\lim_{N,T\to\infty}\left(\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Xi}_{Ti}\right), \tag{A107}$$

and $\boldsymbol{\Xi}_{Ti}$ denotes the variance of $T^{-2}\boldsymbol{X}_i'\boldsymbol{M}_q\boldsymbol{X}_i\varkappa_i$. Further, by independence of $\boldsymbol{\varepsilon}_i$ across i,

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

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

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

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

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

where

$$\boldsymbol{\Phi}=\lim_{N,T\to\infty}\left(\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Phi}_{Ti}\right) \tag{A108}$$

and $\boldsymbol{\Phi}_{Ti}$ denotes the variance of $\left(\boldsymbol{Q}_{2iT}-\bar{\boldsymbol{Q}}_{2T}\right)\boldsymbol{\eta}_i$. Thus, overall by the independence of $\varkappa_i$ and $\boldsymbol{\eta}_i$, it follows that

$$\sqrt{N}\left(\hat{\boldsymbol{b}}_P-\boldsymbol{\beta}\right)\xrightarrow{d}N(\boldsymbol{0},\boldsymbol{\Sigma}_P^*),\ \text{as }(N,T)\xrightarrow{j}\infty, \tag{A109}$$

where, now

$$\boldsymbol{\Sigma}_P^*=\boldsymbol{\Theta}^{-1}\left(\boldsymbol{\Xi}+\boldsymbol{\Phi}\right)\boldsymbol{\Theta}^{-1} \tag{A110}$$

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

Appendix D: Proof of Theorem 3

Using (25) in (15) we have

$$\hat{\boldsymbol{b}}_i-\boldsymbol{\beta}_i=\left(\frac{\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{X}_i}{T}\right)^{-1}\left(\frac{\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{F}}{T}\right)\gamma_i+\left(\frac{\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{X}_i}{T}\right)^{-1}\left(\frac{\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{\varepsilon}_i}{T}\right). \tag{A111}$$

Using (A35) and (A44), and assuming that the rank condition (9) is satisfied we have

$$\hat{\boldsymbol{b}}_i - \boldsymbol{\beta}_i = \left(\frac{\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{X}_i}{T}\right)^{-1}\left(\frac{\boldsymbol{X}_i'\boldsymbol{M}_g\boldsymbol{\varepsilon}_i}{T}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{N}\right). \tag{A112}$$

For N and T sufficiently large, the distribution of $\sqrt{T}\left(\hat{\boldsymbol{b}}_i - \boldsymbol{\beta}_i\right)$ will be asymptotically normal if the rank condition (9) is satisfied and if $\sqrt{T}/N \to 0$ as N and $T \to \infty$. To see why this additional condition is needed, using (A112) note that

$$\sqrt{T}\left(\hat{\boldsymbol{b}}_i - \boldsymbol{\beta}_i\right) = \left(\frac{\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{X}_i}{T}\right)^{-1}\frac{\boldsymbol{X}_i'\boldsymbol{M}_g\boldsymbol{\varepsilon}_i}{\sqrt{T}} + O_p\left(\frac{\sqrt{T}}{N}\right) + O_p\left(\frac{1}{\sqrt{N}}\right), \tag{A113}$$

and the asymptotic distribution of $\sqrt{T}\left(\hat{\boldsymbol{b}}_i - \boldsymbol{\beta}_i\right)$ will be free of nuisance parameters only if $\sqrt{T}/N \to 0$, as $(N,T) \overset{j}{\to} \infty$. We now give the necessary arguments for showing that the first term on the RHS of (A113) is asymptotically normally distributed. We note that

$$\begin{aligned}\frac{\boldsymbol{X}_i'\boldsymbol{M}_g\boldsymbol{\varepsilon}_i}{\sqrt{T}} &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[-\left(\hat{\boldsymbol{\Pi}}_i - \boldsymbol{\Pi}_i\right)'\boldsymbol{g}_t + \boldsymbol{v}_{it}\right]\varepsilon_{it} \\ &= -\left(\hat{\boldsymbol{\Pi}}_i - \boldsymbol{\Pi}_i\right)'\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\boldsymbol{g}_t\varepsilon_{it} + \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\boldsymbol{v}_{it}\varepsilon_{it}.\end{aligned} \tag{A114}$$

But, it is straightforward to show that the first term of (A114) is $O_p(T^{-1/2})$ when $\boldsymbol{g}_t$ is $I(1)$. Then, we need to obtain a central limit theorem for the second term of (A114). But, by the martingale difference assumption on ε_{it}, it follows that $\boldsymbol{v}_{it}\varepsilon_{it}$ is also a martingale difference sequence with finite variance given by $\sigma_i^2\boldsymbol{\Sigma}_{\boldsymbol{v}_i}$. Then, by Theorem 24.3 of Davidson (1994), it follows that

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\boldsymbol{v}_{it}\varepsilon_{it} \overset{d}{\to} N(\boldsymbol{0}, \sigma_i^2\boldsymbol{\Sigma}_{\boldsymbol{v}_i}). \tag{A115}$$

Further, by (A34) and noting that by Assumptions 5 and 6, $\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{X}_i/T$ and $\boldsymbol{X}_i'\boldsymbol{M}_g\boldsymbol{X}_i/T$ are nonsingular, we also have

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

and, by Lemma 5, it follows that

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

finally implying that

$$\sqrt{T}\left(\hat{\boldsymbol{b}}_i - \boldsymbol{\beta}_i\right) \overset{d}{\to} N(0, \sigma_i^2\boldsymbol{\Sigma}_{\boldsymbol{v}_i}^{-1}), \tag{A116}$$

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

$$\mathring{\sigma}_i^2\left(\frac{\boldsymbol{X}_i'\bar{\boldsymbol{M}}\boldsymbol{X}_i}{T}\right)^{-1}, \tag{A117}$$

where

$$\mathring{\sigma}_i^2 = \frac{\left(\boldsymbol{y}_i - \boldsymbol{X}_i\hat{\boldsymbol{b}}_i\right)'\bar{\boldsymbol{M}}\left(\boldsymbol{y}_i - \boldsymbol{X}_i\hat{\boldsymbol{b}}_i\right)}{T - (n + 2k + 1)}. \tag{A118}$$