ECONSTOR
*Make Your Publications Visible.*

A Service of

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Pesaran, Mohammad Hashem

**Working Paper**

# Estimation and inference in large heterogeneous panels with a multifactor error structure

CESifo Working Paper, No. 1331

**Provided in Cooperation with:**
Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

*Suggested Citation:* Pesaran, Mohammad Hashem (2004) : Estimation and inference in large heterogeneous panels with a multifactor error structure, CESifo Working Paper, No. 1331, Center for Economic Studies and ifo Institute (CESifo), Munich

This Version is available at:
https://hdl.handle.net/10419/18696

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

# ESTIMATION AND INFERENCE IN LARGE HETEROGENEOUS PANELS WITH A MULTIFACTOR ERROR STRUCTURE

M. HASHEM PESARAN

CESIFO WORKING PAPER NO. 1331
CATEGORY 10: EMPIRICAL AND THEORETICAL METHODS
NOVEMBER 2004

*An electronic version of the paper may be downloaded*
- *from the SSRN website: www.SSRN.com*
- *from the CESifo website: www.CESifo.de*

CESifo *Working Paper No. 1331*

# ESTIMATION AND INFERENCE IN LARGE HETEROGENEOUS PANELS WITH A MULTIFACTOR ERROR STRUCTURE

## Abstract

This paper presents a new approach to estimation and inference in panel data models with a multifactor error structure where the unobserved common factors are (possibly) correlated with exogenously given individual-specific regressors, and the factor loadings differ over the cross section units. The basic idea behind the proposed estimation procedure is to filter the individual-specific regressors by means of (weighted) cross-section aggregates such that asymptotically as the cross-section dimension (N) tends to infinity the differential effects of unobserved common factors are eliminated. The estimation procedure has the advantage that it can be computed by OLS applied to an auxiliary regression where the observed regressors are augmented by (weighted) cross sectional averages of the dependent variable and the individual specific regressors. Two different but related problems are addressed: one that concerns the coefficients of the individual-specific regressors, and the other that focusses on the mean of the individual coefficients assumed random. In both cases appropriate estimators, referred to as common correlated effects (CCE) estimators, are proposed and their asymptotic distribution as $N \to \infty$, with T (the time-series dimension) fixed or as N and $T \to \infty$ (jointly) are derived under different regularity conditions. One important feature of the proposed CCE mean group (CCEMG) estimator is its invariance to the (unknown but fixed) number of unobserved common factors as N and $T \to \infty$ (jointly). The small sample properties of the various pooled estimators are investigated by Monte Carlo experiments that confirm the theoretical derivations and show that the pooled estimators have generally satisfactory small sample properties even for relatively small values of N and T.

JEL Code: C12, C13, C33.

Keywords: cross section dependence, large panels, common correlated effects, heterogeneity, estimation and inference.

*M. Hashem Pesaran*
*Faculty of Economics and Politics*
*University of Cambridge*
*Sidgwick Avenue*
*Cambridge, CB3 9DD*
*United Kingdom*
*mhp1@econ.cam.ac.uk*

I am most grateful to a Co-Editor and three anonymous referees for their helpful suggestions and constructive comments on an earlier version which focussed on a one-factor error structure. I would also like to thank George Kapetanios, Yongcheol Shin, Ron Smith, Til Schuermann and Takashi Yamagata for helpful comments and discussions on the current version. Takashi Yamagata also carried out the computations of the Monte Carlo results reported in the paper, most efficiently and beyond the call of duty. Financial support from the ESRC (Grant No. RES-000-23-0135) is gratefully acknowledged.

# 1 Introduction

A number of different approaches have been advanced for the analysis of cross section dependence. In the case of spatial problems where a natural immutable distance measure is available the dependence is captured through "spatial lags" using techniques familiar from time series literature. In economic applications spatial techniques are often adapted using alternative measures of "economic distance". See, for example, Lee and Pesaran (1993), Conley and Topa (2002), Conley and Dupor (2003), 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. This approach allows for general (time-invariant) correlation patterns across the errors in the different cross section equations.

There are also a number of contributions in the literature that allow for time-varying individual effects in the case of panels with homogeneous slopes where $T$ is fixed as $N \to \infty$. Holtz-Eakin, Newey and Rosen (1988) use a quasi-differencing procedure to eliminate the time-varying effects and then estimate the model by instrumental variables. This procedure eliminates the individual-specific effects but yields regression equations with time-varying coefficients that are generally difficult to estimate and is likely to work only when $T$ is quite small. Ahn, Lee and Schmidt (2001), building on the earlier contributions of Kiefer (1980) and Lee (1991) propose a number of different generalized method of moments (GMM) estimators depending on whether first as well as second-order moment restrictions are utilized. In the case where idiosyncratic errors are homoskedastic and nonautocorrelated, they show that the GMM estimator that makes use of all the first and second order moment restrictions dominates the maximum likelihood estimator (which is also the generalized within estimator) originally proposed by Kiefer (1980). However, their analysis assumes that the regressors are identically and independently distributed across the individuals, which may not be valid in practice. In addition, none of these approaches are appropriate when both $N$ and $T$ are large and of the same order of magnitude, as is often the case in cross-country (region) studies.

The application of an unrestricted SURE-GLS approach to large $N$ and $T$ panels involves nuisance parameters that increase at a quadratic rate as the cross section dimension of the panel is allowed to rise. To deal with this problem a number of authors including Robertson and Symons (2000), Coakley, Fuertes and Smith (2002), and Phillips and Sul (2003) propose restricting the covariance matrix of the errors using a common factor specification with a fixed number of unobserved factors. Phillips and Sul (2003) adopt a GLS-SURE procedure for estimation of autoregressive models with heterogeneous slopes (but without exogenous regressors) using a single factor structure for the residuals, but do not provide any large $N$ asymptotic results. Coakley, Fuertes and Smith (2002)

propose a principal components approach that is arguably simpler to implement than Robertson and Symons's full maximum likelihood procedure.[1] These authors also claim that their procedure is valid even if the unobserved common factors and the observed individual effects are correlated, possibly due to omitted global variables or common shocks that are correlated with the included regressors.

In this paper we first establish that in general the estimation procedure proposed by Coakley, Fuertes and Smith (CFS) will not be consistent if the unobserved factors and the included regressors are correlated. We also show that the satisfactory simulation results reported in the paper are due to the paper's special Monte Carlo design where the cross-section average of the included regressor and the unobserved common effect become perfectly correlated as $N \to \infty$. We shall then propose a new approach that yields consistent and asymptotically normal parameter estimates even in the presence of correlated unobserved common effects both when $T$ is fixed and $N \to \infty$, and as $(N, T) \to \infty$, jointly.

We consider a multifactor residual model and distinguish between individual-specific regressors, as well as observed and unobserved common effects. We permit the common effects to have differential impacts on individual units, while at the same time allowing them to exhibit an arbitrary degree of correlation amongst themselves and with the individual-specific regressors. We allow for error variance heterogeneity and do not require the individual-specific regressors to be identically and/or independently distributed over the cross-section units, which is particularly relevant to the analysis of cross-country panels. However, in this paper we assume the individual-specific regressors and the common factors to be stationary and exogenous. Allowing for unit roots and other extensions is currently the subject of further research.

The basic idea behind the proposed estimation procedure is to filter the individual specific regressors by means of cross section aggregates such that asymptotically (as $N \to \infty$) the differential effects of unobserved common factors are eliminated. This is in contrast with the various approaches adopted in the literature that focus on estimation of factor loadings as an input into the GLS algorithm. The estimation approach has the added advantage that it can be computed by ordinary least squares (OLS) applied to an auxiliary regression where the observed regressors are augmented by cross section (weighted) averages of the dependent variable and the individual specific regressors. Using this approach we consider two different but related estimation and inference problems; one that concerns the coefficients of the individual-specific regressors, and the other that focusses on the means of the individual coefficients assumed random as in Swamy (1970). We refer to these as common correlated effects (CCE) estimators and derive their asymptotic distributions under certain regularity conditions.

[1] Similar issues are also discussed in the analysis of (dynamic) factor models by Forni and Lippi (1997), Forni and Reichlin (1998), Stock and Watson (1998), and Bai and Ng (2002), among others.

We show that the CCE estimator of the individual-specific coefficients are asymptotically unbiased as $N \to \infty$ both for $T$ fixed and $T \to \infty$, so long as a certain rank condition concerning the factor loadings is satisfied. In this case the asymptotic distribution of the CCE estimator is shown to be free of nuisance parameters when $T$ is fixed as $N \to \infty$, or if $\sqrt{T}/N \to 0$ , as $N, T \to \infty$, jointly. Building on these results we then show that the mean group estimator based on the individual-specific CCE estimators (referred to as CCEMG) is also asymptotically unbiased as $N \to \infty$ both for $T$ fixed and $T \to \infty$, and derive its asymptotic distribution as $N, T \to \infty$, with no particular restrictions on the convergence rates of $N$ and $T$. The CCEMG estimator continues to hold under slope homogeneity. Remarkably, these results hold for any fixed number of unobserved common effects, which is an important consideration in practice where in general little is known about the unobserved common effects.

Similar results are also obtained for a standard pooled version of the CCE estimator (referred to as CCEP). The CCEP estimator is asymptotically unbiased as $N \to \infty$ both for $T$ fixed and as $T \to \infty$, but under slope homogeneity the derivation of its asymptotic distribution requires $T/N \to 0$ as $N$ and $T \to \infty$. This requirement, however, is not unduly restrictive in micro panels where $T$ is typically small and $N$ relatively large.

The above theoretical results are confirmed by a number of Monte Carlo experiments some of which are summarized in Section 8. Tests based on the CCEMG estimator is shown to have the correct size even for samples as small as $N = 30$ and $T = 20$, with the empirical size being controlled as $(N, T) \to \infty$, jointly. The CCEP estimator behaves similarly, although under slope homogeneity there is evidence of size distortions when $T > N$ (as predicted by the theory). A modified test based on the CCEP estimator is proposed where the variance formula for the heterogeneous slope case is used even if it is believed that the slope coefficients are homogeneous.[2] The resultant test, denoted by CCEP(hetro), shows little size distortions for $N, T \geq 20$, and has better small sample properties than the CCEMG estimator. Both estimators also perform well relative to the infeasible estimator that uses data on the unobserved common effects and assumes a complete knowledge of the residual factor structure. The CCE type estimators come close to replicating the properties of the infeasible estimators without knowledge of the residual factor structure and/or the realizations of the unobserved effects. The Monte Carlo results also illustrate the substantial bias and size distortions that results if error cross section dependence is ignored, which in turn highlight the importance of testing for error cross section dependence in panel data models.[3]

The plan of the paper is as follows: Section 2 sets out the multifactor residual model and its assumptions. Section 3 shows the general inconsistency of the principal components estimator proposed by Coakley, Fuertes and Smith (2002). Section 4 motivates the idea of approximating the

[2]In reality one is, of course, never sure of the validity of the slope homogeneity assumption.

[3]General tests of error cross section dependence are discussed in Pesaran (2004).

unobserved common factor by linear combination of the cross section averages of the dependent and the individual-specific regressors. The CCE estimators of the coefficients of the individual-specific regressors are presented in Section 5, and their pooled counterpart in Section 6. The mean group estimator based on the individual CCE estimators (i.e. CCEMG) is discussed in sub-section (6.1), and the pooled version (i.e. CCEP) in sub-section (6.2). The problems of how best to choose the weights for the construction of the cross-section aggregates and in the formation of the pooled estimator are discussed in Section 7. Section 8 reports the results of the Monte Carlo experiments. Section 9 concludes by identifying important areas for extensions and further developments.

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

## 2 A Multifactor Residual Model

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

$$y_{it} = \boldsymbol{\alpha}_i' \mathbf{d}_t + \boldsymbol{\beta}_i' \mathbf{x}_{it} + e_{it}, \tag{2.1}$$

where $\mathbf{d}_t$ is a $n \times 1$ vector of observed common effects (including deterministics such as intercepts or seasonals dummies), $\mathbf{x}_{it}$ is a $k \times 1$ vector of observed individual-specific regressors on the $i^{th}$ cross section unit at time $t$, and the errors have the multifactor structure

$$e_{it} = \boldsymbol{\gamma}_i' \mathbf{f}_t + \varepsilon_{it}, \tag{2.2}$$

in which $\mathbf{f}_t$ is the $m \times 1$ vector of unobserved common effects, and $\varepsilon_{it}$ are the individual-specific (idiosyncratic) errors assumed to be independently distributed of $(\mathbf{d}_t, \mathbf{x}_{it})$. In general, however, the unobserved factors, $\mathbf{f}_t$, could be correlated with $(\mathbf{d}_t, \mathbf{x}_{it})$, and to allow for such a possibility we adopt the following fairly general model for the individual specific regressors

$$\mathbf{x}_{it} = \mathbf{A}_i' \mathbf{d}_t + \Gamma_i' \mathbf{f}_t + \mathbf{v}_{it}, \tag{2.3}$$

where $\mathbf{A}_i$ and $\Gamma_i$ are $n \times k$ and $m \times k$, factor loading matrices with fixed components, $\mathbf{v}_{it}$ are the specific components of $\mathbf{x}_{it}$ distributed independently of the common effects and across $i$, but

assumed to follow general covariance stationary processes. Unit roots and deterministic trends can be considered in $\mathbf{x}_{it}$ and $y_{it}$ by allowing one or more of the the common effects in $\mathbf{d}_t$ or $\mathbf{f}_t$ to have unit roots and/or deterministic trends. In what follows, however, we focus on the case where $\mathbf{d}_t$ and $\mathbf{f}_t$ are covariance stationary.

Combining (2.1), (2.2) and (2.3) we now have the following system of equations

$$\underset{(k+1)\times 1}{\mathbf{z}_{it}} = \begin{pmatrix} y_{it} \\ \mathbf{x}_{it} \end{pmatrix} = \underset{(k+1)\times n}{\mathbf{B}_i'} \underset{n\times 1}{\mathbf{d}_t} + \underset{(k+1)\times m}{\mathbf{C}_i'} \underset{m\times 1}{\mathbf{f}_t} + \underset{(k+1)\times 1}{\mathbf{u}_{it}}, \tag{2.4}$$

where

$$\mathbf{u}_{it} = \begin{pmatrix} \varepsilon_{it} + \boldsymbol{\beta}_i' \mathbf{v}_{it} \\ \mathbf{v}_{it} \end{pmatrix}, \tag{2.5}$$

$$\mathbf{B}_i = \begin{pmatrix} \boldsymbol{\alpha}_i & \mathbf{A}_i \end{pmatrix} \begin{pmatrix} 1 & \mathbf{0} \\ \boldsymbol{\beta}_i & \mathbf{I}_k \end{pmatrix}, \mathbf{C}_i = \begin{pmatrix} \boldsymbol{\gamma}_i & \Gamma_i \end{pmatrix} \begin{pmatrix} 1 & \mathbf{0} \\ \boldsymbol{\beta}_i & \mathbf{I}_k \end{pmatrix}, \tag{2.6}$$

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

$$\tilde{\Gamma}_i = \begin{pmatrix} \boldsymbol{\gamma}_i & \Gamma_i \end{pmatrix}. \tag{2.7}$$

Throughout we shall assume that $\|\mathbf{B}_i\|$ and $\|\mathbf{C}_i\|$ or their expectations (if assumed random) are bounded.

The above set up is sufficiently general and renders a variety of panel data models as special cases. (i) The familiar fixed or random effects models correspond to the case where $\mathbf{d}_t = 1$, $\boldsymbol{\beta}_i = \boldsymbol{\beta}$ and $\boldsymbol{\gamma}_i = \mathbf{0}$, for all $i$. (ii) The time-varying effects models of Kiefer (1980), Lee (1991) and Ahn, Lee and Schmidt (2001) allow for error cross section dependence through a single unobserved factor but, in addition to assuming that $\mathbf{d}_t = 1$, $\boldsymbol{\beta}_i = \boldsymbol{\beta}$, also require the individual specific regressors to be cross sectionally independent, namely $\mathbf{A}_i = \mathbf{0}$ and $\Gamma_i = \mathbf{0}$. In most applications of interest, however, the individual specific regressors are likely to be cross sectionally dependent and a formulation such as (2.3) will be far more widely applicable. (iii) The random coefficient model of Swamy (1970) allows for slope heterogeneity but assumes $\boldsymbol{\gamma}_i = \mathbf{0}$, for all $i$. (iv) In the special case where $\boldsymbol{\gamma}_i = \boldsymbol{\gamma}$, the multifactor structure reduces to $\gamma_t = \boldsymbol{\gamma}' \mathbf{f}_t$, and (2.1) and (2.2) become the familiar panel data model with time dummies. In this case the estimation of $\boldsymbol{\beta}$ can be achieved using standard panel data estimators based on cross sectionally de-meaned observations. (v) The large $N$ and $T$ factor models recently analyzed by Stock and Watson (1998) and Bai and Ng (2002) focus on consistent estimation of $\mathbf{f}_t$ (including its dimension $m$) and the factor loadings, $\boldsymbol{\gamma}_i$, and are not concerned with the estimation of the "structural" parameters $\boldsymbol{\beta}_i$, and in effect set them to zero.[4]

[4]Note that $\boldsymbol{\beta}_i$ is unidentified if, as maintained in the factor models, the variance matrix of $\mathbf{u}_{it}$ is unrestricted. The assumption that $\mathbf{v}_{it}$ and $\mathbf{u}_{it}$ in (2.5) are uncorrelated provides the $k$ restrictions needed for the exact identification of $\boldsymbol{\beta}_i$.

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. In this paper we shall focus on the estimation and inference problems relating to $E(\boldsymbol{\beta}_i) = \boldsymbol{\beta}$, and discuss the circumstances under which the individual slope coefficients, $\boldsymbol{\beta}_i$, can also be consistently estimated and tested. To this end we make the following assumptions:

**Assumption 1** (common effects): The $(n + m) \times 1$ vector of common effects, $\mathbf{g}_t = (\mathbf{d}_t', \mathbf{f}_t')'$, is covariance stationary with absolute summable autocovariances, distributed independently of the individual-specific errors, $\varepsilon_{it'}$ and $\mathbf{v}_{it'}$ for all $i$, $t$ and $t'$.

**Assumption 2** (individual-specific errors): The individual specific errors $\varepsilon_{it}$ and $\mathbf{v}_{jt}$ are distributed independently for all $i, j$ and $t$. For each $i$, $\varepsilon_{it}$ is serially uncorrelated with mean zero, a finite variance $\sigma_i^2 < K$, and a finite fourth-order cumulant. $\mathbf{v}_{it}$ follows a linear stationary process with absolute summable autocovariances given by

$$\mathbf{v}_{it} = \sum_{\ell=0}^{\infty} \mathbf{S}_{i\ell} \boldsymbol{\nu}_{i,t-\ell}, \tag{2.8}$$

where $\boldsymbol{\nu}_{it}$ are $k \times 1$ vectors of identically, independently distributed (*IID*) random variables with mean zero, the variance matrix, $\mathbf{I}_k$, and finite fourth-order cumulants. In particular, the $k \times k$ coefficient matrices $\mathbf{S}_{i\ell}$ satisfy the condition

$$Var\left(\mathbf{v}_{it}\right) = \sum_{\ell=0}^{\infty} \mathbf{S}_{i\ell} \mathbf{S}_{i\ell}' = \Sigma_i \leq \mathbf{K} < \infty, \tag{2.9}$$

for all $i$ and some constant matrix $\mathbf{K}$, where $\Sigma_i$ is a positive definite matrix.

**Assumption 3** (factor loadings): The unobserved factor loadings, $\boldsymbol{\gamma}_i$ and $\Gamma_i$, are independently and identically distributed across $i$, and of the individual specific errors, $\varepsilon_{jt}$ and $\mathbf{v}_{jt}$, the common factors, $\mathbf{g}_t = (\mathbf{d}_t', \mathbf{f}_t')$, for all $i, j$ and $t$ with fixed means $\boldsymbol{\gamma}$ and $\Gamma$, respectively, and finite variances. In particular,

$$\boldsymbol{\gamma}_i = \boldsymbol{\gamma} + \boldsymbol{\eta}_i, \;\; \boldsymbol{\eta}_i \backsim IID\left(\mathbf{0}, \Omega_{\boldsymbol{\eta}}\right), \text{ for } i = 1, 2, ..., N, \tag{2.10}$$

where $\Omega_{\boldsymbol{\eta}}$ is a $m \times m$ symmetric non-negative definite matrix, and $\|\boldsymbol{\gamma}\| < K$, $\|\Gamma\| < K$, and $\|\Omega_{\boldsymbol{\eta}}\| < K$.

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

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{\upsilon}_i, \;\; \boldsymbol{\upsilon}_i \backsim IID\left(\mathbf{0}, \Omega_{\boldsymbol{\upsilon}}\right), \text{ for } i = 1, 2, ..., N, \tag{2.11}$$

where $\|\boldsymbol{\beta}\| < K$, $\|\Omega_{\boldsymbol{\upsilon}}\| < K$, $\Omega_{\boldsymbol{\upsilon}}$ is a $k \times k$ symmetric non-negative definite matrix, and the random deviations, $\boldsymbol{\upsilon}_i$, are distributed independently of $\boldsymbol{\gamma}_j, \Gamma_j, \varepsilon_{jt}$, $\mathbf{v}_{jt}$, and $\mathbf{g}_t$ for all $i$, $j$ and $t$.

**Assumption 5**: (identification of $\boldsymbol{\beta}_i$ and $\boldsymbol{\beta}$): Consider the cross section averages of the individual specific variables, $\mathbf{z}_{it}$, defined by $\bar{\mathbf{z}}_{wt} = \sum_{j=1}^{N} w_j \mathbf{z}_{jt}$, with the weights $\{w_j\}$ satisfying the conditions[5]

$$\text{(i): } w_i = O\left(\frac{1}{N}\right), \text{ (ii): } \sum_{i=1}^{N} |w_i| < K, \tag{2.12}$$

and let

$$\bar{\mathbf{M}}_w = \mathbf{I}_T - \bar{\mathbf{H}}_w \left(\bar{\mathbf{H}}_w' \bar{\mathbf{H}}_w\right)^{-} \bar{\mathbf{H}}_w', \tag{2.13}$$

and

$$\mathbf{M}_g = \mathbf{I}_T - \mathbf{G}\left(\mathbf{G}'\mathbf{G}\right)^{-} \mathbf{G}', \tag{2.14}$$

where $\bar{\mathbf{H}}_w = (\mathbf{D}, \bar{\mathbf{Z}}_w)$, $\mathbf{G} = (\mathbf{D}, \mathbf{F})$, $\mathbf{D} = (\mathbf{d}_1, \mathbf{d}_2, ..., \mathbf{d}_T)'$, $\mathbf{F} = (\mathbf{f}_1, \mathbf{f}_2, ..., \mathbf{f}_T)'$ are $T \times n$ and $T \times m$ data matrices on observed and unobserved common factors, respectively, $\bar{\mathbf{Z}}_w = (\bar{\mathbf{z}}_{w1}, \bar{\mathbf{z}}_{w2}, ..., \bar{\mathbf{z}}_{wT})'$ is the $T \times (k+1)$ matrix of observations on the cross section averages, and $\left(\bar{\mathbf{H}}_w' \bar{\mathbf{H}}_w\right)^{-}$ and $(\mathbf{G}'\mathbf{G})^{-}$ denote the generalized inverses of $\bar{\mathbf{H}}_w' \bar{\mathbf{H}}_w$ and $\mathbf{G}'\mathbf{G}$, respectively. Also denote the $T \times k$ observation matrix on individual specific regressors by $\mathbf{X}_i = (\mathbf{x}_{i1}, \mathbf{x}_{i2}, ..., \mathbf{x}_{iT})'$.

5a: (identification of $\boldsymbol{\beta}_i$): The $k \times k$ matrices $\hat{\Psi}_{iT} = T^{-1}\left(\mathbf{X}_i' \bar{\mathbf{M}}_w \mathbf{X}_i\right)$ and $\Psi_{ig} = T^{-1}\left(\mathbf{X}_i' \mathbf{M}_g \mathbf{X}_i\right)$ are non-singular and $\hat{\Psi}_{iT}^{-1}$ and $\Psi_{ig}^{-1}$ have finite second order moments, for all $i$.

5b: (identification of $\boldsymbol{\beta}$): The $k \times k$ pooled observation matrix $\hat{\Psi}_{NT}$ defined by

$$\hat{\Psi}_{NT} = \sum_{i=1}^{N} \theta_i \left(\frac{\mathbf{X}_i' \bar{\mathbf{M}}_w \mathbf{X}_i}{T}\right) \tag{2.15}$$

is non-singular for the scaler weights, $\theta_i$ satisfying the conditions

$$\text{(i): } \theta_i = O\left(\frac{1}{N}\right), \text{ (ii): } \sum_{i=1}^{N} |\theta_i| < K. \tag{2.16}$$

**Remark 2.1** *The residual factor model specified by (2.1), (2.2) and (2.3) is quite general and allows the unobserved common factors, $\mathbf{f}_t$, to be correlated with the individual specific regressors, $\mathbf{x}_{it}$, and permits a general degree of error cross section dependence by considering a multifactor structure with differential factor loadings over the cross section units.*

**Remark 2.2** *In addition to intercepts, seasonal dummies, and observed stationary variables such as asset returns or oil price changes, it is also possible to include deterministic trends in $\mathbf{d}_t$, by suitable scaling of the trend variables.. For example, to include a linear deterministic trend in the model, one of the elements of $\mathbf{d}_t$, say its $s^{th}$ element could be specified as $d_{st} = t/T$, with appropriate*

[5]Note that the conditions in (2.12) also imply that $\sum_{i=1}^{N} w_i^2 = O\left(N^{-1}\right)$.

*adjustments to the rate of convergence of the CCE estimator of the associated trend coefficient. The main results of the paper also hold if there are unit root processes amongst the elements of* $\mathbf{d}_t$ *and/or* $\mathbf{f}_t$*, which in turn would introduce unit roots in the individual specific regressors,* $\mathbf{x}_{it}$*. The technical details of this case can be found in Kapetanios, Pesaran and Yamagata (2004), which is currently under preparation.*

**Remark 2.3** *The weights,* $w_i$*, are not unique and, as it turns out, do not affect the asymptotic distribution of the estimators advanced in this paper. In small samples, however, they might be important, a topic which we do not address here. In practice, when* $N$ *is reasonably large one could use the equal weights* $w_i = 1/N$*. Otherwise, measures of economic distance such as output shares or trade weights could be considered, as in Pesaran, Schuermann and Weiner (2004), for example.*

**Remark 2.4** *The number of observed factors,* $n$*, and the number of individual specific regressors,* $k$*, are assumed fixed and known. The number of unobserved factors,* $m$*, is also assumed fixed, but need not be known.*

**Remark 2.5** *Finally, it is worth noting that the common feature dynamics across* $i$ *are captured through the serial correlation structure of the common effects. The assumption that the idiosyncractic errors,* $\varepsilon_{it}$*, are serially uncorrelated can also be relaxed, although in this case the CCE type estimators proposed in the paper continue to be consistent, but will no longer be efficient. Other more general individual specific dynamics can be introduced by relaxing Assumptions 1 and 2 so that lagged values of* $y_{it}$ *can also be included amongst* $\mathbf{x}_{it}$*. However, this is beyond the scope of the present paper.*

# 3 The Principal Components Estimator

To deal with the residual cross section dependence, Coakley, Fuertes and Smith (2002), hereafter referred to as CFS, propose a principal components estimator by augmenting the regression of $y_{it}$ on $\mathbf{x}_{it}$ with one or more principal components of the estimated OLS residuals, $\hat{e}_{it}$, $i = 1, 2, ..., N$, $t = 1, 2, ..., T$ obtained from the first stage regression of $y_{it}$ on $\mathbf{x}_{it}$ for each $i$. By means of a simple example we shall demonstrate that the CFS's estimator will not be consistent, unless $f_t$ and $\bar{\mathbf{x}}_t$ (the simple cross section average of $x_{it}$) are uncorrelated or if they are perfectly correlated.

For this purpose we shall focus on the simple case of only one individual-specific regressor ($k = 1$) and assume that all the coefficients of the underlying data generating process are homogeneous across $i$, namely $\alpha_i = 0$, $\beta_i = \beta$, $\gamma_i = \gamma$, and $\sigma_i^2 = \sigma^2$. This is the set up considered by CFS in the analytical discussion of their estimator. In this case the first principal component is given by

$e_t = N^{-1}\sum_{i=1}^{N} e_{it}$. CFS suggest estimating $e_t$ using the pooled estimator of $\beta$, given by

$$\hat{\beta}_{PE} = \frac{\sum_{t=1}^{T}\sum_{i=1}^{N} y_{it}x_{it}}{\sum_{t=1}^{T}\sum_{i=1}^{N} x_{it}^2}. \tag{3.1}$$

This yields $\hat{e}_t = N^{-1}\sum_{i=1}^{N}(y_{it} - \hat{\beta}_{PE}x_{it}) = \bar{y}_t - \hat{\beta}_{PE}\ \bar{x}_t$, for $t = 1, 2, ..., T$ which are then used in the augmented OLS regression of $y_{it}$on $x_{it}$ and $\hat{e}_t$ to obtain the principal components estimate of $\beta$, which we denote by $\hat{\beta}_{PC}$.

To examine the asymptotic properties of $\hat{\beta}_{PC}$ as $T$ and $N \to \infty$, using the following vector notations:

$$\begin{aligned}
\mathbf{y}_i &= (y_{i1}, y_{i2}, ..., y_{iT})',\ \mathbf{x}_i = (x_{i1}, x_{i2}, ..., x_{iT})',\ \boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \varepsilon_{i2}, ..., \varepsilon_{iT})' \\
\bar{\mathbf{y}} &= (\bar{y}_1, \bar{y}_2, ..., \bar{y}_T)',\ \bar{\mathbf{x}} = (\bar{x}_1, \bar{x}_2, ..., \bar{x}_T)',\ \bar{\boldsymbol{\varepsilon}} = (\bar{\varepsilon}_1, \bar{\varepsilon}_2, ..., \bar{\varepsilon}_T)' \\
\hat{\mathbf{e}} &= (\hat{e}_1, \hat{e}_2, ..., \hat{e}_T)',\ \mathbf{f} = (f_1, f_2, ...., f_T)',
\end{aligned}$$

we first note that

$$\hat{\beta}_{PC} = \frac{N^{-1}\sum_{i=1}^{N}(\frac{\mathbf{x}_i'\mathbf{y}_i}{T}) - (\frac{\bar{\mathbf{x}}'\hat{\mathbf{e}}}{T})(\frac{\hat{\mathbf{e}}'\hat{\mathbf{e}}}{T})^{-1}(\frac{\hat{\mathbf{e}}'\bar{\mathbf{y}}}{T})}{D_{NT}}, \tag{3.2}$$

where

$$D_{NT} = N^{-1}\sum_{i=1}^{N}(\frac{\mathbf{x}_i'\mathbf{x}_i}{T}) - (\frac{\bar{\mathbf{x}}'\hat{\mathbf{e}}}{T})(\frac{\hat{\mathbf{e}}'\hat{\mathbf{e}}}{T})^{-1}(\frac{\hat{\mathbf{e}}'\bar{\mathbf{x}}}{T})$$

In the present simple case, $\mathbf{y}_i = \beta\mathbf{x}_i + \gamma\mathbf{f} + \boldsymbol{\varepsilon}_i$, and averaging across $i$, $\bar{\mathbf{y}} = \beta\bar{\mathbf{x}} + \gamma\mathbf{f} + \bar{\boldsymbol{\varepsilon}}$. Using these in (3.2) we obtain

$$\hat{\beta}_{PC} - \beta = \gamma\frac{(\frac{\bar{\mathbf{x}}'\mathbf{f}}{T}) - (\frac{\bar{\mathbf{x}}'\hat{\mathbf{e}}}{T})(\frac{\hat{\mathbf{e}}'\hat{\mathbf{e}}}{T})^{-1}(\frac{\hat{\mathbf{e}}'\mathbf{f}}{T})}{D_{NT}} + \frac{N^{-1}\sum_{i=1}^{N}(\frac{\mathbf{x}_i'\boldsymbol{\varepsilon}_i}{T}) - (\frac{\bar{\mathbf{x}}'\hat{\mathbf{e}}}{T})(\frac{\hat{\mathbf{e}}'\hat{\mathbf{e}}}{T})^{-1}(\frac{\hat{\mathbf{e}}'\bar{\boldsymbol{\varepsilon}}}{T})}{D_{NT}}. \tag{3.3}$$

To derive the probability limit of $\hat{\beta}_{PC}$, as $N$ and $T \to \infty$, we first note that

$$\begin{aligned}
\frac{\hat{\mathbf{e}}'\bar{\boldsymbol{\varepsilon}}}{T} &= (\beta - \hat{\beta}_{PE})(\frac{\bar{\mathbf{x}}'\bar{\boldsymbol{\varepsilon}}}{T}) + \boldsymbol{\gamma}(\frac{\mathbf{f}'\bar{\boldsymbol{\varepsilon}}}{T}) + (\frac{\bar{\boldsymbol{\varepsilon}}'\bar{\boldsymbol{\varepsilon}}}{T}), \\
\frac{\hat{\mathbf{e}}'\bar{\mathbf{x}}}{T} &= (\beta - \hat{\beta}_{PE})(\frac{\bar{\mathbf{x}}'\bar{\mathbf{x}}}{T}) + \boldsymbol{\gamma}(\frac{\bar{\mathbf{x}}'\mathbf{f}}{T}) + (\frac{\bar{\mathbf{x}}'\bar{\boldsymbol{\varepsilon}}}{T}), \\
\frac{\hat{\mathbf{e}}'\mathbf{f}}{T} &= (\beta - \hat{\beta}_{PE})(\frac{\bar{\mathbf{x}}'\mathbf{f}}{T}) + \boldsymbol{\gamma}(\frac{\mathbf{f}'\mathbf{f}}{T}) + (\frac{\mathbf{f}'\bar{\boldsymbol{\varepsilon}}}{T}),
\end{aligned}$$

and finally

$$\begin{aligned}
\frac{\hat{\mathbf{e}}'\hat{\mathbf{e}}}{T} &= (\beta - \hat{\beta}_{PE})^2(\frac{\bar{\mathbf{x}}'\bar{\mathbf{x}}}{T}) + 2\gamma(\beta - \hat{\beta}_{PE})(\frac{\bar{\mathbf{x}}'\mathbf{f}}{T}) + \boldsymbol{\gamma}^2(\frac{\mathbf{f}'\mathbf{f}}{T}) \\
&\quad + (\frac{\bar{\boldsymbol{\varepsilon}}'\bar{\boldsymbol{\varepsilon}}}{T}) + 2\gamma(\beta - \hat{\beta}_{PE})(\frac{\bar{\mathbf{x}}'\bar{\boldsymbol{\varepsilon}}}{T}) + 2\gamma(\frac{\mathbf{f}'\bar{\boldsymbol{\varepsilon}}}{T}).
\end{aligned}$$

Under CFS's assumptions $T^{-1}\bar{\boldsymbol{\varepsilon}}'\bar{\boldsymbol{\varepsilon}}$, $T^{-1}\bar{\mathbf{x}}'\bar{\boldsymbol{\varepsilon}}$, $T^{-1}\mathbf{f}'\bar{\boldsymbol{\varepsilon}}$ and $N^{-1}\sum_{i=1}^{N}T^{-1}\mathbf{x}_i'\boldsymbol{\varepsilon}_i$ all converge to zero in probability as $N$ and $T \to \infty$ (in no particular order) and the following probability limits exist and are bounded

$$(\frac{\bar{\mathbf{x}}'\bar{\mathbf{x}}}{T}) \xrightarrow{p} \sigma_{\bar{x}}^2 \geq 0,\ (\frac{\bar{\mathbf{x}}'\mathbf{f}}{T}) \xrightarrow{p} \sigma_{\bar{x}f},\ (\frac{\mathbf{f}'\mathbf{f}}{T}) \xrightarrow{p} \sigma_f^2 > 0,$$

and

$$\frac{1}{N}\sum_{i=1}^{N}(\frac{\mathbf{x}_i'\mathbf{x}_i}{T}) \xrightarrow{p} \lim_{N\to\infty}\left(\frac{1}{N}\sum_{i=1}^{N}\sigma_{ix}^2\right) = \sigma_x^2 > 0.$$

Also using (3.1)

$$\beta - \hat{\beta}_{PE} \xrightarrow{p} -\gamma\left(\frac{\sigma_{\bar{x}f}}{\sigma_x^2}\right).$$

Substituting these probability limits in (3.3) and after some algebra we have

$$\hat{\beta}_{PC} - \beta \xrightarrow{p} \frac{\gamma\left(\sigma_{\bar{x}f}/\sigma_x^2\right)\left(\sigma_f^2\sigma_{\bar{x}}^2 - \sigma_{\bar{x}f}^2\right)}{\sigma_x^2\sigma_f^2 - \sigma_{\bar{x}f}^2\left[\sigma_{\bar{x}}^4/\sigma_x^4 - 3\sigma_{\bar{x}}^2/\sigma_x^2 + 3\right]}. \tag{3.4}$$

Therefore, in the presence of common effects ($\gamma \neq 0$) the CFS's principal components estimator is consistent only under the two extremes of zero correlation between the common factor and the cross-section average of the included regressor, namely if $\sigma_{\bar{x}f} = 0$, or when the common factor and the cross section average of the included regressor are perfectly correlated, namely $\sigma_{\bar{x}f}^2 = \sigma_f^2\sigma_{\bar{x}}^2$. This result also explains CFS's Monte Carlo simulations and the small sample evidence that they seem to provide in support of their proposed estimator. The processes used to generate $f_t$ and $x_{it}$ are given by

$$\begin{aligned} f_t &= 0.9\ f_{t-1} + \varepsilon_{ft}, \\ x_{it} &= \lambda_i\ f_t + \mathrm{v}_{it}, \\ \mathrm{v}_{it} &= 0.9\ \mathrm{v}_{i,t-1} + \varepsilon_{\mathrm{v}i,t}, \end{aligned}$$

and the shocks $\varepsilon_{ft}$ and $\varepsilon_{\mathrm{v}i,t}$ are $IID$ draws from the normal distribution. It is now easily seen that

$$\bar{x}_t = \bar{\lambda}f_t + \bar{\mathrm{v}}_t,$$

where $\bar{\mathrm{v}}_t$ and $\bar{\lambda}$ are the cross section means of $d_{it}$ and $\lambda_i$, respectively. Also

$$\bar{\mathrm{v}}_t = 0.9\ \bar{\mathrm{v}}_{t-1} + \bar{\varepsilon}_{dt},$$

and since the shocks, $\varepsilon_{\mathrm{v}i,t}$, are $IID$ it then readily follows that $Var(\bar{\varepsilon}_{\mathrm{v}t}) \to 0$ and hence $Var(\bar{\mathrm{v}}_t) \to 0$ for each $t$ as $N \to \infty$. Therefore, $\bar{x}_t$ and $f_t$ will become perfectly correlated if $N$ is sufficiently large.

# 4 A General Approach to Estimation of Panels with Common Effects

The main difficulty with the CFS's estimator lies in the fact that it makes use of an inconsistent estimator of $\boldsymbol{\beta}_i$ to obtain the principal components which are then used as proxies for the unobserved common effects. One way of overcoming this problem would be to estimate $\boldsymbol{\beta}_i$ directly, using suitable proxies for the unobserved factors that do not depend on an initial estimate of $\boldsymbol{\beta}_i$. To see how this can be done consider the cross section averages of the equations in (2.4), using the weights $w_j$:[6]

$$\bar{\mathbf{z}}_{wt} = \bar{\mathbf{B}}'_w \mathbf{d}_t + \bar{\mathbf{C}}'_w \mathbf{f}_t + \bar{\mathbf{u}}_{wt}, \tag{4.1}$$

where as before, $\bar{\mathbf{z}}_{wt} = \sum_{j=1}^{N} w_j \mathbf{z}_{jt}$ and

$$\bar{\mathbf{B}}_w = \sum_{i=1}^{N} w_i \mathbf{B}_i,\ \bar{\mathbf{C}}_w = \sum_{i=1}^{N} w_i \mathbf{C}_i,\ \bar{\mathbf{u}}_{wt} = \sum_{i=1}^{N} w_i \mathbf{u}_{it}, \tag{4.2}$$

and suppose that

$$Rank(\bar{\mathbf{C}}_w) = m \leq k+1, \text{ for all } N. \tag{4.3}$$

Then we have

$$\mathbf{f}_t = \left(\bar{\mathbf{C}}_w \bar{\mathbf{C}}'_w\right)^{-1} \bar{\mathbf{C}}_w \left(\bar{\mathbf{z}}_{wt} - \bar{\mathbf{B}}'_w \mathbf{d}_t - \bar{\mathbf{u}}_{wt}\right). \tag{4.4}$$

But using Lemma A.1 in Appendix A, we have

$$\bar{\mathbf{u}}_{wt} \overset{q.m.}{\to} \mathbf{0}, \text{ as } N \to \infty, \text{ for each } t, \tag{4.5}$$

and

$$\bar{\mathbf{C}}_w \xrightarrow{p} \mathbf{C} = \tilde{\Gamma} \begin{pmatrix} 1 & \mathbf{0} \\ \boldsymbol{\beta} & \mathbf{I}_k \end{pmatrix}, \text{ as } N \to \infty, \tag{4.6}$$

where

$$\tilde{\Gamma} = \left(E\left(\boldsymbol{\gamma}_i\right), E\left(\Gamma_i\right)\right) = \left(\boldsymbol{\gamma}, \Gamma\right). \tag{4.7}$$

[6] In principle the weights used in the construction of the aggregates, $\bar{\mathbf{z}}_{wt}$, could be individual-specific, namely for individual $i$ one could use $\bar{\mathbf{z}}_{w_i t} = \sum_{j=1}^{N} w_{ij} \mathbf{z}_{jt}$, with $w_{ii} = 0$. As we shall see later in small samples the optimal choice of these weights will depend on the unknown parameters, $\gamma_j$ and $\sigma_j^2$, $j = 1, 2, ..., N$. But for consistent estimation it is only required that the chosen weights satisfy the conditions in (2.12), in particular that for each $i$, $\sum_{j=1}^{N} w_{ij}^2 \to 0$ as $N \to \infty$.

Therefore, assuming that $\text{Rank}(\tilde{\Gamma}) = m$ we obtain

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

This suggests using $\bar{\mathbf{h}}_{wt} = (\mathbf{d}_t', \bar{\mathbf{z}}_{wt}')'$ as observable proxies for $\mathbf{f}_t$. Whilst consistent estimation of $\mathbf{f}_t$ using the above results still requires knowledge of the underlying parameters, the individual slope coefficients of interest, $\boldsymbol{\beta}_i$ and their means, $\boldsymbol{\beta}$, can be consistently estimated by augmenting the OLS or pooled regressions of $y_{it}$ on $\mathbf{x}_{it}$ with $\mathbf{d}_t$ and the cross section averages, $\bar{\mathbf{z}}_{wt}$. We shall refer to such estimators as the "common correlated effect estimator" (CCE). As we shall see later the basic idea of augmenting the regressions with cross section averages continues to work even if the rank condition, (4.3), is not satisfied. Rank deficiency in $\mathbf{C}$ induces exact linear dependencies amongst the elements of $\bar{\mathbf{h}}_{wt}$, as $N \to \infty$. For example, in the extreme case where $\mathbf{C} = \mathbf{0}$, using (4.1), we have

$$\bar{\mathbf{z}}_{wt} - \bar{\mathbf{B}}_w'\mathbf{d}_t \overset{q.m.}{\to} \mathbf{0}, \text{ as } N \to \infty,$$

and a full augmentation of regressions of $y_{it}$ on $\mathbf{x}_{it}$ with all the elements of $\bar{\mathbf{h}}_{wt}$ would not be necessary. But augmenting the individual regressions with $\bar{\mathbf{h}}_{wt}$ would still be effective in reducing residual cross section correlations, even though in this case the elements of $\bar{\mathbf{h}}_{wt}$ will be perfectly correlated as $N \to \infty$. But as we shall show the CCE estimators of $\boldsymbol{\beta}$ are not affected by rank deficiency problem and continue to be asymptotically invariant to the factor loadings, $\boldsymbol{\gamma}_i$, for any fixed $m$.

# 5 Common Correlated Effects Estimators: Individual Specific Coefficients

For the individual slope coefficients the CCE is given by

$$\hat{\mathbf{b}}_i = (\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{X}_i)^{-1}\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{y}_i, \tag{5.8}$$

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

$$\bar{\mathbf{M}}_w = \mathbf{I}_T - \bar{\mathbf{H}}_w\left(\bar{\mathbf{H}}_w'\bar{\mathbf{H}}_w\right)^{-1}\bar{\mathbf{H}}_w', \tag{5.9}$$

and as before $\bar{\mathbf{H}}_w = (\mathbf{D}, \bar{\mathbf{Z}}_w)$, $\mathbf{D}$ and $\bar{\mathbf{Z}}_w$ being, respectively, the $T \times n$ and $T \times (k+1)$ matrices of observations on $\mathbf{d}_t$ and $\bar{\mathbf{z}}_{wt}$. The rank condition, $\text{Rank}(\tilde{\Gamma}) = m$, ensures that under Assumptions 1-4, $T^{-1}\left(\bar{\mathbf{H}}_w'\bar{\mathbf{H}}_w\right)$ converges to a positive definite matrix, for a fixed $T$ as $N \to \infty$, as well as when $(N,T) \xrightarrow{j} \infty$. But $T^{-1}(\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{X}_i)$ and its limit as $(N,T) \xrightarrow{j} \infty$ exits even if the rank condition is not satisfied. This is because $T^{-1}(\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{X}_i)$ is invariant to the choice of a $g$-inverse for $\bar{\mathbf{H}}_w'\bar{\mathbf{H}}_w$, and as we shall see its limit under $(N,T) \xrightarrow{j} \infty$ will be positive definite so long as $\Sigma_i$, is positive definite.

For each $i$ and $t = 1, 2, ..., T$, writing (2.1) and (2.2) in matrix notations we have

$$\mathbf{y}_i = \mathbf{D}\boldsymbol{\alpha}_i + \mathbf{X}_i\boldsymbol{\beta}_i + \mathbf{F}\boldsymbol{\gamma}_i + \boldsymbol{\varepsilon}_i, \tag{5.10}$$

where $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \varepsilon_{i2}, ..., \varepsilon_{iT})'$, and as set out in Assumption 5, $\mathbf{D} = (\mathbf{d}_1, \mathbf{d}_2, ..., \mathbf{d}_T)'$ and $\mathbf{F} = (\mathbf{f}_1, \mathbf{f}_2, ..., \mathbf{f}_T)'$. Using (5.10) in (5.8) we have

$$\hat{\mathbf{b}}_i - \boldsymbol{\beta}_i = \left(\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{X}_i}{T}\right)^{-1}\left(\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{F}}{T}\right)\boldsymbol{\gamma}_i + \left(\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{X}_i}{T}\right)^{-1}\left(\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\boldsymbol{\varepsilon}_i}{T}\right), \tag{5.11}$$

which shows the direct dependence of $\hat{\mathbf{b}}_i$ on the unobserved factors through $T^{-1}\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{F}$. To examine the properties of this component, writing (2.3) and (4.1) in matrix notations, we first note that

$$\mathbf{X}_i = \mathbf{G}\Pi_i + \mathbf{V}_i, \tag{5.12}$$

and

$$\bar{\mathbf{H}}_w = \mathbf{G}\bar{\mathbf{P}}_w + \bar{\mathbf{U}}_w^*, \tag{5.13}$$

where $\mathbf{G} = (\mathbf{D}, \mathbf{F})$, $\Pi_i = (\mathbf{A}_i', \Gamma_i')'$, $\mathbf{V}_i = (\mathbf{v}_{i1}, \mathbf{v}_{i2}, ..., \mathbf{v}_{iT})'$,

$$\underset{(n+m)\times(n+k+1)}{\bar{\mathbf{P}}_w} = \begin{pmatrix} \mathbf{I}_n & \bar{\mathbf{B}}_w \\ \mathbf{0} & \bar{\mathbf{C}}_w \end{pmatrix}, \ \bar{\mathbf{U}}_w^* = (\mathbf{0}, \bar{\mathbf{U}}_w), \tag{5.14}$$

$\bar{\mathbf{U}}_w = (\bar{\mathbf{u}}_{w1}, \bar{\mathbf{u}}_{w2}, ..., \bar{\mathbf{u}}_{wT})'$. Also

$$\left\|\bar{\mathbf{B}}_w\right\| = \sum_{i=1}^{N}|w_i|\left\|\mathbf{B}_i\right\| < K, \text{ and }, \ \left\|\bar{\mathbf{C}}_w\right\| = \sum_{i=1}^{N}|w_i|\left\|\mathbf{C}_i\right\| < K, \tag{5.15}$$

under (2.12) and noting that $\|\mathbf{B}_i\|$ and $\|\mathbf{C}_i\|$ are bounded. Furthermore, under Assumptions 1 and 2, $(\mathbf{G}, \mathbf{V}_i)$ is covariance stationary and

$$\begin{aligned} \frac{\mathbf{X}_i'\mathbf{G}}{T} &= \Pi_i'\left(\frac{\mathbf{G}'\mathbf{G}}{T}\right) + \frac{\mathbf{V}_i'\mathbf{G}}{T} = O_p(1), \\ \frac{\mathbf{G}'\mathbf{G}}{T} &= O_p(1), \ \frac{\mathbf{G}'\mathbf{F}}{T} = O_p(1). \end{aligned}$$

Using results in Lemmas A.2 and A.3, it is now easily seen that

$$\frac{\mathbf{X}_i'\bar{\mathbf{H}}_w}{T} = \left(\frac{\mathbf{X}_i'\mathbf{G}}{T}\right)\bar{\mathbf{P}}_w + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right), \tag{5.16}$$

$$\frac{\bar{\mathbf{H}}_w'\bar{\mathbf{H}}_w}{T} = \bar{\mathbf{P}}_w'\left(\frac{\mathbf{G}'\mathbf{G}}{T}\right)\bar{\mathbf{P}}_w + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right), \tag{5.17}$$

$$\frac{\bar{\mathbf{H}}_w'\mathbf{F}}{T} = \bar{\mathbf{P}}_w'\left(\frac{\mathbf{G}'\mathbf{F}}{T}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right), \tag{5.18}$$

Hence, we obtain the following result which is critical to many of the derivations in this paper and does not require the rank condition (4.3):

$$\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{F}}{T} = \frac{\mathbf{X}_i'\bar{\mathbf{M}}_q\mathbf{F}}{T} + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right), \tag{5.19}$$

where

$$\bar{\mathbf{M}}_q = \mathbf{I}_T - \bar{\mathbf{Q}}_w\left(\bar{\mathbf{Q}}_w'\bar{\mathbf{Q}}_w\right)^{-}\bar{\mathbf{Q}}_w', \text{ with } \bar{\mathbf{Q}}_w = \mathbf{G}\bar{\mathbf{P}}_w. \tag{5.20}$$

When the rank condition (4.3) is satisfied, using familiar results on generalized inverse, we have

$$\bar{\mathbf{M}}_q = \mathbf{M}_g = \mathbf{I}_T - \mathbf{G}\left(\mathbf{G}'\mathbf{G}\right)^{-}\mathbf{G}',$$

and since $\mathbf{F} \subset \mathbf{G}$ then $\bar{\mathbf{M}}_q\mathbf{F} = \mathbf{M}_g\mathbf{F} = \mathbf{0}$, and

$$\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{F}}{T} = O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right). \tag{5.21}$$

If the rank condition is not satisfied, we still have $\mathbf{X}_i'\bar{\mathbf{M}}_q\bar{\mathbf{Q}}_w = \mathbf{0}$, and since $\bar{\mathbf{Q}}_w = \mathbf{G}\bar{\mathbf{P}}_w = (\mathbf{D}, \mathbf{D}\bar{\mathbf{B}}_w + \mathbf{F}\bar{\mathbf{C}}_w)$, it follows that

$$\left(\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{F}}{T}\right)\bar{\mathbf{C}}_w = O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right). \tag{5.22}$$

Also, using (2.6) and (2.11) we have

$$\bar{\mathbf{C}}_w = \left(\bar{\boldsymbol{\gamma}}_w + \bar{\Gamma}_w\boldsymbol{\beta} + \sum_{i=1}^{N} w_i\Gamma_i\boldsymbol{\upsilon}_i,\ \bar{\Gamma}_w\right),$$

where $\bar{\Gamma}_w = \sum_{i=1}^{N} w_i\Gamma_i$. Substituting this result in (5.22) now yields

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

which in turn leads to

$$\frac{\sqrt{N}\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{F}}{T}\left(\bar{\boldsymbol{\gamma}}_w + \sum_{i=1}^{N} w_i\Gamma_i\boldsymbol{\upsilon}_i\right) = O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right).$$

But under Assumption 4 and (2.12), $\sum_{i=1}^{N} w_i\Gamma_i\boldsymbol{\upsilon}_i = O_p\left(N^{-1/2}\right)$, and therefore

$$\frac{\sqrt{N}\left(\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{F}\right)\bar{\boldsymbol{\gamma}}_w}{T} = O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right). \tag{5.23}$$

This result is clearly implied by (5.21), irrespective of whether the factor loadings are random or just bounded. But the reverse is not true; (5.23) does not imply (5.21) if the rank condition is not satisfied.

Similarly, irrespective of the rank of $\bar{\mathbf{C}}_w$, it can be established that

$$\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{X}_i}{T} = \frac{\mathbf{X}_i'\bar{\mathbf{M}}_q\mathbf{X}_i}{T} + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right), \tag{5.24}$$

and

$$\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\boldsymbol{\varepsilon}_i}{T} = \frac{\mathbf{X}_i'\bar{\mathbf{M}}_q\boldsymbol{\varepsilon}_i}{T} + O_p\left(\frac{1}{N}\right). \tag{5.25}$$

When the rank condition is satisfied, however, the matrices $\mathbf{X}_i'\bar{\mathbf{M}}_q\mathbf{X}_i$ and $\mathbf{X}_i'\bar{\mathbf{M}}_q\boldsymbol{\varepsilon}_i$ would simplify to $\mathbf{X}_i'\mathbf{M}_g\mathbf{X}_i$ and $\mathbf{X}_i'\mathbf{M}_g\boldsymbol{\varepsilon}_i$, respectively.

Using the above results in (5.11) and noting that $T^{-1}\mathbf{X}_i'\bar{\mathbf{M}}_q\mathbf{X}_i = O_p(1)$, and assuming that the rank condition (4.3) is satisfied we have[7]

$$\hat{\mathbf{b}}_i - \boldsymbol{\beta}_i = \left(\frac{\mathbf{X}_i'\mathbf{M}_g\mathbf{X}_i}{T}\right)^{-1}\left(\frac{\mathbf{X}_i'\mathbf{M}_g\boldsymbol{\varepsilon}_i}{T}\right) + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right). \tag{5.26}$$

Since $\boldsymbol{\varepsilon}_i$ is independently distributed of $\mathbf{X}_i$ and $\mathbf{G} = (\mathbf{D}, \mathbf{F})$, then for a fixed $T$, and as $N \to \infty$, $E\left(\hat{\mathbf{b}}_i - \boldsymbol{\beta}_i\right) = \mathbf{0}$. The finite-$T$ distribution of $\hat{\mathbf{b}}_i - \boldsymbol{\beta}_i$ will be free of nuisance parameters as $N \to \infty$, but will depend on the probability density of $\boldsymbol{\varepsilon}_i$. For $N$ and $T$ sufficiently large, the distribution of $\sqrt{T}\left(\hat{\mathbf{b}}_i - \boldsymbol{\beta}_i\right)$ will be asymptotically normal if the rank condition (4.3) is satisfied and if $N$ and $T$ are of the same order of magnitudes, namely, if $T/N \to \kappa$ as $N$ and $T \to \infty$, where $\kappa$ is a positive finite constant. To see why this additional condition is needed, using (5.26) note that

$$\sqrt{T}\left(\hat{\mathbf{b}}_i - \boldsymbol{\beta}_i\right) = \left(\frac{\mathbf{X}_i'\mathbf{M}_g\mathbf{X}_i}{T}\right)^{-1}\frac{\mathbf{X}_i'\mathbf{M}_g\boldsymbol{\varepsilon}_i}{\sqrt{T}} + O_p\left(\frac{\sqrt{T}}{N}\right) + O_p\left(\frac{1}{\sqrt{N}}\right), \tag{5.27}$$

and the asymptotic distribution of $\sqrt{T}\left(\hat{\mathbf{b}}_i - \boldsymbol{\beta}_i\right)$ will be free of nuisance parameters only if $\sqrt{T}/N \to 0$, as $(N,T) \xrightarrow{j} \infty$. For this condition to be satisfied it is sufficient that $T/N \to \kappa$, as $(N,T) \xrightarrow{j} \infty$, where $\kappa$ is a finite non-negative constant.

The following theorem provides a formal statement of these results and the associated asymptotic distributions in the case where the rank condition is satisfied.

**Theorem 5.1** *Consider the panel data model (2.1) and (2.2) and suppose that* $\|\boldsymbol{\beta}_i\| < K$, $\|\Pi_i\| < K$, *Assumptions 1,2, and 5a hold, and the rank condition (4.3) is satisfied.*

[7]Note also that under Assumption 5a, $T^{-1}\left(\mathbf{X}_i'\mathbf{M}_g\mathbf{X}_i\right)$ is a positive definite matrix.

*(a) - (N-asymptotic) The common correlated effects estimator, $\hat{\mathbf{b}}_i$, defined by (5.8) is unbiased for a fixed $T > n + 2k + 1$ and $N \to \infty$, in the sense that $lim_{N\to\infty} E\left(\hat{\mathbf{b}}_i\right) = \boldsymbol{\beta}_i$. Under the additional assumption that $\varepsilon_{it} \sim IIDN(0, \sigma_i^2)$,*

$$\hat{\mathbf{b}}_i - \boldsymbol{\beta}_i \xrightarrow{d} N(\mathbf{0}, \Sigma_{T,b_i}), \tag{5.28}$$

*as $N \to \infty$, where*

$$\Sigma_{T,b_i} = T^{-1}\sigma_i^2\Psi_{ig}^{-1}, \quad \Psi_{ig} = T^{-1}\left(\mathbf{X}_i'\mathbf{M}_g\mathbf{X}_i\right), \tag{5.29}$$

$$\mathbf{M}_g = \mathbf{I}_T - \mathbf{G}(\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}', \tag{5.30}$$

*and $\mathbf{G} = (\mathbf{g}_1, \mathbf{g}_2, ..., \mathbf{g}_T) = (\mathbf{F}, \mathbf{D})$.*

*(b) - (Joint asymptotic) As $(N, T) \xrightarrow{j} \infty$ (in no particular order), $\hat{\mathbf{b}}_i$ is a consistent estimator of $\boldsymbol{\beta}_i$. If it is further assumed that $\sqrt{T}/N \to 0$ as $(N, T) \xrightarrow{j} \infty$, then*

$$\sqrt{T}\left(\hat{\mathbf{b}}_i - \boldsymbol{\beta}_i\right) \xrightarrow{d} N(\mathbf{0}, \Sigma_{b_i}), \tag{5.31}$$

*where*

$$\Sigma_{b_i} = \sigma_i^2\Sigma_i^{-1}. \tag{5.32}$$

An asymptotically unbiased estimator of $\Sigma_{T,b_i}$, as $N \to \infty$ for a fixed $T > n + 2k + 1$, is given by (See Appendix B for a proof):

$$\hat{\Sigma}_{T,b_i} = \hat{\sigma}_i^2\left(\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{X}_i\right)^{-1}, \tag{5.33}$$

where

$$\hat{\sigma}_i^2 = \frac{\left(\mathbf{y}_i - \mathbf{X}_i\hat{\mathbf{b}}_i\right)'\bar{\mathbf{M}}_w\left(\mathbf{y}_i - \mathbf{X}_i\hat{\mathbf{b}}_i\right)}{T - (n + m + k)}. \tag{5.34}$$

In the case where $(N, T) \xrightarrow{j} \infty$, a consistent estimator of $\Sigma_{b_i}$ is given by

$$\hat{\Sigma}_{b_i} = \mathring{\sigma}_i^2\left(\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{X}_i}{T}\right)^{-1}, \tag{5.35}$$

where

$$\mathring{\sigma}_i^2 = \frac{\left(\mathbf{y}_i - \mathbf{X}_i\hat{\mathbf{b}}_i\right)'\bar{\mathbf{M}}_w\left(\mathbf{y}_i - \mathbf{X}_i\hat{\mathbf{b}}_i\right)}{T - (n + 2k + 1)}. \tag{5.36}$$

Here we have approximated $m$ in (5.34) by its upper bound under the rank condition (4.3), namely $k + 1$. For $T$ sufficiently large the difference between $\hat{\sigma}_i^2$ and $\mathring{\sigma}_i^2$ will be negligible, but the latter has the advantage of not requiring an *a priori* knowledge of $m$.

When the rank condition, (4.3), is not satisfied consistent estimation of the individual slope coefficients is not possible. But as we shall, the mean of $\boldsymbol{\beta}_i$ can be consistently estimated irrespective of the rank of $\bar{\mathbf{C}}_w$ under the random coefficient Assumptions 3 and 4.

# 6 Pooled Estimators

In this section we shall assume that the parameters of interest are the cross-section means of the slope coefficients $\boldsymbol{\beta}_i$, namely $\boldsymbol{\beta}$ defined by (2.11), and consider two alternative estimators, the Mean Group (MG) estimator proposed in Pesaran and Smith (1995) and a generalization of the fixed effects estimator that allow for the possibility of cross section dependence. We shall refer to the former as the "Common Correlated Effects Mean Group" (CCEMG) estimator, and the latter as the "Common Correlated Effects Pooled" (CCEP) estimator.

## 6.1 Common Correlated Effects Mean Group Estimator

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

$$\hat{\mathbf{b}}_{MG} = N^{-1} \sum_{i=1}^{N} \hat{\mathbf{b}}_i. \tag{6.37}$$

As an alternative one could also consider Swamy's Random Coefficient (RC) estimator defined by the weighted average of the individual estimates with the weights being inversely proportional to the individual variances (see, for example, Swamy (1970)):

$$\hat{\mathbf{b}}_{RC} = \sum_{i=1}^{N} \hat{\Theta}_i \hat{\mathbf{b}}_i, \tag{6.38}$$

where

$$\hat{\Theta}_i = \left\{ \sum_{j=1}^{N} \left[ \hat{\Sigma}_{T,b_j} + \hat{\Omega}_{\boldsymbol{v}} \right]^{-1} \right\}^{-1} \left[ \hat{\Sigma}_{T,b_i} + \hat{\Omega}_{\boldsymbol{v}} \right]^{-1}, \tag{6.39}$$

$\hat{\Sigma}_{T,b_j}$ is given by (5.33) and $\hat{\Omega}_{\boldsymbol{v}}$ is a consistent estimator of $\Omega_{\boldsymbol{v}}$, the variance of $\boldsymbol{v}_i$ defined by (2.11). A comparative analysis of the MG and the RC estimators in the context of dynamic panel data models without unobserved common effects is provided in Hsiao, Pesaran and Tahmiscioglu (1999). It is shown that, for $N$ and $T$ sufficiently large, both of these estimators are consistent and asymptotically equivalent. These results continue to apply in the more general setting of this paper. Here we shall focus on the MG estimator, and note that under Assumption 4 and using (5.11) we have

$$\begin{aligned} \sqrt{N}\left(\hat{\mathbf{b}}_{MG} - \boldsymbol{\beta}\right) &= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \boldsymbol{v}_i + \frac{1}{N} \sum_{i=1}^{N} \hat{\Psi}_{iT}^{-1} \left( \frac{\sqrt{N} \mathbf{X}_i' \bar{\mathbf{M}}_w \mathbf{F}}{T} \right) \boldsymbol{\gamma}_i + \\ &\quad \frac{1}{N} \sum_{i=1}^{N} \hat{\Psi}_{iT}^{-1} \left( \frac{\sqrt{N} \mathbf{X}_i' \bar{\mathbf{M}}_w \boldsymbol{\varepsilon}_i}{T} \right), \end{aligned} \tag{6.40}$$

where by assumption $\hat{\Psi}_{iT}^{-1} = \left(T^{-1}\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{X}_i\right)^{-1}$ has second order moments. In the case where the rank condition (4.3) is satisfied, using (5.21) we have

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

and it is easily seen that for all bounded values of the factor loadings, $\boldsymbol{\gamma}_i$, that

$$\frac{1}{N}\sum_{i=1}^{N}\hat{\Psi}_{iT}^{-1}\left(\frac{\sqrt{N}\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{F}}{T}\right)\boldsymbol{\gamma}_i \xrightarrow{p} \mathbf{0}, \text{ as } (N,T) \xrightarrow{j} \infty.$$

Similarly, using (5.24) and (5.25)

$$\frac{1}{N}\sum_{i=1}^{N}\hat{\Psi}_{iT}^{-1}\left(\frac{\sqrt{N}\mathbf{X}_i'\bar{\mathbf{M}}_w\boldsymbol{\varepsilon}_i}{T}\right) = \Delta_{NT} + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right),$$

where

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

However, since $\boldsymbol{\varepsilon}_i$ is distributed independently of $\mathbf{X}_i$ and $\mathbf{G}$, and by Assumption 5a, $E\left(\Psi_{ig}^{-1}\right) < K$, we have

$$Var\left(\Delta_{NT}\right) = \frac{1}{NT}\sum_{i=1}^{N}\sigma_i^2 E\left(\Psi_{ig}^{-1}\right) = O\left(\frac{1}{T}\right),$$

and

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

Hence

$$\sqrt{N}\left(\hat{\mathbf{b}}_{MG} - \boldsymbol{\beta}\right) \xrightarrow{d} N(\mathbf{0}, \Sigma_{MG}), \text{ as } (N,T) \xrightarrow{j} \infty. \tag{6.41}$$

In the present case $\Sigma_{MG} = \Omega_{\upsilon}$, and can be consistently estimated non-parametrically by

$$\hat{\Sigma}_{MG} = \frac{1}{N-1}\sum_{i=1}^{N}\left(\hat{\mathbf{b}}_i - \hat{\mathbf{b}}_{MG}\right)\left(\hat{\mathbf{b}}_i - \hat{\mathbf{b}}_{MG}\right)'. \tag{6.42}$$

It is also interesting to note that (6.41) holds even if the rank condition is not satisfied, so long as the factor loadings satisfy the random coefficient model, (2.10). In this case using (2.10) we note that the second term in (6.40) can be written as

$$\boldsymbol{\chi}_{NT} = \frac{1}{N}\sum_{i=1}^{N}\hat{\Psi}_{iT}^{-1}\left(\frac{\sqrt{N}\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{F}}{T}\right)\left(\bar{\boldsymbol{\gamma}}_w + \boldsymbol{\eta}_i - \bar{\boldsymbol{\eta}}_w\right), \tag{6.43}$$

where $\bar{\boldsymbol{\gamma}}_w = \sum_{i=1}^{N} w_i\boldsymbol{\gamma}_i$, and $\bar{\boldsymbol{\eta}}_w = \sum_{i=1}^{N} w_i\boldsymbol{\eta}_i$. Also using (5.19), (5.23), and (5.24) we have

$$\boldsymbol{\chi}_{NT} = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\frac{\mathbf{X}_i'\bar{\mathbf{M}}_q\mathbf{X}_i}{T}\right)^{-1}\left(\frac{\mathbf{X}_i'\bar{\mathbf{M}}_q\mathbf{F}}{T}\right)\left(\boldsymbol{\eta}_i - \bar{\boldsymbol{\eta}}_w\right) + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right),$$

which establishes that for $N$ and $T$ large

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

The two terms on the right hand side of the above expression are independently distributed and both tend to Normal densities with mean zero and finite variances.[8] In this case the asymptotic variance of $\sqrt{N}\left(\hat{\mathbf{b}}_{MG}-\boldsymbol{\beta}\right)$ is given by

$$\Sigma_{MG}=\Omega_{\upsilon}+\lim_{N\to\infty}\left[\frac{1}{N}\sum_{i=1}^{N}\left(\Sigma_{iq}^{-1}\mathbf{Q}_{if}\Omega_{\eta}\mathbf{Q}_{if}'\Sigma_{iq}^{-1}\right)\right], \tag{6.44}$$

where

$$\Sigma_{iq}=p\lim_{T\to\infty}\left(T^{-1}\mathbf{X}_i'\bar{\mathbf{M}}_q\mathbf{X}_i\right)\text{ and }\mathbf{Q}_{if}=p\lim_{T\to\infty}\left(T^{-1}\mathbf{X}_i'\bar{\mathbf{M}}_q\mathbf{F}\right), \tag{6.45}$$

and depends on the unobserved factors. Nevertheless, it can be consistently estimated non-parametrically using (6.42). To see this first note that

$$\hat{\mathbf{b}}_i-\boldsymbol{\beta}=\boldsymbol{\upsilon}_i+\mathbf{h}_{iT}+O_p\left(\frac{1}{\sqrt{N}}\right)+O_p\left(\frac{1}{\sqrt{T}}\right), \tag{6.46}$$

where

$$\mathbf{h}_{iT}=\left(\frac{\mathbf{X}_i'\bar{\mathbf{M}}_q\mathbf{X}_i}{T}\right)^{-1}\frac{\mathbf{X}_i'\bar{\mathbf{M}}_q\left[\mathbf{F}\left(\boldsymbol{\eta}_i-\bar{\boldsymbol{\eta}}_w\right)+\boldsymbol{\varepsilon}_i\right]}{T}, \tag{6.47}$$

and

$$\hat{\mathbf{b}}_i-\hat{\mathbf{b}}_{MG}=\left(\boldsymbol{\upsilon}_i-\bar{\boldsymbol{\upsilon}}\right)+\left(\mathbf{h}_{iT}-\bar{\mathbf{h}}_T\right)+O_p\left(\frac{1}{\sqrt{N}}\right)+O_p\left(\frac{1}{\sqrt{T}}\right). \tag{6.48}$$

Since by assumption $\boldsymbol{\upsilon}_i$ and $\mathbf{h}_{iT}$ are independently distributed across $i$, then

$$E\left[\frac{1}{N-1}\sum_{i=1}^{N}\left(\hat{\mathbf{b}}_i-\hat{\mathbf{b}}_{MG}\right)\left(\hat{\mathbf{b}}_i-\hat{\mathbf{b}}_{MG}\right)'\right]=\Sigma_{MG}+O\left(\frac{1}{\sqrt{N}}\right)+O\left(\frac{1}{\sqrt{T}}\right).$$

The above results are summarized in the following general theorem:

**Theorem 6.1** *Consider the panel data model (2.1) and (2.2) and suppose that Assumptions 1-4, and 5a hold. Then the Common Correlated Effects Mean Group estimator, $\mathbf{b}_{MG}$ defined by (6.37), is asymptotically (for a fixed $T$ and as $N\to\infty$) unbiased for $\boldsymbol{\beta}$, and as $(N,T)\overset{j}{\to}\infty$*

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

*where $\Sigma_{MG}$ is given by (6.44), which is consistently estimated by (6.42).*

[8]The latter result follows using Lemma A.4 and noting that as $T\to\infty$, $T^{-1}\mathbf{X}_i'\bar{\mathbf{M}}_q\mathbf{X}_i\overset{p}{\to}\Sigma_i$, which is a positive definite matrix by assumption.

This theorem does not require the rank condition, (4.3), holds for any number, $m$, of unobserved factors so long as $m$ is fixed, and 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 Assumption 3 can be relaxed and the factor loadings, $\boldsymbol{\gamma}_i$, need not follow the random coefficient model. It would be sufficient that they are bounded.

### 6.2 Common Correlated Effects Pooled Estimators

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. In what follows we developed a pooled estimator of $\boldsymbol{\beta}$ that assumes (possibly incorrectly) that $\boldsymbol{\beta}_i = \boldsymbol{\beta}$, and $\sigma_i^2 = \sigma^2$, although it allows the slope coefficients of the common effects (whether observed or not) to differ across $i$. Such a pooled estimator of $\boldsymbol{\beta}$, denoted by CCEP, is given by

$$\hat{\mathbf{b}}_P = \left( \sum_{i=1}^{N} \theta_i \mathbf{X}_i' \bar{\mathbf{M}}_w \mathbf{X}_i \right)^{-1} \sum_{i=1}^{N} \theta_i \mathbf{X}_i' \bar{\mathbf{M}}_w \mathbf{y}_i. \tag{6.49}$$

Typically, the (pooling) weights $\theta_i$ are set equal to $1/N$, although in the general case where $\sigma_i^2$ differ across $i$ as we shall see it will be optimal to set $\theta_i = \sigma_i^{-2} / \sum_{j=1}^{N} \sigma_j^{-2}$. However, in practice where $\sigma_i^2$ is unknown the efficiency gain from using an estimate of $\sigma_i^2$ is likely to be limited particularly when $T$ is small. In the present context it also turns out that when the rank condition (4.3) is not satisfied the pooling weights, $\theta_i$, must equal the aggregating weights, $w_i$; otherwise the CCEP estimator will not be consistent. The asymptotic results for $\hat{\mathbf{b}}_P$ is summarized in the following theorem, with proofs provided in Appendix B.

**Theorem 6.2** *Consider the panel data model (2.1) and (2.2) and suppose that Assumptions 1-4 and 5b hold, and $\theta_i = w_i$. Then the Common Correlated Effects Pooled estimator, $\hat{\mathbf{b}}_P$, defined by (6.49) is asymptotically unbiased for $\boldsymbol{\beta}$, and as $(N,T) \xrightarrow{j} \infty$ we have*

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

*where*

$$\Sigma_P^* = \Psi^{*-1} \mathbf{R}^* \Psi^{*-1}, \tag{6.50}$$

$$\Psi^* = \lim_{N \to \infty} \left( \sum_{i=1}^{N} w_i \Sigma_{iq} \right), \ \mathbf{R}^* = \lim_{N \to \infty} \left[ N^{-1} \sum_{i=1}^{N} \tilde{w}_i^2 \left( \Sigma_{iq} \Omega_\upsilon \Sigma_{iq} + \mathbf{Q}_{if} \Omega_\eta \mathbf{Q}_{if}' \right) \right], \tag{6.51}$$

$$\tilde{w}_i = \frac{w_i}{\sqrt{N^{-1} \sum_{i=1}^{N} w_i^2}}, \tag{6.52}$$

*and $\Sigma_{iq}$ and $\mathbf{Q}_{if}$ are defined by (6.45).*

Although the asymototic variance matrix of $\hat{\mathbf{b}}_P$ depends on the unobserved factors and their loadings, it is nevertheless possible to estimate it consistently along the lines similar to that followed in the case of CCEMG. Using (5.24) and (6.48) we first note that

$$\left(\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{X}_i}{T}\right)\left(\hat{\mathbf{b}}_i-\hat{\mathbf{b}}_{MG}\right)=\left(\frac{\mathbf{X}_i'\bar{\mathbf{M}}_q\mathbf{X}_i}{T}\right)(\boldsymbol{\upsilon}_i-\bar{\boldsymbol{\upsilon}})+\left(\frac{\mathbf{X}_i'\bar{\mathbf{M}}_q\mathbf{X}_i}{T}\right)\left(\mathbf{h}_{iT}-\bar{\mathbf{h}}_T\right)+O_p\left(\frac{1}{\sqrt{N}}\right)+O_p\left(\frac{1}{\sqrt{T}}\right),$$

and since $(\boldsymbol{\upsilon}_i-\bar{\boldsymbol{\upsilon}})$ and $\left(\mathbf{h}_{iT}-\bar{\mathbf{h}}_T\right)$ are independently distributed across $i$ we then have

$$E\left[\frac{1}{N-1}\sum_{i=1}^{N}\tilde{w}_i^2\left(\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{X}_i}{T}\right)\left(\hat{\mathbf{b}}_i-\hat{\mathbf{b}}_{MG}\right)\left(\hat{\mathbf{b}}_i-\hat{\mathbf{b}}_{MG}\right)'\left(\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{X}_i}{T}\right)\right]=\mathbf{R}^*+O\left(\frac{1}{\sqrt{N}}\right)+O\left(\frac{1}{\sqrt{T}}\right).$$

Therefore, $\mathbf{R}^*$ can be consistently estimated by

$$\hat{\mathbf{R}}^*=\frac{1}{N-1}\sum_{i=1}^{N}\tilde{w}_i^2\left(\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{X}_i}{T}\right)\left(\hat{\mathbf{b}}_i-\hat{\mathbf{b}}_{MG}\right)\left(\hat{\mathbf{b}}_i-\hat{\mathbf{b}}_{MG}\right)'\left(\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{X}_i}{T}\right). \tag{6.53}$$

Using (5.24) we also note that $\Psi^*$ can be consistently estimated by

$$\hat{\Psi}^*=\sum_{i=1}^{N}w_i\left(\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{X}_i}{T}\right). \tag{6.54}$$

Hence

$$\widehat{AVar}\left(\hat{\mathbf{b}}_P\right)=\left(\sum_{i=1}^{N}w_i^2\right)\hat{\Psi}^{*-1}\hat{\mathbf{R}}^*\hat{\Psi}^{*-1}. \tag{6.55}$$

**Remark 6.1** *It can also be shown that when the rank condition (4.3) is satisfied Theorem 6.2 holds even if $\theta_i\neq w_i$. Further, in this case Assumption 3 can be relaxed by requiring the factor loadings, $\boldsymbol{\gamma}_i$, to be bounded. The expression for the asymptotic variance of $\left(\sum_{i=1}^{N}\theta_i^2\right)^{-1/2}\left(\hat{\mathbf{b}}_P-\boldsymbol{\beta}\right)$ also simplifies to*

$$\Sigma_P=\Psi^{-1}\mathbf{R}\Psi^{-1}, \tag{6.56}$$

*where*

$$\Psi=\lim_{N\to\infty}\left(\sum_{i=1}^{N}\theta_i\Sigma_i\right),\ \mathbf{R}=\lim_{(N,T)\overset{j}{\to}\infty}\left[N^{-1}\sum_{i=1}^{N}\tilde{\theta}_i^2\left(\Sigma_i\Omega_\upsilon\Sigma_i+T^{-1}\sigma_i^2\Sigma_i\right)\right], \tag{6.57}$$

*and*[9]

$$\widehat{AVar}\left(\hat{\mathbf{b}}_P\right)=\left(\sum_{i=1}^{N}\theta_i\hat{\Psi}_{iT}\right)^{-1}\left[\sum_{i=1}^{N}\theta_i^2\left(\hat{\Psi}_{iT}\hat{\Omega}_{\boldsymbol{\upsilon}}\hat{\Psi}_{iT}+T^{-1}\hat{\sigma}_i^2\hat{\Psi}_{iT}\right)\right]\left(\sum_{i=1}^{N}\theta_i\hat{\Psi}_{iT}\right)^{-1}, \tag{6.58}$$

[9]Although the second term of $\mathbf{R}$ in (6.57) is negligible when $T$ is sufficiently large, Monte Carlo experiments suggest that its inclusion could be beneficial when $T$ is small.

*where* $\hat{\Psi}_{iT} = T^{-1}\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{X}_i$, *and* $\hat{\sigma}_i^2$ *is defined by (5.36). To obtain* $\hat{\Omega}_{\boldsymbol{v}}$ *we use (6.48) and note that when the rank condition is satisfied, (6.47) reduces to*

$$\mathbf{h}_{iT} = \left(\frac{\mathbf{X}_i'\mathbf{M}_g\mathbf{X}_i}{T}\right)^{-1}\frac{\mathbf{X}_i'\mathbf{M}_g\boldsymbol{\varepsilon}_i}{T},$$

*and we have*

$$\hat{\Omega}_{\boldsymbol{v}} = \frac{1}{N-1}\sum_{i=1}^{N}\left(\hat{\mathbf{b}}_i - \hat{\mathbf{b}}_{MG}\right)\left(\hat{\mathbf{b}}_i - \hat{\mathbf{b}}_{MG}\right)' - \frac{1}{TN}\sum_{i=1}^{N}\hat{\sigma}_i^2\hat{\Psi}_{iT}^{-1}. \tag{6.59}$$

*As with Swamy type standard errors, it is possible for* $\hat{\Omega}_{\boldsymbol{v}}$ *to become non-negative definite when* $T$ *is small.*[10] *To avoid this possibility the second term in (6.59) which is of order* $T^{-1}$ *can be ignored. Alternatively, one could use the non-parametric estimator, (6.55), which is valid irrespective of whether the rank condition (4.3) is satisfied.*

Finally, the case where $\boldsymbol{\beta}_i$'s are homogeneous, namely when $\Omega_{\boldsymbol{v}} = \mathbf{0}$, requires special treatment. In this case $\hat{\mathbf{b}}_P$ converges to $\boldsymbol{\beta}$ at a faster rate and its asymptotic covariance matrix is no longer given by (6.50). Under $\boldsymbol{\beta}_i = \boldsymbol{\beta}$, and using (B.12) and (B.14) we have (noting that in this case $\boldsymbol{v}_i = \mathbf{0}$)

$$\left(\frac{\sum_{i=1}^{N}w_i^2}{T}\right)^{-1/2}\left(\hat{\mathbf{b}}_P - \boldsymbol{\beta}\right) \overset{d}{\sim} \Psi^{*-1}\left[\frac{1}{\sqrt{TN}}\sum_{i=1}^{N}\tilde{w}_i\mathbf{X}_i'\bar{\mathbf{M}}_w\left(\mathbf{F}\boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_i\right)\right], \tag{6.60}$$

where we have also multiplied both sides of (B.12) by $\sqrt{T}$ in order to avoid a degenerate asymptotic distribution. It is easily seen that $\hat{\mathbf{b}}_P$ continues to be consistent for $\boldsymbol{\beta}$ so long as $N \to \infty$, irrespective of whether $T$ is fixed or $\to \infty$. In general, however, its asymptotic distribution will depend on the nuisance parameters, with at least one important exception summarized in the following theorem.[11]

**Theorem 6.3** *Consider the panel data model (2.1) and (2.2) and suppose that Assumptions 1-4 and 5b hold,* $m = 1$, *the rank condition (4.3) is satisfied,* $\theta_i = w_i$, *and* $\boldsymbol{\beta}_i = \boldsymbol{\beta}$ *for all* $i$, *and* $T/N \to 0$, *as* $(N,T) \xrightarrow{j} \infty$. *Then*

$$\left(\frac{\sum_{i=1}^{N}w_i^2}{T}\right)^{-1/2}\left(\hat{\mathbf{b}}_P - \boldsymbol{\beta}\right) \xrightarrow{d} N(\mathbf{0}, \Sigma_{PH}), \tag{6.61}$$

*where*

$$\Sigma_{PH} = \Psi^{-1}\dot{\mathbf{R}}\Psi^{-1}, \tag{6.62}$$

[10] But the inclusion of $T^{-1}\hat{\sigma}_i^2\hat{\Psi}_{iT}$ in (6.58), which is also of order $T^{-1}$, should help compensate for the possible negative effect of $\hat{\Omega}_v$ on $AVar\left(\hat{\mathbf{b}}_P\right)$.

[11] See Appendix B for a proof.

$$\Psi = \lim_{N\to\infty}\left(\sum_{i=1}^{N} w_i\Sigma_i\right), \dot{\mathbf{R}} = \lim_{N\to\infty}\left(\frac{1}{N}\sum_{i=1}^{N}\tilde{w}_i^2\sigma_i^2\Sigma_i\right), \tag{6.63}$$

*and*

$$\tilde{w}_i = \frac{w_i}{\sqrt{N^{-1}\sum_{i=1}^{N} w_i^2}}.$$

This theorem also applies to the standard homogenous slope panel data models when $T$ is fixed and $N \to \infty$. But it is clearly not as general as Theorem 6.1 for the CCEMG estimator.

Under assumptions of Theorem 6.3, the asymptotic variance matrix of $\hat{\mathbf{b}}_P$ is given by

$$AVar\left(\hat{\mathbf{b}}_P\right) = \frac{1}{T}\left(\sum_{i=1}^{N} w_i\Sigma_i\right)^{-1}\left(\sum_{i=1}^{N} w_i^2\sigma_i^2\Sigma_i\right)\left(\sum_{i=1}^{N} w_i\Sigma_i\right)^{-1}, \tag{6.64}$$

which can be consistently estimated by

$$\widehat{AVar}(\hat{\mathbf{b}}_P) = \frac{1}{T}\left(\sum_{i=1}^{N} w_i\hat{\Psi}_{iT}\right)^{-1}\left(\sum_{i=1}^{N} w_i^2\tilde{\sigma}_i^2\hat{\Psi}_{iT}\right)\left(\sum_{i=1}^{N} w_i\hat{\Psi}_{iT}\right)^{-1}, \tag{6.65}$$

where

$$\tilde{\sigma}_i^2 = \frac{\left(\mathbf{y}_i - \mathbf{X}_i\hat{\mathbf{b}}_P\right)'\bar{\mathbf{M}}_w\left(\mathbf{y}_i - \mathbf{X}_i\hat{\mathbf{b}}_P\right)}{T}. \tag{6.66}$$

In general, however, where the conditions of theorem 6.3 might not be satisfied, one could use the non-parametric variance estimator of $\hat{\mathbf{b}}_P$, given by (6.55). The Monte Carlo experiments to be reported in Section 8 support such a strategy.

# 7 Determination of Optimal Weights

Our asymptotic results hold for all weights, $w_i$, that satisfy the atomistic conditions in (2.12). Clearly, these conditions do not uniquely determine these weights and the issue of an optimal choice for $w_i$'s naturally arises. One possible approach would be to determine the weights such that the asymptotic variance of the estimators of interest are minimized (in a suitable sense) subject to the conditions in (2.12). For the individual coefficients, $\hat{\mathbf{b}}_i$, with $T$ fixed, the variance matrix is given by (5.29), and does not depend on $w_i$'s, and the asymptotic (large $N$) properties of the CCE estimator would be invariant to the choice of the weights used in the construction of the cross section aggregates. By implication the same also applies to the CCEMG estimator, $\hat{\mathbf{b}}_{MG}$, defined by (6.37).

Consider now the CCE pooled estimator, $\hat{\mathbf{b}}_P$, under slope homogeneity. The asymptotic variance matrix of $\hat{\mathbf{b}}_P$ in this case is given by (6.64), and is minimized with $w_i$ set at

$$w_i^* = \frac{\sigma_i^{-2}}{\sum_{j=1}^{N}\sigma_j^{-2}}, \tag{7.67}$$

yielding

$$AVar\left(\hat{\mathbf{b}}_P(w^*)\right) = \frac{1}{T}\left(\sum_{i=1}^{N}\sigma_i^{-2}\Sigma_i\right)^{-1}. \tag{7.68}$$

Noting that $\Sigma_i$ is a positive definite matrix we can write

$$\begin{aligned} & T\left[AVar\left(\hat{\mathbf{b}}_P(w^*)\right)^{-1} - AVar\left(\hat{\mathbf{b}}_P\right)^{-1}\right] \\ = & \left(\sum_{i=1}^{N}\mathcal{X}_i\mathcal{X}_i'\right) - \left(\sum_{i=1}^{N}\mathcal{X}_i\mathcal{Y}_i'\right)\left(\sum_{i=1}^{N}\mathcal{Y}_i\mathcal{Y}_i'\right)^{-1}\left(\sum_{i=1}^{N}\mathcal{Y}_i\mathcal{X}_i'\right) \geq \mathbf{0}, \end{aligned}$$

where

$$\mathcal{X}_i = \sigma_i^{-1}\Sigma_i^{1/2}, \text{ and } \mathcal{Y}_i = w_i\sigma_i\Sigma_i^{1/2}.$$

This now establishes that $\left[AVar\left(\hat{\mathbf{b}}_P(w^*)\right)^{-1} - AVar\left(\hat{\mathbf{b}}_P\right)^{-1}\right]$ is a non-negative definite matrix, with $\{w_i^*\}$ providing an optimal choice in the sense that $AVar\left(\hat{\mathbf{b}}_P(w^*)\right) \leq AVar\left(\hat{\mathbf{b}}_P\right)$.

Not surprisingly the pooled estimator computed using $w_i^*$ reduces to the generalized least squares estimator

$$\hat{\mathbf{b}}_P(w^*) = \left(\sum_{i=1}^{N}\sigma_i^{-2}\mathbf{X}_i'\bar{\mathbf{M}}_{w^*}\mathbf{X}_i\right)^{-1}\sum_{i=1}^{N}\sigma_i^{-2}\mathbf{X}_i'\bar{\mathbf{M}}_{w^*}\mathbf{y}_i, \tag{7.69}$$

with its feasible counterpart obtained by replacing $\sigma_i^2$ with the estimates, $\hat{\sigma}_i^2$, given by (5.36) and computed using an initial consistent estimator of $\boldsymbol{\beta}$ based on (say) $w_i = 1/N$. Recall, however, for the pooled estimator to remain asymptotically valid the weights used for the construction of the aggregates must be the same as the ones used in the formation of the pooled estimator.

# 8 Small Sample Properties of CCE Estimators: Monte Carlo Experiments

This section provides Monte Carlo evidence on the small sample properties of the CCEMG and the CCEP estimators defined by (6.37) and (6.49), respectively, using the weights $w_i = \theta_i = 1/N$, and the following data generating process (DGP):

$$y_{it} = \alpha_{i1}d_{1t} + \beta_{i1}x_{1it} + \beta_{i2}x_{2it} + \gamma_{i1}f_{1t} + \gamma_{i2}f_{2t} + \varepsilon_{it}, \tag{8.1}$$

and

$$x_{ijt} = a_{ij1}d_{1t} + a_{ij2}d_{2t} + \gamma_{ij1}f_{1t} + \gamma_{ij3}f_{3t} + \mathrm{v}_{ijt},\ j = 1, 2, \tag{8.2}$$

for $i = 1, 2, ..., N$, and $t = 1, 2, ..., T$. This DGP is a restricted version of the general linear model considered in the paper, and sets $n = k = 2$, and $m = 3$, with $\boldsymbol{\alpha}_i' = (\alpha_{i1}, 0)$, $\boldsymbol{\beta}_i' = (\beta_{i1}, \beta_{i2})$, and $\boldsymbol{\gamma}_i' = (\gamma_{i1}, \gamma_{i2}, 0)$, imposed on (2.1) and (2.2), and

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

on the (2.3). The common factors and the individual specific errors of $\mathbf{x}_{it}$ are generated as independent stationary AR(1) processes with zero means and unit variances:

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

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

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

and

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

The individual specific errors of $y_{it}$ are generated as

$$\varepsilon_{it} \sim IIDN\left(0, \sigma_i^2\right), \ \sigma_i^2 \sim IIDU\left[0.5, 1.5\right].$$

The factor loadings of the observed common effects, $\alpha_{i1}$, and $vec(\mathbf{A}_i) = (a_{i11}, a_{i21}, a_{i12}, a_{i22})'$ are generated as $IIDN(1, 1)$, and $IIDN(0.5\boldsymbol{\tau}_4, 0.5\ \mathbf{I}_4)$, 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 $\mathbf{x}_{it}$ equation are generated independently across replications as

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

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 (4.3) 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

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

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

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

and the rank condition is *not* satisfied. For each set we conducted two different experiments:[12]

- **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 $\mathcal{A}1$ and $\mathcal{B}1$, and those of experiment 2 by $\mathcal{A}2$, and $\mathcal{B}2$.[13] For each experiment we 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 $(d_{1t},\mathbf{x}_{it})$, and the "naive" estimators that excludes these factors. The infeasible MG (Pooled) estimator provides an upper bound to the efficiency of the CCEMG (CCEP) estimator under slope heterogeneity (homogeneity), whilst the naive estimators illustrate the extent of bias and size distortions that can occur if the error cross section dependence is ignored. Each experiment was replicated 2000 times for the $(N,T)$ pairs with $N,T=20,30,50,100,200$. In what follows we shall focus on $\beta_1$ (the cross section mean of $\beta_{i1}$). Results for $\beta_2$ are very similar and will not be reported.

### 8.1 Bias and RMSE

Results of experiments $\mathcal{A}1$ and $\mathcal{B}1$ are summarized in Tables A1(i)-A1(iv) and B1(i)-B1(iv), respectively. Not surprisingly, as can be seen from Tables A1(i)-A1(iv) the naive estimator is substantially

[12]We also carried out a number of experiments with $\gamma_{ij}\sim IIDN\left(0.5,0.2\right)$, for $j=1,2$, that give a lower degree of error cross section dependence as compared to $\gamma_{ij}\sim IIDN\left(1,0.2\right)$, but obtained very similar results. We decided to report the outcomes of the experiments with the higher cross section dependence, as they are likely to provide a more demanding check on the validity of the CCE estimators.

[13]We also carried out a third set of experiments with $\beta_{i2}=0$, so that $k+1<m$. Once again the results turned out to be qualitatively the same. The failure of the order or rank condition does not seem to play a significant role in the outcomes.

biased, performs very poorly and is subject to large size distortions; an outcome that continues to apply in the case of other experiments. To save space we provide results for the naive estimators only in the case of experiment $\mathcal{A}1$. In contrast, the bias of the CCEMG and CCEP estimators are very small and comparable to the bias of the associated infeasible estimators. A comparison of the bias estimates in Tables A1(i) and B1(i) also shows that the bias of the CCE type estimators does not depend on whether the rank condition, (4.3), is satisfied.[14]

Table A1(ii) provides the root mean squared errors (RMSE) of the various estimators for experiment $\mathcal{A}1$ (full rank+heterogenous slopes). Under this experiment the lower bound to CCEMG's RMSE is given by the RMSE of the infeasible MG estimator. For $T = N = 20$, the RMSE of the CCEMG is 32.1% higher than that of the infeasible MG, and falls steadily with $N$ and $T$, and ends up being only 2.5% higher for $T = N = 200$. The Monte Carlo results also confirm the asymptotic efficiency of the MG type estimators relative to the pooled estimators under slope heterogeneity. This seems to occur for $T \geq 30$. It is also interesting to note that the CCEP estimator in fact dominates the infeasible pooled estimator for $N \geq 30$ and $T \geq 50$. For example, for $N = 50$ and $T = 100$ the RMSE of the CCEP estimator is 9% lower than the RMSE of the infeasible pooled estimator. Overall, both CCEMG and CCEP provide reasonably efficient estimators, particularly for relatively large $N$ and $T$, with the CCEP doing slightly better in small samples. This general conclusion also holds in the rank deficient case, as can be seen from the results summarized in Table B1(ii). In the rank deficient case, however, the efficiency loss of the CCEMG relative to the infeasible MG is higher, being 69% (compared to 32.1% under full rank) at $N = T = 20$ and 11.5% (compared to 2.5% under full rank) at $N = T = 200$.

The RMSE results for the homogeneous slope experiments, $\mathcal{A}2$ and $\mathcal{B}2$, are summarized in Tables A2(i) and B2(i). For these experiments the pooled estimators are expected to be more efficient than the MG estimators, and this is corroborated by the results in these Tables, although the differences between MG and pooled estimators become very small as $N$ and $T$ are increased. The efficiency loss of the CCE estimators relative to their infeasible counterparts also tends to be slightly higher in the case of the homogeneous slope experiments, as compared to the heterogenous slope case discussed above. Once again the same qualitative conclusions follow under rank deficiency, although the efficiency loss of not knowing the true error factor model is now even greater. See Table B2(i).

Of course, in reality the true error factor model is not known even if other proxies could be found for the unobserved factors, $\mathbf{f}_t$. It is not clear how this can be accomplished in the present experimental set up. Therefore, within the realm of feasible estimators the choice is between CCEMG and CCEP. The simulation results tend to favour the CCEP for small to moderate sample sizes and CCEMG when $N$ and $T$ are relatively large. This conclusion seems to be robust and

[14]To save space we are not reporting the bias estimates for the homogeneous slope experiments $\mathcal{A}2$ and $\mathcal{B}2$.

stands for homogeneous as well as heterogeneous slope experiments, and does not seem to depend on whether the rank condition is satisfied.

Finally, it is worth emphasizing that knowing the factors or having good proxies for them is not enough; one must also know which of them influence $y_{it}$ and which of them influence $\mathbf{x}_{it}$. This would involve specification searches that are not required by the CCE estimators.

### 8.2 Size and Power

For the full rank and heterogenous experiments $\mathcal{A}1$, size and power of a two-sided test of $\beta_1 = 1$ are reported in Tables A1(iii) and A1(iv), respectively. The variance of the CCEMG estimator is computed using (6.42), both under heterogeneous and homogeneous slope coefficients. The empirical size of the test based on the CCEMG estimator is very close to the nominal size of 5%, for all values of $N$ and $T$ except for $T = 20$, which is slightly over-sized. As can be seen from Tables B1(iii), A2(ii), and B2(ii), this conclusion continues to hold for all other experiments and does not seem to depend on the rank condition or the homogeneity/heterogeneity of the slopes. This is in line with our theoretical results set out in Theorem 6.1.

By comparison, tests based on the CCEP estimator are less robust and depend on the choice of the variance estimator, namely whether (6.55) or (6.65) is used. Under heterogenous slopes the appropriate variance estimator is (6.55), which is the one used to produce the results in Table A1(iii). In this case the size of the CCEP test is very similar to those obtained using CCEMG. As can be seen from Table B1(iii), this conclusion holds even if the rank condition is not satisfied. However, as predicted by Theorem 6.3, under slope homogeneity, $\boldsymbol{\beta}_i = \boldsymbol{\beta}$, the validity of a test based on CCEP using the variance estimator (6.65) requires $T/N$ to be relatively small, even if the rank condition is satisfied. This can be clearly seen in the empirical sizes of the CCEP test summarized in Tables A2(ii) and B2(ii). It is also interesting that rank deficiency now seems to make a noticeable difference to the results. The empirical sizes for CCEP in Table B2(ii) are generally higher than those in Table A2(ii).

Given the efficiency of CCEP estimator relative to the CCEMG estimator under slope homogeneity, and the fact that CCEP is asymptotically unbiased as $N \to \infty$, the over-rejection tendency of the CCEP test is most likely due to inappropriate standard errors. One possible alternative would be to use the heterogenous variance estimator, (6.55), even under slope homogeneity.[15] We denote this test by CCEP(hetro), and report its empirical size in Tables A2(ii) and B2(ii). The CCEP(hetro) test results all have the correct size for $N, T \geq 20$, and the outcomes no longer depend on the rank condition.

The power of the various tests are computed under the alternative, $\beta_1 = 0.95$ and reported in

[15] It is unlikely that it would be known with certainty that $\boldsymbol{\beta}_i = \boldsymbol{\beta}$, and in practice the use of CCEP(hetro) might be advisable on *a priori* grounds.

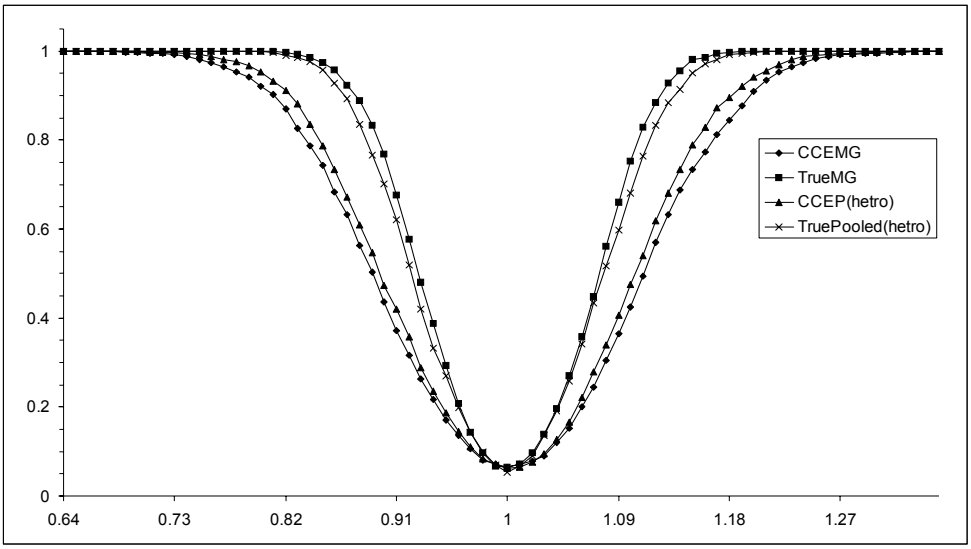


Figure 1: Power Function for Experiment B1, N=50, T=30

Tables A1(iv) and B1(iv) under slope heterogeneity, and in Tables A2(iii) and B2(iii) under slope homogeneity, respectively. Given the size distortion of the CCEP test under slope homogeneity, we only report the power of CCEP(hetro) in these tables. CCEP(hetro) tends to be more powerful than CCEMG for moderate values of $N$ and $T$, particularly for $T \leq 30$.

A comparison of the power of the CCE type tests with the tests based on the infeasible estimators shows, perhaps not surprisingly, that not knowing the true error factor process would result in some loss of power, although the power differentials tend to die out relatively rapidly with increases in $N$ and $T$.

Finally, as can be seen from Figure 1, the power function of the tests tend to be symmetric and have the familiar inverted bell shape. As an illustration, Figure 1 shows the power function of CCEMG and CCEP(hetro) tests, as well as the associated infeasible tests, in the case of experiment $\mathcal{B}1$ for $N = 50$ and $T = 30$. The figure clearly shows that for this sample size the CCEP(hetro) test performs slightly better than the CCEMG test, and as compared with the tests based on the infeasible estimators the two CCE tests seem to perform reasonably well.

# 9 Concluding Remarks

This paper provides a simple procedure for estimation of panel data models subject to error cross section dependence when the cross section dimension ($N$) of the panel is sufficiently large. The asymptotic theory required for estimation and inference is developed under fairly general conditions both when the time dimension ($T$) is fixed and when $T \to \infty$. Conditions under which the proposed

correlated common effects estimators are consistent and asymptotically normal are provided. The Monte Carlo experiments show that the pooled estimators have satisfactory small sample properties. Further extensions and generalizations are, however, clearly desirable.

The focus of this paper has been on estimation of $\boldsymbol{\beta}_i$ and their means, $\boldsymbol{\beta}$. Our analysis shows that consistent estimation of $\boldsymbol{\beta}$, can be carried out for any fixed but unknown $m$, the number of unobserved factors. A priori knowledge of $m$ is not required. But if the focus of the analysis is on the factor loadings, as is the case, for example, in the multifactor asset pricing models, an estimate of $m$ would be needed. This can be achieved, for example, by application of the Bai and Ng's (2002) procedure to the residuals

$$\hat{\mathbf{e}}_i = \bar{\mathbf{M}}\left(\mathbf{y}_i - \mathbf{X}_i\hat{\mathbf{b}}_i\right), \text{ or } \hat{\mathbf{e}}_i = \bar{\mathbf{M}}\left(\mathbf{y}_i - \mathbf{X}_i\hat{\mathbf{b}}_P\right).$$

Under our assumptions, for any fixed $m$ these residuals provide consistent estimates of $e_{it}$ in the multifactor model (2.1), and could be used as "observed data" to obtain estimates of the factors $\mathbf{f}_t$ (subject to orthonormalization restrictions, for example). It is reasonable to expect these factor estimates (denoted by $\hat{\mathbf{f}}_t$) to be consistent. The factor estimates can then be used directly as (generated) regressors in the regression equation

$$y_{it} = \boldsymbol{\alpha}_i'\mathbf{d}_t + \boldsymbol{\beta}_i'\mathbf{x}_{it} + \boldsymbol{\gamma}_i'\hat{\mathbf{f}}_t + \zeta_{it},$$

to obtain the estimates of the factor loadings, $\boldsymbol{\gamma}_i$, or their means, $\boldsymbol{\gamma}$. The small sample properties of such a two-stage procedure would also be of interest.

Further, it is desirable to see if the results of this paper carry over to the case where lagged values of $y_{it}$ are allowed to be included amongst the individual-specific regressors. The regression model (2.1) allows for dynamics only through the general dynamics of the common effects in $e_{it}$, and the fact that these effects could have differential impacts on different groups. This is restrictive and its relaxation is clearly important for a wider applicability of the approach advanced in this paper. Pesaran (2003) provides an application of the CCE approach to testing for unit roots in the presence of error cross section dependence. But more general treatments would be desirable.

Another important extension is to multi-variate panel data models such as Panel Vector Autoregressions (PVAR) of the type discussed, for example, in Binder, Hsiao and Pesaran (2004).

These further developments are beyond the scope of the present paper and will be the subject of separate studies.

# Appendix A: Lemmas: Statements and Proofs

**Lemma A.1** *Suppose that either $\|\boldsymbol{\beta}_i\| < K$, or that the random coefficient Assumption 4 holds. Then under Assumption 2 for each $t$, we have*

$$E\left(\bar{\mathbf{u}}_{wt}\right) = 0, \tag{A.1}$$

$$Var(\bar{\mathbf{u}}_{wt}) = O\left(\sum_{i=1}^{N} w_i^2\right) = O\left(\frac{1}{N}\right), \tag{A.2}$$

$$\bar{\mathbf{u}}_{wt} \overset{q.m.}{\to} 0, \text{ as } N \to \infty, \tag{A.3}$$

$$E\left\|\bar{\mathbf{u}}_{wt}\right\|^2 = O\left(\frac{1}{N}\right), \text{ and } E\left\|\bar{\mathbf{u}}_{wt}\right\| = O\left(\frac{1}{\sqrt{N}}\right), \tag{A.4}$$

*where $\bar{\mathbf{u}}_{wt} = \sum_{i=1}^{N} w_i \mathbf{u}_{it}$, $\mathbf{u}_{it}$ is defined by (2.5) and the weights, $w_i$, satisfy the conditions in (2.12).*

**Proof**: First note that

$$\bar{\mathbf{u}}_{wt} = \begin{pmatrix} \bar{\varepsilon}_{wt} + \sum_{i=1}^{N} w_i \boldsymbol{\beta}_i' \mathbf{v}_{it} \\ \bar{\mathbf{v}}_{wt} \end{pmatrix}, \tag{A.5}$$

where $\bar{\mathbf{v}}_{wt} = \sum_{i=1}^{N} \sum_{\ell=0}^{\infty} w_i \mathbf{S}_{i\ell} \boldsymbol{\nu}_{i,t-\ell}$. Since $\boldsymbol{\nu}_{it} \sim IID(\mathbf{0}, \mathbf{I}_k)$, then conditional on $w_i$ and $\mathbf{S}_{i\ell}$, $Var\left(\bar{\mathbf{v}}_{wt}\right) = \sum_{i=1}^{N} w_i^2 \left(\sum_{\ell=0}^{\infty} \mathbf{S}_{i\ell} \mathbf{S}_{i\ell}'\right)$, and using (2.9) and (2.12) we have (unconditionally)

$$Var\left(\bar{\mathbf{v}}_{wt}\right) \leq \mathbf{K}\left(\sum_{i=1}^{N} w_i^2\right) = O\left(\frac{1}{N}\right). \tag{A.6}$$

Similarly,

$$Var\left(\bar{\varepsilon}_{wt}\right) = O\left(\frac{1}{N}\right), \tag{A.7}$$

and

$$Var\left(\sum_{i=1}^{N} w_i \boldsymbol{\beta}_i' \mathbf{v}_{it}\right) = \sum_{i=1}^{N} w_i^2 E\left(\boldsymbol{\beta}_i' \Sigma_i \boldsymbol{\beta}_i\right) \leq \sum_{i=1}^{N} w_i^2 E\left(\boldsymbol{\beta}_i' \boldsymbol{\beta}_i\right) E\left[\boldsymbol{\lambda}_{\max}(\Sigma_i)\right]$$

where $\boldsymbol{\lambda}_{\max}(\Sigma_i)$ is the maximum eigen value of $\Sigma_i$ which is bounded by Assumption 2. Also, either $\boldsymbol{\beta}_i' \boldsymbol{\beta}_i = \|\boldsymbol{\beta}_i\|^2 < K$, or under Assumption 4 we have $E\left(\boldsymbol{\beta}_i' \boldsymbol{\beta}_i\right) < K$, and therefore

$$Var\left(\sum_{i=1}^{N} w_i \boldsymbol{\beta}_i' \mathbf{v}_{it}\right) = O\left(\sum_{i=1}^{N} w_i^2\right) = O\left(\frac{1}{N}\right). \tag{A.8}$$

Using (A.6), (A.7) and (A.8) in connection with (A.5), and noting that

$$Cov\left(\bar{\varepsilon}_{wt} + \sum_{i=1}^{N} w_i \boldsymbol{\beta}_i' \mathbf{v}_{it}, \bar{\mathbf{v}}_{wt}\right) = \sum_{i=1}^{N} w_i^2 E\left(\boldsymbol{\beta}_i'\right) \Sigma_i = O\left(\sum_{i=1}^{N} w_i^2\right) = O\left(\frac{1}{N}\right),$$

it also readily follows that

$$Var\left(\bar{\mathbf{u}}_{wt}\right) = O\left(\sum_{i=1}^{N} w_i^2\right) = O\left(\frac{1}{N}\right), \tag{A.9}$$

which establishes (A.3), considering that $E\left(\bar{\mathbf{u}}_{wt}\right) = \mathbf{0}$.

To prove (A.4), note that by assumption $E\left(\mathbf{v}_{it}'\mathbf{v}_{it}\right) = Tr\left(\Sigma_i\right) < K$, and $\sigma_i^2 + E\left(\boldsymbol{\beta}_i'\Sigma_i\boldsymbol{\beta}_i\right) < K$, and hence using (A.5):

$$E\left\|\bar{\mathbf{u}}_{wt}\right\|^2 = \sum_{i=1}^{N} w_i^2\left[\sigma_i^2 + E\left(\boldsymbol{\beta}_i'\Sigma_i\boldsymbol{\beta}_i\right) + E\left(\mathbf{v}_{it}'\mathbf{v}_{it}\right)\right] = O\left(\sum_{i=1}^{N} w_i^2\right) = O\left(\frac{1}{N}\right).$$

Further,

$$E\left\|\bar{\mathbf{u}}_{wt}\right\| \leq \left[E\left\|\bar{\mathbf{u}}_{wt}\right\|^2\right]^{1/2} = O\left(\frac{1}{\sqrt{N}}\right). \blacksquare$$

**Lemma A.2** *Suppose that either $\|\boldsymbol{\beta}_i\| < K$, or that the random coefficient Assumption 4 holds. Then under Assumptions 1 and 2*

$$\frac{\bar{\mathbf{U}}_w'\bar{\mathbf{U}}_w}{T} = O_p\left(\frac{1}{N}\right), \tag{A.10}$$

$$\frac{\mathbf{F}'\bar{\mathbf{U}}_w}{T} = O_p\left(\frac{1}{\sqrt{NT}}\right), \ \frac{\mathbf{D}'\bar{\mathbf{U}}_w}{T} = O_p\left(\frac{1}{\sqrt{NT}}\right), \tag{A.11}$$

$$\frac{\mathbf{V}_i'\mathbf{D}}{T} = O_p\left(\frac{1}{\sqrt{T}}\right), \ \frac{\mathbf{V}_i'\mathbf{F}}{T} = O_p\left(\frac{1}{\sqrt{T}}\right), \tag{A.12}$$

$$\frac{\mathbf{V}_i'\bar{\mathbf{U}}_w}{T} = O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right), \ \frac{\boldsymbol{\varepsilon}_i'\bar{\mathbf{U}}_w}{T} = O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right), \tag{A.13}$$

*where $\bar{\mathbf{U}}_w = \left(\bar{\mathbf{u}}_{w1}, \bar{\mathbf{u}}_{w2}, ..., \bar{\mathbf{u}}_{wT}\right)'$, $\bar{\mathbf{u}}_{wt}$ is defined by (A.5), the weights, $w_i$, satisfy the conditions in (2.12), $\mathbf{V}_i = \left(\mathbf{v}_{i1}, \mathbf{v}_{i2}, ..., \mathbf{v}_{iT}\right)'$, $\mathbf{D}$ and $\mathbf{F}$ are $T \times n$ and $T \times m$, data matrices on observed and unobserved common factors.*

**Proof**: Note that $T^{-1}\bar{\mathbf{U}}_w'\bar{\mathbf{U}}_w = T^{-1}\left(\sum_{t=1}^{T}\bar{\mathbf{u}}_{wt}\bar{\mathbf{u}}_{wt}'\right)$, where the cross-product terms in $\bar{\mathbf{u}}_{wt}\bar{\mathbf{u}}_{wt}'$, being functions of linear stationary processes with fourth-order cumulants, are themselves stationary with finite means and variances. Also, $E\left\|T^{-1}\bar{\mathbf{U}}_w'\bar{\mathbf{U}}_w\right\| \leq T^{-1}\sum_{t=1}^{T}E\left\|\bar{\mathbf{u}}_{wt}\right\|^2$, and by (A.4) $E\left\|T^{-1}\bar{\mathbf{U}}_w'\bar{\mathbf{U}}_w\right\| = O\left(N^{-1}\right)$, which establishes (A.10).

Consider the $\ell^{th}$ row of $T^{-1}\left(\mathbf{F}'\bar{\mathbf{U}}_w\right)$ and note that it can be written as $T^{-1}\left(\sum_{t=1}^{T}f_{\ell t}\bar{\mathbf{u}}_{wt}'\right)$. Since by assumption $f_{\ell t}$ and $\bar{\mathbf{u}}_{wt}$ are independently distributed covariance stationary processes then

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

where $E\left(\bar{\mathbf{u}}_{wt}\bar{\mathbf{u}}_{wt'}'\right) = O\left(N^{-1}\right)$. Hence,

$$\begin{aligned} Var\left(\frac{\sum_{t=1}^{T}f_{\ell t}\bar{\mathbf{u}}_{wt}}{T}\right) &= O\left(\frac{1}{N}\right)\left\{\frac{\sum_{t=1}^{T}\sum_{t'=1}^{T}E\left(f_{\ell t}f_{\ell t'}\right)}{T^2}\right\} \\ &= O\left(\frac{1}{N}\right)\left\{\frac{\sum_{t=1}^{T}\sum_{t'=1}^{T}\Gamma_{f\ell}\left(|t-t'|\right)}{T^2}\right\}, \end{aligned}$$

where $\Gamma_{f\ell}\left(|t-t'|\right)$ is the autocovariance function of the stationary process, $f_{\ell t}$, which decays exponentially in $|t-t'|$. Therefore,

$$Var\left(\frac{\sum_{t=1}^{T}f_{\ell t}\bar{\mathbf{u}}_{wt}}{T}\right) = O\left(\frac{1}{NT}\right), \tag{A.14}$$

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 $O_p\left(1/\sqrt{NT}\right)$. Consider now the limit of $T^{-1}\sum_{t=1}^{T}f_{\ell t}\bar{\mathbf{u}}_{wt}$ and note that since $f_{\ell t}$ and $\bar{\mathbf{u}}_{wt}$ are independently distributed covariance stationary processes,

$$\frac{\sum_{t=1}^{T}f_{\ell t}\bar{\mathbf{u}}_{wt}}{T} = O_p\left(\frac{1}{\sqrt{T}}\right), \text{ for any fixed } N,$$

and by (A.4)

$$E\left\|\frac{\sum_{t=1}^{T} f_{\ell t}\bar{\mathbf{u}}_{wt}}{T}\right\| \leq \frac{\sum_{t=1}^{T} E\left\|f_{\ell t}\right\| E\left\|\bar{\mathbf{u}}_{wt}\right\|}{T} = O_p\left(\frac{1}{\sqrt{N}}\right), \text{ for any fixed } T.$$

Furthermore, since for each $t$, $\mathbf{u}_{it}$'s are cross sectionally independent, then by standard central limit theorems for independent but not identically distributed random variables we have $\sqrt{N}\bar{\mathbf{u}}_{wt} \xrightarrow{d} O_p(1)$, as $N \to \infty$. Therefore,

$$\frac{\sum_{t=1}^{T} f_{\ell t}\sqrt{N}\bar{\mathbf{u}}'_{wt}}{\sqrt{T}} \xrightarrow{d} O_p(1) \text{ as } (N,T) \xrightarrow{j} \infty,$$

as required. The second result in (A.11) follows similarly.

The results in (A.12) are standard in the literature on independent stationary processes.

To establish the results in (A.13), using (A.5) first note that

$$T^{-1}\mathbf{V}_i'\bar{\mathbf{U}}_w = \left(T^{-1}\mathbf{V}_i'\bar{\boldsymbol{\varepsilon}}_w + T^{-1}\mathbf{V}_i'\sum_{j=1}^{N} w_j\mathbf{V}_j\boldsymbol{\beta}_j,\ T^{-1}\mathbf{V}_i'\bar{\mathbf{V}}_w\right), \tag{A.15}$$

where $\bar{\boldsymbol{\varepsilon}}_w = \sum_{j=1}^{N} w_j\boldsymbol{\varepsilon}_j$ and $\bar{\mathbf{V}}_w = \sum_{j=1}^{N} w_j\mathbf{V}_j$. Since, by assumption $\mathbf{v}_{it}$ and $\bar{\varepsilon}_{wt}$ are independently distributed covariance stationary processes, then by following the same line of reasoning as used for the proof of (A.11) we have

$$T^{-1}\mathbf{V}_i'\bar{\boldsymbol{\varepsilon}}_w = O_p\left(\frac{1}{\sqrt{NT}}\right). \tag{A.16}$$

Consider the second term in (A.15) and note that

$$T^{-1}\mathbf{V}_i'\sum_{j=1}^{N} w_j\mathbf{V}_j\boldsymbol{\beta}_j = w_i\left(\frac{\mathbf{V}_i'\mathbf{V}_i}{T}\right)\boldsymbol{\beta}_i + \left(\frac{\mathbf{V}_i'\bar{\mathbf{V}}^*_{w,-i}}{T}\right), \tag{A.17}$$

where $\bar{\mathbf{V}}^*_{w,-i} = \sum_{j=1,j\neq i}^{N} w_j\mathbf{V}_j\boldsymbol{\beta}_j$. Since $w_i = O(N^{-1})$, $\boldsymbol{\beta}_i$ is either bounded or satisfy the conditions of Assumption 4 and the elements of $\mathbf{V}_i$ are covariance stationary, then

$$w_i\left(\frac{\mathbf{V}_i'\mathbf{V}_i}{T}\right)\boldsymbol{\beta}_i = O_p\left(\frac{1}{N}\right). \tag{A.18}$$

Also since the elements of $\mathbf{V}_i$ and $\bar{\mathbf{V}}^*_{w,-i}$ are independently distributed and covariance stationary, using the same line of reasoning as above we have

$$\frac{\mathbf{V}_i'\bar{\mathbf{V}}^*_{w,-i}}{T} = O_p\left(\frac{1}{\sqrt{NT}}\right). \tag{A.19}$$

Using (A.18) and (A.19) in (A.17) now yields

$$T^{-1}\mathbf{V}_i'\sum_{j=1}^{N} w_j\mathbf{V}_j\boldsymbol{\beta}_j = O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right). \tag{A.20}$$

Finally, since the last term of (A.15) can be written as

$$T^{-1}\mathbf{V}_i'\bar{\mathbf{V}}_w = w_i\left(\frac{\mathbf{V}_i'\mathbf{V}_i}{T}\right) + \frac{\mathbf{V}_i'\bar{\mathbf{V}}_{w,-i}}{T}$$

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

$$T^{-1}\mathbf{V}_i'\bar{\mathbf{V}}_w = O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right). \tag{A.21}$$

Using (A.16), (A.20) and (A.21) in (A.15) now establishes the first result in (A.13). The second result also follows similarly. ■

**Lemma A.3** *Suppose that the conditions of Lemma A.2 hold, and* $\|\Pi_i\| \leq K$, *where* $\Pi_i = (\mathbf{A}_i', \Gamma_i')'$ *and* $\mathbf{A}_i$ *and* $\Gamma_i$ *are the parameters of the* $\mathbf{x}_{it}$ *process defined by (2.3). Then*

$$\frac{\mathbf{X}_i'\bar{\mathbf{U}}_w}{T} = O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right) \tag{A.22}$$

**Proof**: Using (5.12) we have

$$\frac{\mathbf{X}_i'\bar{\mathbf{U}}_w}{T} = \Pi_i'\left(\frac{\mathbf{G}'\bar{\mathbf{U}}_w}{T}\right) + \left(\frac{\mathbf{V}_i'\bar{\mathbf{U}}_w}{T}\right),$$

and (A.22) follows from (A.11) and (A.13), and since by assumption the elements of $\Pi_i$ are bounded. ■

**Lemma A.4** *Suppose that Assumption 3, and conditions (2.12) and (2.16) hold and* $\mathbf{Q}_{iT}$ *is a* $k \times m$ *matrix, distributed independently of* $\boldsymbol{\eta}_i \sim IID\left(\mathbf{0}, \Omega_\eta\right)$, $\|\Omega_\eta\| < K$, *and* $E\|\mathbf{Q}_{iT}\| < K$. *Let*

$$\mathbf{q}_{NT} = \left(\sum_{i=1}^N \theta_i^2\right)^{-1/2} \sum_{i=1}^N \theta_i \mathbf{Q}_{iT}\left(\boldsymbol{\eta}_i - \bar{\boldsymbol{\eta}}_w\right),$$

*where* $\bar{\boldsymbol{\eta}}_w = \sum_{i=1}^N w_i \boldsymbol{\eta}_i$, *and* $\boldsymbol{\eta}_i$, $w_i$ *and* $\theta_i$ *are defined by (2.10), (2.12), and (2.16), respectively. Then*

$$\mathbf{q}_{NT} \xrightarrow{d} N(\mathbf{0}, \Sigma_{qT}), \text{ as } N \to \infty,$$

*where*

$$\Sigma_{qT} = \lim_{N\to\infty}\left(N^{-1}\sum_{i=1}^N \mathbf{P}_{iT}\Omega_\eta \mathbf{P}_{iT}'\right) < \mathbf{K},$$

*and*

$$\mathbf{P}_{iT} = \frac{\theta_i}{\sqrt{N^{-1}\sum_{i=1}^N \theta_i^2}}\mathbf{Q}_{iT} - \frac{w_i}{\sqrt{N^{-1}\sum_{i=1}^N \theta_i^2}}\bar{\mathbf{Q}}_{\theta T}, \quad \bar{\mathbf{Q}}_{\theta T} = \sum_{i=1}^N \theta_i \mathbf{Q}_{iT}.$$

**Proof**: The result follows observing that

$$\left(\sum_{i=1}^N \theta_i^2\right)^{-1/2} \sum_{i=1}^N \theta_i \mathbf{Q}_{iT}\left(\boldsymbol{\eta}_i - \bar{\boldsymbol{\eta}}_w\right) = \sum_{i=1}^N \mathbf{P}_{iT}\boldsymbol{\eta}_i,$$

$$E\|\mathbf{P}_{iT}\| < \frac{|\theta_i|}{\sqrt{\sum_{i=1}^N \theta_i^2}} E\|\mathbf{Q}_{iT}\| + \frac{|w_i|}{\sqrt{\sum_{i=1}^N \theta_i^2}} E\left\|\bar{\mathbf{Q}}_{\theta T}\right\|,$$

$$E\left\|\bar{\mathbf{Q}}_{\theta T}\right\| < \sum_{i=1}^N |\theta_i| E\|\mathbf{Q}_{iT}\| < K,$$

and since by assumption

$$\frac{|\theta_i|}{\sqrt{N^{-1}\sum_{i=1}^N \theta_i^2}} = O(1), \text{ and } \frac{|w_i|}{\sqrt{N^{-1}\sum_{i=1}^N \theta_i^2}} = O(1). \blacksquare$$

# Appendix B: Mathematical Proofs

**Proof of Asymptotic Unbiasedness of** $\hat{\Sigma}_{T,b_i}$

Here $T$ is fixed and the rank condition (4.3) is satisfied. $\hat{\sigma}_i^2$ given by (5.34) can be written as

$$\hat{\sigma}_i^2 = \frac{\mathbf{y}_i'\bar{\mathbf{S}}_w\mathbf{y}_i}{T-(n+m+k)}, \tag{B.1}$$

where

$$\bar{\mathbf{S}}_w = \bar{\mathbf{M}}_w - \bar{\mathbf{M}}_w\mathbf{X}_i\left(\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{X}_i\right)^{-1}\mathbf{X}_i'\bar{\mathbf{M}}_w. \tag{B.2}$$

Under (5.10)

$$\mathbf{y}_i'\bar{\mathbf{S}}_w\mathbf{y}_i = \boldsymbol{\gamma}_i'\mathbf{F}'\bar{\mathbf{S}}_w\mathbf{F}\boldsymbol{\gamma}_i - 2\boldsymbol{\gamma}_i'\mathbf{F}'\bar{\mathbf{S}}_w\boldsymbol{\varepsilon}_i + \boldsymbol{\varepsilon}_i'\bar{\mathbf{S}}_w\boldsymbol{\varepsilon}_i, \tag{B.3}$$

where

$$\boldsymbol{\gamma}_i'\mathbf{F}'\bar{\mathbf{S}}_w\mathbf{F}\boldsymbol{\gamma}_i = \boldsymbol{\gamma}_i'\mathbf{F}'\bar{\mathbf{M}}_w\mathbf{F}\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_i'\mathbf{F}'\bar{\mathbf{M}}_w\mathbf{X}_i\left(\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{X}_i\right)^{-1}\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{F}\boldsymbol{\gamma}_i, \tag{B.4}$$

$$\boldsymbol{\gamma}_i'\mathbf{F}'\bar{\mathbf{S}}_w\boldsymbol{\varepsilon}_i = \boldsymbol{\gamma}_i'\mathbf{F}'\bar{\mathbf{M}}_w\boldsymbol{\varepsilon}_i - \boldsymbol{\gamma}_i'\mathbf{F}'\bar{\mathbf{M}}_w\mathbf{X}_i\left(\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{X}_i\right)^{-1}\mathbf{X}_i'\bar{\mathbf{M}}_w\boldsymbol{\varepsilon}_i.$$

and

$$\boldsymbol{\varepsilon}_i'\bar{\mathbf{S}}_w\boldsymbol{\varepsilon}_i = \boldsymbol{\varepsilon}_i'\bar{\mathbf{M}}_w\boldsymbol{\varepsilon}_i - \boldsymbol{\varepsilon}_i'\bar{\mathbf{M}}_w\mathbf{X}_i\left(\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{X}_i\right)^{-1}\mathbf{X}_i'\bar{\mathbf{M}}_w\boldsymbol{\varepsilon}_i.$$

Using (5.21), (5.24) and (5.25), and noting also that

$$\frac{\mathbf{F}'\bar{\mathbf{M}}_w\mathbf{F}}{T} = O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right), \tag{B.5}$$

it follows that $T^{-1}\boldsymbol{\gamma}_i'\mathbf{F}'\bar{\mathbf{S}}_w\mathbf{F}\boldsymbol{\gamma}_i$ and $T^{-1}\boldsymbol{\gamma}_i'\mathbf{F}'\bar{\mathbf{S}}_w\boldsymbol{\varepsilon}_i$ are $O_p\left(N^{-1}\right) + O_p\left[(NT)^{-1/2}\right]$, and hence

$$\hat{\sigma}_i^2 = \frac{\boldsymbol{\varepsilon}_i'\bar{\mathbf{M}}_w\boldsymbol{\varepsilon}_i - \boldsymbol{\varepsilon}_i'\bar{\mathbf{M}}_w\mathbf{X}_i\left(\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{X}_i\right)^{-1}\mathbf{X}_i'\bar{\mathbf{M}}_w\boldsymbol{\varepsilon}_i}{T-(n+m+k)} + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right).$$

Also under the rank condition (4.3), using (5.24) and (5.25) we have

$$\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{X}_i}{T} = \frac{\mathbf{X}_i'\mathbf{M}_g\mathbf{X}_i}{T} + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right), \tag{B.6}$$

$$\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\boldsymbol{\varepsilon}_i}{T} = \frac{\mathbf{X}_i'\mathbf{M}_g\boldsymbol{\varepsilon}_i}{T} + O_p\left(\frac{1}{N}\right). \tag{B.7}$$

Similarly

$$\frac{\boldsymbol{\varepsilon}_i'\bar{\mathbf{M}}_w\boldsymbol{\varepsilon}_i}{T-(n+m+k)} = \frac{\boldsymbol{\varepsilon}_i'\mathbf{M}_g\boldsymbol{\varepsilon}_i}{T-(n+m+k)} + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right).$$

Hence

$$\hat{\sigma}_i^2 = \frac{\boldsymbol{\varepsilon}_i'\mathbf{M}_g\boldsymbol{\varepsilon}_i - \boldsymbol{\varepsilon}_i'\mathbf{M}_g\mathbf{X}_i\left(\mathbf{X}_i'\mathbf{M}_g\mathbf{X}_i\right)^{-1}\mathbf{X}_i'\mathbf{M}_g\boldsymbol{\varepsilon}_i}{T-(n+m+k)} + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right). \tag{B.8}$$

This result in conjunction with (B.6) now yields

$$\begin{aligned}\hat{\Sigma}_{T,b_i} &= T^{-1}\hat{\sigma}_i^2\left(\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{X}_i}{T}\right)^{-1} = T^{-1}\left(\frac{\boldsymbol{\varepsilon}_i'\mathbf{M}_g\boldsymbol{\varepsilon}_i - \boldsymbol{\varepsilon}_i'\mathbf{M}_g\mathbf{X}_i\left(\mathbf{X}_i'\mathbf{M}_g\mathbf{X}_i\right)^{-1}\mathbf{X}_i'\mathbf{M}_g\boldsymbol{\varepsilon}_i}{T-(n+m+k)}\right)\left(\frac{\mathbf{X}_i'\mathbf{M}_g\mathbf{X}_i}{T}\right)^{-1} \\ &\quad + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right).\end{aligned}$$

and, conditioning on $\mathbf{X}_i$ and $\mathbf{G}$, for a fixed $T$, we have

$$\lim_{N\to\infty} E(\hat{\Sigma}_{T,b_i}) = T^{-1}\sigma_i^2\left(\frac{\mathbf{X}_i'\mathbf{M}_g\mathbf{X}_i}{T}\right)^{-1}. \blacksquare$$

**Proof of Consistency of $\hat{\Sigma}_{b_i}$**

Using (5.12) in (B.6)

$$\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{X}_i}{T} = \frac{\mathbf{V}_i'\mathbf{M}_g\mathbf{V}_i}{T} + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right). \tag{B.9}$$

Also

$$\frac{\mathbf{V}_i'\mathbf{M}_g\mathbf{V}_i}{T} = \Sigma_i + O_p\left(\frac{1}{\sqrt{T}}\right), \tag{B.10}$$

and since $\mathring{\sigma}_i^2 = \hat{\sigma}_i^2 + O(T^{-1})$, from (B.8)

$$\mathring{\sigma}_i^2 = \sigma_i^2 + O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right). \tag{B.11}$$

Using this result and (B.9) in (5.35) we have

$$\mathring{\sigma}_i^2\left(\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{X}_i}{T}\right)^{-1} = \sigma_i^2\Sigma_i^{-1} + O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right),$$

as required. ■

**Proof of Theorem 6.2**

Under (2.1) and (2.2), $\hat{\mathbf{b}}_P$ defined by (6.49), can be written as

$$\left(\sum_{i=1}^N \theta_i^2\right)^{-1/2}\left(\hat{\mathbf{b}}_P - \boldsymbol{\beta}\right) = \left(\sum_{i=1}^N \theta_i\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{X}_i}{T}\right)^{-1}\left[\frac{1}{\sqrt{N}}\sum_{i=1}^N \tilde{\theta}_i\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w(\mathbf{X}_i\boldsymbol{\upsilon}_i + \boldsymbol{\varepsilon}_i)}{T} + \mathbf{q}_{NT}\right]. \tag{B.12}$$

where

$$\tilde{\theta}_i = \frac{\theta_i}{\sqrt{N^{-1}\sum_{i=1}^N \theta_i^2}} = O(1), \tag{B.13}$$

and

$$\mathbf{q}_{NT} = \frac{1}{\sqrt{N}}\sum_{i=1}^N \tilde{\theta}_i\frac{\left(\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{F}\right)\boldsymbol{\gamma}_i}{T}. \tag{B.14}$$

Using (2.10) we first note that $\boldsymbol{\gamma}_i = \bar{\gamma}_w + \boldsymbol{\eta}_i - \bar{\boldsymbol{\eta}}_w$, where $\bar{\boldsymbol{\eta}}_w = \sum_{i=1}^N w_i\eta_i$. Hence

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

Recall that $N^{-1}\sum_{i=1}^N \theta_i^2 = O(1)$, and in general when the rank condition is not satisfied $T^{-1}\left(\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{F}\right) = O_p(1)$. (See (5.24)). Hence, the first term of $\mathbf{q}_{NT}$ will be unbounded, unless $\theta_i = w_i$. But when this condition is satisfied, since $\bar{\mathbf{X}}_w'\bar{\mathbf{M}}_w = \mathbf{0}$, we have

$$\sum_{i=1}^N w_i\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{F}\left(\bar{\gamma}_w - \bar{\boldsymbol{\eta}}_w\right) = \bar{\mathbf{X}}_w'\bar{\mathbf{M}}_w\mathbf{F}\left(\bar{\gamma}_w - \bar{\boldsymbol{\eta}}_w\right) = \mathbf{0},$$

and using (5.24) it follows that

$$\begin{aligned}\mathbf{q}_{NT} &= \frac{1}{\sqrt{N}}\sum_{i=1}^N \tilde{w}_i\left(\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{F}}{T}\right)\boldsymbol{\eta}_i \qquad (B.15)\\ &= \frac{1}{\sqrt{N}}\sum_{i=1}^N \tilde{w}_i\left(\frac{\mathbf{X}_i'\bar{\mathbf{M}}_q\mathbf{F}}{T}\right)\boldsymbol{\eta}_i + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right),\end{aligned}$$

where $\tilde{w}_i = w_i/\left(N^{-1}\sum_{i=1}^N w_i^2\right)^{1/2}$. Substituting this result in (B.12), and making use of (5.24) and (5.25) we have

$$\begin{aligned}\left(\sum_{i=1}^N w_i^2\right)^{-1/2}\left(\hat{\mathbf{b}}_P - \boldsymbol{\beta}\right) &= \left(\sum_{i=1}^N w_i\frac{\mathbf{X}_i'\bar{\mathbf{M}}_q\mathbf{X}_i}{T}\right)^{-1}\left[\frac{1}{\sqrt{N}}\sum_{i=1}^N \tilde{w}_i\frac{\mathbf{X}_i'\bar{\mathbf{M}}_q(\mathbf{X}_i\boldsymbol{\upsilon}_i + \boldsymbol{\varepsilon}_i + \mathbf{F}\boldsymbol{\eta}_i)}{T}\right] + \\ &\quad O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right).\end{aligned} \tag{B.16}$$

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

$$\left(\sum_{i=1}^{N} w_i^2\right)^{-1/2} \left(\hat{\mathbf{b}} - \boldsymbol{\beta}\right) \xrightarrow{d} N(\mathbf{0}, \Sigma_P^*),$$

where

$$\Sigma_P^* = \Psi^{*-1} \mathbf{R}^* \Psi^{*-1}, \tag{B.17}$$

$$\Psi^* = \lim_{N\to\infty} \left(\sum_{i=1}^{N} w_i \Sigma_{iq}\right), \ \mathbf{R}^* = \lim_{N\to\infty} \left[N^{-1} \sum_{i=1}^{N} \tilde{w}_i^2 \left(\Sigma_{iq} \Omega_v \Sigma_{iq} + \mathbf{Q}_{if} \Omega_\eta \mathbf{Q}_{if}'\right)\right], \tag{B.18}$$

and $\Sigma_{iq}$ and $\mathbf{Q}_{if}$ are defined by (6.45). ■

**Proof of Theorem 6.3 (Pooled Homogeneous Slope)**

As in proof of Theorem 6.2, we first note that under $\theta_i = w_i$

$$\frac{1}{\sqrt{TN}} \sum_{i=1}^{N} \tilde{\theta}_i \left(\mathbf{X}_i' \bar{\mathbf{M}}_w \mathbf{F}\right) \boldsymbol{\gamma}_i = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \tilde{w}_i \left(\mathbf{X}_i' \bar{\mathbf{M}}_w \mathbf{F}\right) \boldsymbol{\eta}_i.$$

Also since the rank condition (4.3) is satisfied, using (4.4) we have

$$\mathbf{X}_i' \bar{\mathbf{M}}_w \mathbf{F} = -\left(\mathbf{X}_i' \bar{\mathbf{M}}_w \bar{\mathbf{U}}_w\right) \bar{\mathbf{C}}_w' \left(\bar{\mathbf{C}}_w \bar{\mathbf{C}}_w'\right)^{-1},$$

where $\bar{\mathbf{C}}_w' \left(\bar{\mathbf{C}}_w \bar{\mathbf{C}}_w'\right)^{-1}$ is bounded for all $N$. Hence (noting that here $\eta_i$ is an scaler):

$$\frac{1}{\sqrt{TN}} \sum_{i=1}^{N} \tilde{w}_i \left(\mathbf{X}_i' \bar{\mathbf{M}}_w \mathbf{F}\right) \boldsymbol{\eta}_i = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \tilde{w}_i \eta_i \left(\mathbf{X}_i' \bar{\mathbf{M}}_w \bar{\mathbf{U}}_w\right) \bar{\mathbf{C}}_w' \left(\bar{\mathbf{C}}_w \bar{\mathbf{C}}_w'\right)^{-1}.$$

But

$$\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \tilde{w}_i \eta_i \left(\mathbf{X}_i' \bar{\mathbf{M}}_w \bar{\mathbf{U}}_w\right) = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \tilde{w}_i \eta_i \left(\mathbf{X}_i' \bar{\mathbf{U}}_w\right) - \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \tilde{w}_i \eta_i \left(\frac{\mathbf{X}_i' \bar{\mathbf{H}}_w}{T}\right) \left(\frac{\bar{\mathbf{H}}_w' \bar{\mathbf{H}}_w}{T}\right)^{-1} \bar{\mathbf{H}}_w' \bar{\mathbf{U}}_w, \tag{B.19}$$

where $\bar{\mathbf{H}}_w$ is defined by (5.13). Writing the first term as

$$\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \tilde{w}_i \eta_i \left(\mathbf{X}_i' \bar{\mathbf{U}}_w\right) = \frac{1}{N} \sum_{i=1}^{N} \tilde{w}_i \eta_i \left(\frac{\sqrt{N} \mathbf{X}_i' \bar{\mathbf{U}}_w}{\sqrt{T}}\right),$$

and noting from (A.22) that $\left(\sqrt{N}\mathbf{X}_i' \bar{\mathbf{U}}_w / \sqrt{T}\right) = O_p\left(\sqrt{T/N}\right) + O_p(1)$, it readily follows that

$$\frac{1}{N} \sum_{i=1}^{N} \tilde{w}_i \eta_i \left(\frac{\sqrt{N} \mathbf{X}_i' \bar{\mathbf{U}}_w}{\sqrt{T}}\right) \xrightarrow{p} 0, \text{ as } N \to \infty, \text{ for all } T/N \to 0, \tag{B.20}$$

since $\tilde{w}_i = O(1)$, $\eta_i$ are IID and distributed independently of $\sqrt{N}\mathbf{X}_i' \bar{\mathbf{U}}_w / \sqrt{T}$, with the terms $\eta_i \left(\sqrt{N}\mathbf{X}_i' \bar{\mathbf{U}}_w / \sqrt{T}\right)$ having finite second order moments. Consider the second term of (B.19) and note that it can be written as

$$\frac{1}{N} \sum_{i=1}^{N} \tilde{w}_i \eta_i \left(\frac{\mathbf{X}_i' \bar{\mathbf{H}}_w}{T}\right) \left(\frac{\bar{\mathbf{H}}_w' \bar{\mathbf{H}}_w}{T}\right)^{-1} \left(\frac{\sqrt{N} \bar{\mathbf{H}}_w' \bar{\mathbf{U}}_w}{\sqrt{T}}\right).$$

Also

$$\frac{\bar{\mathbf{H}}_w' \bar{\mathbf{U}}_w}{T} = \frac{\bar{\mathbf{P}}_w' \mathbf{G}' \bar{\mathbf{U}}_w}{T} + \frac{\bar{\mathbf{U}}_w^{*\prime} \bar{\mathbf{U}}_w}{T} = \frac{\bar{\mathbf{P}}_w' \mathbf{G}' \bar{\mathbf{U}}_w}{T} + O_p\left(\frac{1}{N}\right),$$

which in conjunction with (5.16) and (5.17) yields

$$\begin{aligned} & \left(\frac{\mathbf{X}_i' \bar{\mathbf{H}}_w}{T}\right) \left(\frac{\bar{\mathbf{H}}_w' \bar{\mathbf{H}}_w}{T}\right)^{-1} \left(\frac{\sqrt{N} \bar{\mathbf{H}}_w' \bar{\mathbf{U}}_w}{\sqrt{T}}\right) \\ = & \left(\frac{\mathbf{X}_i' \mathbf{G}}{T}\right) \left(\frac{\mathbf{G}' \mathbf{G}}{T}\right)^{-1} \left(\frac{\sqrt{N} \mathbf{G}' \bar{\mathbf{U}}_w}{\sqrt{T}} + O_p\left(\sqrt{\frac{T}{N}}\right)\right). \end{aligned}$$

Therefore

$$\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{N}\tilde{w}_i\eta_i\left(\frac{\mathbf{X}_i'\bar{\mathbf{H}}_w}{T}\right)\left(\frac{\bar{\mathbf{H}}_w'\bar{\mathbf{H}}_w}{T}\right)^{-1}\left(\frac{\sqrt{N}\bar{\mathbf{H}}_w'\bar{\mathbf{U}}_w}{\sqrt{T}}\right)\\
= \;&\frac{1}{N}\sum_{i=1}^{N}\tilde{w}_i\eta_i\left(\frac{\mathbf{X}_i'\mathbf{G}}{T}\right)\left(\frac{\mathbf{G}'\mathbf{G}}{T}\right)^{-1}\left(\frac{\sqrt{N}\mathbf{G}'\bar{\mathbf{U}}_w}{\sqrt{T}}\right)+O_p\left(\sqrt{\frac{T}{N}}\right),
\end{aligned}$$

where $\sqrt{N/T}\mathbf{G}'\bar{\mathbf{U}}_w = O_p(1)$, and $\eta_i$ are IID and distributed independently of $\mathbf{G}$ and $\bar{\mathbf{U}}_w$. Hence, under the condition that $T/N \to 0$ as $(N,T) \xrightarrow{j} \infty$, we also obtain

$$\frac{1}{N}\sum_{i=1}^{N}\tilde{w}_i\eta_i\left(\frac{\mathbf{X}_i'\bar{\mathbf{H}}_w}{T}\right)\left(\frac{\bar{\mathbf{H}}_w'\bar{\mathbf{H}}_w}{T}\right)^{-1}\left(\frac{\sqrt{N}\bar{\mathbf{H}}_w'\bar{\mathbf{U}}_w}{\sqrt{T}}\right)\xrightarrow{p}\mathbf{0},$$

Using this result and (B.20) in (B.19) now yields

$$\frac{1}{\sqrt{TN}}\sum_{i=1}^{N}\tilde{w}_i\eta_i\left(\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{F}\right)=O_p\left(\sqrt{\frac{T}{N}}\right),$$

and

$$\left(\frac{\sum_{i=1}^{N}w_i^2}{T}\right)^{-1/2}\left(\hat{\mathbf{b}}_P-\boldsymbol{\beta}\right)=\Psi^{-1}\left[\frac{1}{\sqrt{TN}}\sum_{i=1}^{N}\tilde{w}_i\mathbf{X}_i'\bar{\mathbf{M}}_w\boldsymbol{\varepsilon}_i\right]+O_p\left(\sqrt{\frac{T}{N}}\right).$$

But since the rank condition (4.3) is satisfied, using (5.25) we have

$$\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\boldsymbol{\varepsilon}_i}{T}=\frac{\mathbf{X}_i'\mathbf{M}_g\boldsymbol{\varepsilon}_i}{T}+O_p\left(\frac{1}{N}\right),$$

and

$$\left(\frac{\sum_{i=1}^{N}w_i^2}{T}\right)^{-1/2}\left(\hat{\mathbf{b}}_P-\boldsymbol{\beta}\right)=\Psi^{-1}\left[\frac{1}{\sqrt{TN}}\sum_{i=1}^{N}\tilde{w}_i\mathbf{X}_i'\mathbf{M}_g\boldsymbol{\varepsilon}_i\right]+O_p\left(\sqrt{\frac{T}{N}}\right),$$

which establishes the validity of (6.61). ■

# References


[1] Ahn, S.G., Y-H. Lee and Schmidt, P. (2001), GMM Estimation of Linear Panel Data Models with Time-varying Individual Effects, *Journal of Econometrics*, 102, 219-255.

[2] Anselin, L. (2001), "Spatial Econometrics", in B. Baltagi (ed.), *A Companion to Theoretical Econometrics*, Blackwell, Oxford.

[3] Bai, J. and Ng, S. (2002), "Determining the Number of Factors in Approximate Factor Models", *Econometrica*, 70, 191-221.

[4] Binder, M., C. Hsiao, Pesaran, M.H. (2004), "Estimation and Inference in Short Panel Vector Autoregressions with Unit Roots and Cointegration", University of Cambridge DAE Working Paper No.0003, (revised).

[5] Coakley, J., Fuertes, A., and Smith, R.P. (2002), "A Principal Components Approach to Cross-Section Dependence in Panels", Unpublished manuscript, *Birkbeck College*, University of London.

[6] Conley, T.G. and Dupor, B. (2003), "A Spatial Analysis of Sectoral Complementarity", *Journal of Political Economy*, 111, 311-352.

[7] Conley, T.G. and Topa, G. (2002), "Socio-economic Distance and Spatial Patterns in Unemployment", *Journal of Applied Econometrics*, 17, 303 - 327.

[8] Forni, M., and Lippi, M. (1997), *Aggregation and the Microfoundations of Dynamic Macroeconomics*, Calrendon Press, Oxford.

[9] Forni, M., and Reichlin, L. (1998), "Let's Get Real: A Factor Analytical Approach to Disaggregated Business Cycle Dynamics", *Review of Economic Studies*, 65, 453-73.

[10] Holtz_Eakin, D, Newey, W.K., and Rosen, H., (1988), "Estimating Vector Autoregressions with Panel Data", *Econometrica*, 56, 1371-1395.

[11] Hsiao, C., Pesaran, M.H., and Tahmiscioglu, A.K. (1999), Bayes Estimation of Short-run coefficients in Dynamic Panel Data Models in C. Hsiao, K. Lahiri, L-F Lee and M.H. Pesaran (eds), *Analysis of Panels and Limited Dependent Variables: A Volume in Honour of G S Maddala*, Cambridge University Press, Cambridge.

[12] Kapetanios, G., Pesaran, M.H., and Yamagata, T., (2004), "Analysis of Panel Data Models with Unit Roots and a Multifactor Error Structure", under preparation.

[13] Kiefer, N.M. (1980), "A Time Series-Cross Section Model with Fixed Effects with an Intertemporal Factor Structure. Unpublished manuscript, Department of Economics, Cornell University.

[14] Lee, Y.H. (1991), "Panel Data Models with Multiplicative Individual and Time Effects_ Application to Compensation and Frontier Production Functions", Unpublished Ph.D. Dissertation, Mishigan State University.

[15] Lee, K.C., and Pesaran, M.H. (1993), "The Role of Sectoral Interactions in Wage Determination in the UK Economy", *The Economic Journal*, 103, 21-55.

[16] Pesaran, M.H. (2003) "A Simple Panel Unit Root Test in the Presence of Cross Section Dependence", Cambridge Working Papers in Economics No. 0346.

[17] Pesaran, M.H. (2004) "General Diagnostic Tests for Cross Section Dependence in Panels", CESifo Working Paper Series No. 1229; IZA Discussion Paper No. 1240.

[18] Pesaran, M.H. and Smith R.P. (1995) "Estimating Long-Run Relationships from Dynamic Heterogeneous Panels", *Journal of Econometrics*, 1995, 68, 79-113.

[19] Pesaran, M.H., Schuermann, T., and Weiner, S.M. (2004), "Modeling Regional Interdependencies using a Global Error-Correcting Macroeconomic Model", *Journal of Business Economics and Statistics* (with Discussions and a Rejoinder), 22, pp 129-181.

[20] Phillips, P.C.B., and Sul, D. (2003), "Dynamic Panel Estimation and Homogeneity Testing Under Cross Section Dependence", *The Econometrics Journal*, 6,217-259.

[21] Robertson, D. and Symons, J. (2000), "Factor Residuals in SUR Regressions: Estimating Panels Allowing for Cross Sectional Correlation", Unpublished manuscript, Faculty of Economics and Politics, University of Cambridge.

[22] Stock and Watson, M.W. (1998), "Diffusion Indexes", NBER Working Paper 6702.

[23] Swamy, P.A.V.B. (1970), Efficient Inference in Random Coefficient Regression Model, *Econometrica*, 38, 311-23.

**Table A1(i): Bias of Estimators of $\beta_1$**
**Experiment A1: Full Rank + Heterogeneous Slope**

| **CCE Type Estimators** | | | | | |
|---|---|---|---|---|---|
| **CCEMG** | T=20 | T=30 | T=50 | T=100 | T=200 |
| N=20 | -0.0012 | -0.0020 | -0.0014 | 0.0015 | 0.0019 |
| N=30 | 0.0003 | 0.0015 | 0.0004 | 0.0006 | -0.0004 |
| N=50 | -0.0022 | 0.0009 | 0.0000 | -0.0011 | -0.0004 |
| N=100 | -0.0001 | -0.0012 | -0.0001 | 0.0007 | 0.0011 |
| N=200 | 0.0001 | -0.0008 | -0.0003 | -0.0005 | 0.0003 |
| **CCEP** | | | | | |
| N=20 | -0.0001 | -0.0012 | -0.0011 | 0.0012 | 0.0021 |
| N=30 | -0.0001 | 0.0012 | 0.0009 | 0.0007 | -0.0006 |
| N=50 | -0.0011 | 0.0006 | 0.0002 | -0.0009 | -0.0003 |
| N=100 | -0.0004 | -0.0015 | -0.0004 | 0.0011 | 0.0013 |
| N=200 | -0.0004 | -0.0010 | -0.0001 | -0.0006 | 0.0002 |
| **Infeasible Estimators (including $f_{1t}$ and $f_{2t}$)** | | | | | |
| **Mean Group** | T=20 | T=30 | T=50 | T=100 | T=200 |
| N=20 | 0.0005 | -0.0010 | -0.0011 | 0.0007 | 0.0014 |
| N=30 | 0.0006 | -0.0010 | 0.0010 | 0.0004 | -0.0002 |
| N=50 | -0.0015 | 0.0007 | 0.0003 | -0.0006 | -0.0005 |
| N=100 | 0.0005 | -0.0005 | -0.0001 | 0.0006 | 0.0010 |
| N=200 | 0.0001 | -0.0007 | -0.0003 | -0.0004 | 0.0003 |
| **Pooled** | | | | | |
| N=20 | 0.0002 | 0.0002 | -0.0006 | 0.0003 | 0.0026 |
| N=30 | 0.0013 | -0.0005 | 0.0009 | 0.0006 | -0.0005 |
| N=50 | -0.0014 | 0.0006 | 0.0006 | -0.0011 | 0.0007 |
| N=100 | 0.0002 | -0.0001 | -0.0005 | 0.0002 | 0.0012 |
| N=200 | -0.0004 | -0.0008 | -0.0006 | -0.0004 | 0.0004 |
| **Naïve Estimators (excluding $f_{1t}$ and $f_{2t}$)** | | | | | |
| **Mean Group** | T=20 | T=30 | T=50 | T=100 | T=200 |
| N=20 | 0.1452 | 0.1449 | 0.1408 | 0.1445 | 0.1449 |
| N=30 | 0.1531 | 0.1494 | 0.1505 | 0.1498 | 0.1486 |
| N=50 | 0.1366 | 0.1391 | 0.1375 | 0.1347 | 0.1360 |
| N=100 | 0.1524 | 0.1518 | 0.1497 | 0.1482 | 0.1486 |
| N=200 | 0.1558 | 0.1524 | 0.1500 | 0.1488 | 0.1454 |
| **Pooled** | | | | | |
| N=20 | 0.1599 | 0.1636 | 0.1608 | 0.1666 | 0.1692 |
| N=30 | 0.1646 | 0.1668 | 0.1665 | 0.1689 | 0.1667 |
| N=50 | 0.1448 | 0.1489 | 0.1507 | 0.1502 | 0.1522 |
| N=100 | 0.1622 | 0.1636 | 0.1638 | 0.1648 | 0.1659 |
| N=200 | 0.1661 | 0.1660 | 0.1657 | 0.1672 | 0.1654 |

**Table A1(ii): Root Mean Squared Errors of Estimators of $\beta_1$**
**Experiment A1: Full Rank + Heterogeneous Slope**

| **CCE Type Estimators** | | | | | |
|---|---|---|---|---|---|
| **CCEMG** | T=20 | T=30 | T=50 | T=100 | T=200 |
| N=20 | 0.0947 | 0.0743 | 0.0619 | 0.0534 | 0.0499 |
| N=30 | 0.0779 | 0.0601 | 0.0496 | 0.0420 | 0.0390 |
| N=50 | 0.0587 | 0.0456 | 0.0375 | 0.0320 | 0.0300 |
| N=100 | 0.0419 | 0.0331 | 0.0268 | 0.0227 | 0.0212 |
| N=200 | 0.0308 | 0.0236 | 0.0192 | 0.0166 | 0.0148 |
| **CCEP** | | | | | |
| N=20 | 0.0880 | 0.0729 | 0.0625 | 0.0560 | 0.0520 |
| N=30 | 0.0698 | 0.0584 | 0.0502 | 0.0435 | 0.0405 |
| N=50 | 0.0526 | 0.0443 | 0.0378 | 0.0325 | 0.0305 |
| N=100 | 0.0367 | 0.0313 | 0.0268 | 0.0232 | 0.0214 |
| N=200 | 0.0269 | 0.0222 | 0.0191 | 0.0168 | 0.0150 |
| **Infeasible Estimators (including $f_{1t}$ and $f_{2t}$)** | | | | | |
| **Mean Group** | T=20 | T=30 | T=50 | T=100 | T=200 |
| N=20 | 0.0717 | 0.0603 | 0.0525 | 0.0481 | 0.0464 |
| N=30 | 0.0591 | 0.0502 | 0.0442 | 0.0391 | 0.0371 |
| N=50 | 0.0448 | 0.0382 | 0.0336 | 0.0301 | 0.0292 |
| N=100 | 0.0323 | 0.0274 | 0.0239 | 0.0217 | 0.0205 |
| N=200 | 0.0238 | 0.0192 | 0.0169 | 0.0157 | 0.0144 |
| **Pooled** | | | | | |
| N=20 | 0.0716 | 0.0627 | 0.0589 | 0.0546 | 0.0531 |
| N=30 | 0.0591 | 0.0532 | 0.0502 | 0.0463 | 0.0439 |
| N=50 | 0.0431 | 0.0408 | 0.0377 | 0.0358 | 0.0347 |
| N=100 | 0.0323 | 0.0296 | 0.0277 | 0.0262 | 0.0259 |
| N=200 | 0.0237 | 0.0210 | 0.0194 | 0.0183 | 0.0178 |
| **Naïve Estimators (excluding $f_{1t}$ and $f_{2t}$)** | | | | | |
| **Mean Group** | T=20 | T=30 | T=50 | T=100 | T=200 |
| N=20 | 0.1907 | 0.1795 | 0.1695 | 0.1659 | 0.1635 |
| N=30 | 0.1914 | 0.1803 | 0.1733 | 0.1667 | 0.1621 |
| N=50 | 0.1767 | 0.1670 | 0.1569 | 0.1489 | 0.1465 |
| N=100 | 0.1751 | 0.1676 | 0.1602 | 0.1556 | 0.1539 |
| N=200 | 0.1738 | 0.1657 | 0.1591 | 0.1543 | 0.1489 |
| **Pooled** | | | | | |
| N=20 | 0.2032 | 0.1970 | 0.1890 | 0.1876 | 0.1873 |
| N=30 | 0.2048 | 0.1979 | 0.1903 | 0.1867 | 0.1805 |
| N=50 | 0.1889 | 0.1816 | 0.1727 | 0.1651 | 0.1630 |
| N=100 | 0.1908 | 0.1832 | 0.1773 | 0.1738 | 0.1718 |
| N=200 | 0.1906 | 0.1836 | 0.1778 | 0.1739 | 0.1696 |

**Table A1(iii): Size of the test ($H_0 : \beta_1 = 1$) at 0.05 level**
**Experiment A1: Full Rank + Heterogeneous Slope**

| **CCE Type Estimators** | | | | | |
|---|---|---|---|---|---|
| **CCEMG** | T=20 | T=30 | T=50 | T=100 | T=200 |
| N=20 | 0.071 | 0.076 | 0.076 | 0.078 | 0.081 |
| N=30 | 0.067 | 0.063 | 0.067 | 0.061 | 0.066 |
| N=50 | 0.046 | 0.058 | 0.068 | 0.057 | 0.056 |
| N=100 | 0.043 | 0.057 | 0.057 | 0.054 | 0.054 |
| N=200 | 0.054 | 0.056 | 0.058 | 0.055 | 0.048 |
| **CCEP(hetero)** | | | | | |
| N=20 | 0.076 | 0.079 | 0.078 | 0.085 | 0.079 |
| N=30 | 0.074 | 0.069 | 0.066 | 0.061 | 0.067 |
| N=50 | 0.063 | 0.060 | 0.065 | 0.054 | 0.053 |
| N=100 | 0.052 | 0.055 | 0.057 | 0.055 | 0.051 |
| N=200 | 0.046 | 0.050 | 0.050 | 0.060 | 0.045 |
| **Infeasible Estimators (including $f_{1t}$ and $f_{2t}$)** | | | | | |
| **Mean Group** | T=20 | T=30 | T=50 | T=100 | T=200 |
| N=20 | 0.071 | 0.069 | 0.062 | 0.062 | 0.061 |
| N=30 | 0.063 | 0.063 | 0.055 | 0.060 | 0.059 |
| N=50 | 0.055 | 0.064 | 0.055 | 0.055 | 0.053 |
| N=100 | 0.048 | 0.064 | 0.055 | 0.052 | 0.046 |
| N=200 | 0.060 | 0.049 | 0.054 | 0.055 | 0.049 |
| **Pooled(hetero)** | | | | | |
| N=20 | 0.072 | 0.066 | 0.070 | 0.067 | 0.068 |
| N=30 | 0.062 | 0.055 | 0.064 | 0.060 | 0.053 |
| N=50 | 0.053 | 0.053 | 0.066 | 0.054 | 0.055 |
| N=100 | 0.053 | 0.055 | 0.058 | 0.055 | 0.060 |
| N=200 | 0.058 | 0.048 | 0.048 | 0.052 | 0.053 |
| **Naïve Estimators (excluding $f_{1t}$ and $f_{2t}$)** | | | | | |
| **Mean Group** | T=20 | T=30 | T=50 | T=100 | T=200 |
| N=20 | 0.314 | 0.369 | 0.406 | 0.476 | 0.499 |
| N=30 | 0.426 | 0.478 | 0.547 | 0.621 | 0.672 |
| N=50 | 0.548 | 0.603 | 0.679 | 0.738 | 0.795 |
| N=100 | 0.754 | 0.843 | 0.921 | 0.961 | 0.985 |
| N=200 | 0.888 | 0.932 | 0.977 | 0.997 | 1.000 |
| **Pooled(hetero)** | | | | | |
| N=20 | 0.389 | 0.441 | 0.466 | 0.535 | 0.593 |
| N=30 | 0.519 | 0.572 | 0.624 | 0.679 | 0.720 |
| N=50 | 0.613 | 0645 | 0.718 | 0.784 | 0.838 |
| N=100 | 0.781 | 0.860 | 0.924 | 0.964 | 0.992 |
| N=200 | 0.896 | 0.940 | 0.976 | 0.995 | 1.000 |

**Table A1(iv): Power of the test ($H_0 : \beta_1 = 0.95$) at 0.05 level**
**Experiment A1: Full Rank + Heterogeneous Slope**

| **CCE Type Estimators** | | | | | |
|---|---|---|---|---|---|
| **CCEMG** | T=20 | T=30 | T=50 | T=100 | T=200 |
| N=20 | 0.108 | 0.128 | 0.166 | 0.196 | 0.242 |
| N=30 | 0.123 | 0.160 | 0.208 | 0.249 | 0.268 |
| N=50 | 0.149 | 0.214 | 0.295 | 0.350 | 0.388 |
| N=100 | 0.223 | 0.330 | 0.470 | 0.606 | 0.677 |
| N=200 | 0.375 | 0.543 | 0.749 | 0.855 | 0.917 |
| **CCEP(hetero)** | | | | | |
| N=20 | 0.132 | 0.138 | 0.168 | 0.199 | 0.231 |
| N=30 | 0.140 | 0.171 | 0.214 | 0.241 | 0.257 |
| N=50 | 0.176 | 0.232 | 0.288 | 0.332 | 0.377 |
| N=100 | 0.262 | 0.349 | 0.458 | 0.595 | 0.670 |
| N=200 | 0.449 | 0.592 | 0.742 | 0.843 | 0.910 |
| **Infeasible Estimators (including $f_{1t}$ and $f_{2t}$)** | | | | | |
| **Mean Group** | T=20 | T=30 | T=50 | T=100 | T=200 |
| N=20 | 0.133 | 0.144 | 0.165 | 0.203 | 0.227 |
| N=30 | 0.148 | 0.183 | 0.238 | 0.244 | 0.256 |
| N=50 | 0.211 | 0.293 | 0.342 | 0.376 | 0.397 |
| N=100 | 0.361 | 0.445 | 0.557 | 0.640 | 0.693 |
| N=200 | 0.574 | 0.721 | 0.835 | 0.897 | 0.936 |
| **Pooled(hetero)** | | | | | |
| N=20 | 0.135 | 0.145 | 0.165 | 0.174 | 0.197 |
| N=30 | 0.166 | 0.168 | 0.193 | 0.219 | 0.216 |
| N=50 | 0.213 | 0.271 | 0.284 | 0.294 | 0.322 |
| N=100 | 0.345 | 0.401 | 0.442 | 0.511 | 0.520 |
| N=200 | 0.574 | 0.658 | 0.721 | 0.770 | 0.808 |
| **Naïve Estimators (excluding $f_{1t}$ and $f_{2t}$)** | | | | | |
| **Mean Group** | T=20 | T=30 | T=50 | T=100 | T=200 |
| N=20 | 0.464 | 0.544 | 0.618 | 0.691 | 0.745 |
| N=30 | 0.618 | 0.671 | 0.759 | 0.834 | 0.882 |
| N=50 | 0.713 | 0.806 | 0.868 | 0.923 | 0.961 |
| N=100 | 0.914 | 0.964 | 0.992 | 0.999 | 1.000 |
| N=200 | 0.975 | 0.991 | 0.999 | 1.000 | 1.000 |
| **Pooled(hetero)** | | | | | |
| N=20 | 0.563 | 0.613 | 0.658 | 0.755 | 0.792 |
| N=30 | 0.694 | 0.745 | 0.800 | 0.856 | 0.910 |
| N=50 | 0.759 | 0.819 | 0.870 | 0.927 | 0.964 |
| N=100 | 0.920 | 0.957 | 0.987 | 0.998 | 0.999 |
| N=200 | 0.967 | 0.987 | 0.998 | 1.000 | 1.000 |

**Table A2(i): Root Mean Squared Errors of Estimators of $\beta_1$**
**Experiment A2: Full Rank + Homogeneous Slope**

| **CCE Type Estimators** | | | | | |
|---|---|---|---|---|---|
| **CCEMG** | T=20 | T=30 | T=50 | T=100 | T=200 |
| N=20 | 0.0837 | 0.0604 | 0.0413 | 0.0285 | 0.0218 |
| N=30 | 0.0698 | 0.0467 | 0.0342 | 0.0221 | 0.0163 |
| N=50 | 0.0525 | 0.0350 | 0.0235 | 0.0162 | 0.0114 |
| N=100 | 0.0374 | 0.0261 | 0.0183 | 0.0112 | 0.0075 |
| N=200 | 0.0282 | 0.0190 | 0.0129 | 0.0080 | 0.0054 |
| **CCEP** | | | | | |
| N=20 | 0.0705 | 0.0535 | 0.0398 | 0.0292 | 0.0228 |
| N=30 | 0.0562 | 0.0408 | 0.0308 | 0.0214 | 0.0169 |
| N=50 | 0.0417 | 0.0303 | 0.0222 | 0.0156 | 0.0115 |
| N=100 | 0.0297 | 0.0223 | 0.0167 | 0.0106 | 0.0074 |
| N=200 | 0.0219 | 0.0158 | 0.0115 | 0.0076 | 0.0053 |
| **Infeasible Estimators (including $f_{1t}$ and $f_{2t}$)** | | | | | |
| **Mean Group** | T=20 | T=30 | T=50 | T=100 | T=200 |
| N=20 | 0.0563 | 0.0406 | 0.0290 | 0.0192 | 0.0138 |
| N=30 | 0.0472 | 0.0340 | 0.0241 | 0.0155 | 0.0107 |
| N=50 | 0.0335 | 0.0240 | 0.0166 | 0.0111 | 0.0077 |
| N=100 | 0.0261 | 0.0183 | 0.0134 | 0.0086 | 0.0058 |
| N=200 | 0.0185 | 0.0132 | 0.0095 | 0.0061 | 0.0042 |
| **Pooled** | | | | | |
| N=20 | 0.0424 | 0.0315 | 0.0242 | 0.0164 | 0.0115 |
| N=30 | 0.0341 | 0.0268 | 0.0193 | 0.0130 | 0.0090 |
| N=50 | 0.0235 | 0.0182 | 0.0133 | 0.0090 | 0.0063 |
| N=100 | 0.0184 | 0.0140 | 0.0105 | 0.0070 | 0.0048 |
| N=200 | 0.0128 | 0.0097 | 0.0074 | 0.0048 | 0.0034 |

**Table A2(ii): Size of the test ($H_0 : \beta_1 = 1$) at 0.05 level**
**Experiment A2: Full Rank + Homogeneous Slope**

| **CCE Type Estimators** | | | | | |
|---|---|---|---|---|---|
| **CCEMG** | T=20 | T=30 | T=50 | T=100 | T=200 |
| N=20 | 0.071 | 0.065 | 0.075 | 0.065 | 0.071 |
| N=30 | 0.063 | 0.056 | 0.078 | 0.065 | 0.063 |
| N=50 | 0.059 | 0.047 | 0.044 | 0.057 | 0.056 |
| N=100 | 0.051 | 0.053 | 0.059 | 0.058 | 0.047 |
| N=200 | 0.055 | 0.052 | 0.052 | 0.051 | 0.053 |
| **CCEP** | | | | | |
| N=20 | 0.076 | 0.078 | 0.094 | 0.109 | 0.160 |
| N=30 | 0.068 | 0.056 | 0.074 | 0.088 | 0.130 |
| N=50 | 0.066 | 0.046 | 0.059 | 0.076 | 0.086 |
| N=100 | 0.055 | 0.053 | 0.067 | 0.053 | 0.051 |
| N=200 | 0.059 | 0.046 | 0.050 | 0.050 | 0.053 |
| **CCEP(hetero)** | | | | | |
| N=20 | 0.076 | 0.076 | 0.074 | 0.068 | 0.074 |
| N=30 | 0.072 | 0.053 | 0.066 | 0.061 | 0.068 |
| N=50 | 0.062 | 0.050 | 0.054 | 0.059 | 0.061 |
| N=100 | 0.056 | 0.057 | 0.066 | 0.056 | 0.045 |
| N=200 | 0.059 | 0.047 | 0.048 | 0.047 | 0.053 |
| **Infeasible Estimators (including $f_{1t}$ and $f_{2t}$)** | | | | | |
| **Mean Group** | T=20 | T=30 | T=50 | T=100 | T=200 |
| N=20 | 0.061 | 0.064 | 0.068 | 0.066 | 0.074 |
| N=30 | 0.053 | 0.063 | 0.061 | 0.066 | 0.059 |
| N=50 | 0.051 | 0.049 | 0.045 | 0.049 | 0.053 |
| N=100 | 0.050 | 0.052 | 0.058 | 0.055 | 0.051 |
| N=200 | 0.047 | 0.048 | 0.054 | 0.050 | 0.046 |
| **Pooled** | | | | | |
| N=20 | 0.055 | 0.040 | 0.057 | 0.053 | 0.054 |
| N=30 | 0.053 | 0.051 | 0.057 | 0.055 | 0.050 |
| N=50 | 0.051 | 0.048 | 0.048 | 0.041 | 0.048 |
| N=100 | 0.050 | 0.049 | 0.052 | 0.052 | 0.051 |
| N=200 | 0.047 | 0.049 | 0.057 | 0.045 | 0.049 |

**Table A2(iii): Power of the test ($H_0 : \beta_1 = 0.95$) at 0.05 level**
**Experiment A2: Full Rank + Homogeneous Slope**

| **CCE Type Estimators** | | | | | |
|---|---|---|---|---|---|
| **CCEMG** | T=20 | T=30 | T=50 | T=100 | T=200 |
| N=20 | 0.121 | 0.175 | 0.268 | 0.455 | 0.684 |
| N=30 | 0.141 | 0.211 | 0.366 | 0.616 | 0.864 |
| N=50 | 0.187 | 0.301 | 0.539 | 0.870 | 0.988 |
| N=100 | 0.277 | 0.505 | 0.806 | 0.993 | 1.000 |
| N=200 | 0.444 | 0.757 | 0.979 | 1.000 | 1.000 |
| **CCEP(hetero)** | | | | | |
| N=20 | 0.141 | 0.194 | 0.303 | 0.466 | 0.654 |
| N=30 | 0.177 | 0.235 | 0.402 | 0.631 | 0.853 |
| N=50 | 0.243 | 0.367 | 0.620 | 0.890 | 0.986 |
| N=100 | 0.396 | 0.629 | 0.872 | 0.996 | 1.000 |
| N=200 | 0.650 | 0.882 | 0.994 | 1.000 | 1.000 |
| **Infeasible Estimators (including $f_{1t}$ and $f_{2t}$)** | | | | | |
| **Mean Group** | T=20 | T=30 | T=50 | T=100 | T=200 |
| N=20 | 0.180 | 0.265 | 0.445 | 0.740 | 0.955 |
| N=30 | 0.222 | 0.338 | 0.580 | 0.886 | 0.995 |
| N=50 | 0.330 | 0.558 | 0.831 | 0.991 | 1.000 |
| N=100 | 0.508 | 0.779 | 0.964 | 1.000 | 1.000 |
| N=200 | 0.772 | 0.959 | 1.000 | 1.000 | 1.000 |
| **Pooled** | | | | | |
| N=20 | 0.230 | 0.334 | 0.562 | 0.873 | 0.994 |
| N=30 | 0.325 | 0.483 | 0.763 | 0.970 | 1.000 |
| N=50 | 0.566 | 0.788 | 0.963 | 1.000 | 1.000 |
| N=100 | 0.778 | 0.950 | 0.998 | 1.000 | 1.000 |
| N=200 | 0.968 | 0.998 | 1.000 | 1.000 | 1.000 |

**Table B1(i): Bias of Estimators of $\beta_1$**
**Experiment B1: Rank Deficient + Heterogeneous Slope**

| **CCE Type Estimators** | | | | | |
|---|---|---|---|---|---|
| **CCEMG** | T=20 | T=30 | T=50 | T=100 | T=200 |
| N=20 | -0.0012 | -0.0012 | 0.0000 | -0.0009 | 0.0014 |
| N=30 | 0.0002 | -0.0002 | 0.0011 | 0.0001 | -0.0005 |
| N=50 | -0.0011 | 0.0012 | 0.0006 | -0.0004 | -0.0010 |
| N=100 | -0.0007 | -0.0017 | -0.0006 | -0.0002 | 0.0013 |
| N=200 | 0.0002 | -0.0010 | -0.0004 | -0.0003 | 0.0005 |
| **CCEP** | | | | | |
| N=20 | -0.0002 | -0.0008 | 0.0005 | -0.0014 | 0.0015 |
| N=30 | -0.0003 | -0.0006 | 0.0019 | 0.0006 | -0.0007 |
| N=50 | -0.0001 | 0.0011 | 0.0006 | -0.0003 | -0.0010 |
| N=100 | -0.0002 | -0.0015 | -0.0009 | 0.0002 | 0.0015 |
| N=200 | -0.0007 | -0.0005 | -0.0002 | -0.0004 | 0.0003 |
| **Infeasible Estimators (including $f_{1t}$ and $f_{2t}$)** | | | | | |
| **Mean Group** | T=20 | T=30 | T=50 | T=100 | T=200 |
| N=20 | 0.0005 | -0.0010 | -0.0011 | 0.0007 | 0.0014 |
| N=30 | 0.0006 | -0.0010 | 0.0010 | 0.0004 | -0.0002 |
| N=50 | -0.0015 | 0.0007 | 0.0003 | -0.0006 | -0.0005 |
| N=100 | 0.0005 | -0.0005 | -0.0001 | 0.0006 | 0.0010 |
| N=200 | 0.0001 | -0.0007 | -0.0003 | -0.0004 | 0.0003 |
| **Pooled** | | | | | |
| N=20 | 0.0002 | 0.0002 | -0.0006 | 0.0003 | 0.0026 |
| N=30 | 0.0013 | -0.0005 | 0.0009 | 0.0006 | -0.0005 |
| N=50 | -0.0014 | 0.0006 | 0.0006 | -0.0011 | 0.0007 |
| N=100 | 0.0002 | -0.0001 | -0.0005 | 0.0002 | 0.0012 |
| N=200 | -0.0004 | -0.0008 | -0.0006 | -0.0004 | 0.0004 |

**Table B1(ii): Root Mean Squared Errors of Estimators of $\beta_1$**
**Experiment B1: Rank Deficient + Heterogeneous Slope**

| **CCE Type Estimators** | | | | | |
|---|---|---|---|---|---|
| **CCEMG** | T=20 | T=30 | T=50 | T=100 | T=200 |
| N=20 | 0.1212 | 0.0932 | 0.0742 | 0.0610 | 0.0551 |
| N=30 | 0.1034 | 0.0773 | 0.0615 | 0.0500 | 0.0435 |
| N=50 | 0.0764 | 0.0603 | 0.0467 | 0.0376 | 0.0329 |
| N=100 | 0.0546 | 0.0435 | 0.0330 | 0.0269 | 0.0236 |
| N=200 | 0.0401 | 0.0310 | 0.0237 | 0.0192 | 0.0161 |
| **CCEP** | | | | | |
| N=20 | 0.1068 | 0.0873 | 0.0724 | 0.0623 | 0.0561 |
| N=30 | 0.0895 | 0.0736 | 0.0599 | 0.0508 | 0.0445 |
| N=50 | 0.0663 | 0.0560 | 0.0455 | 0.0373 | 0.0334 |
| N=100 | 0.0467 | 0.0394 | 0.0320 | 0.0272 | 0.0235 |
| N=200 | 0.0333 | 0.0282 | 0.0228 | 0.0191 | 0.0163 |
| **Infeasible Estimators (including $f_{1t}$ and $f_{2t}$)** | | | | | |
| **Mean Group** | T=20 | T=30 | T=50 | T=100 | T=200 |
| N=20 | 0.0717 | 0.0603 | 0.0525 | 0.0481 | 0.0464 |
| N=30 | 0.0591 | 0.0502 | 0.0442 | 0.0391 | 0.0371 |
| N=50 | 0.0448 | 0.0382 | 0.0336 | 0.0301 | 0.0292 |
| N=100 | 0.0323 | 0.0274 | 0.0239 | 0.0217 | 0.0205 |
| N=200 | 0.0238 | 0.0192 | 0.0169 | 0.0157 | 0.0144 |
| **Pooled** | | | | | |
| N=20 | 0.0716 | 0.0627 | 0.0589 | 0.0546 | 0.0531 |
| N=30 | 0.0591 | 0.0532 | 0.0502 | 0.0463 | 0.0439 |
| N=50 | 0.0431 | 0.0408 | 0.0377 | 0.0358 | 0.0347 |
| N=100 | 0.0323 | 0.0296 | 0.0277 | 0.0262 | 0.0259 |
| N=200 | 0.0237 | 0.0210 | 0.0194 | 0.0183 | 0.0178 |

**Table B1(iii): Size of the test ($H_0 : \beta_1 = 1$) at 0.05 level**
**Experiment B1: Rank Deficient + Heterogeneous Slope**

| **CCE Type Estimators** | | | | | |
|---|---|---|---|---|---|
| **CCEMG** | T=20 | T=30 | T=50 | T=100 | T=200 |
| N=20 | 0.070 | 0.070 | 0.073 | 0.075 | 0.075 |
| N=30 | 0.067 | 0.061 | 0.065 | 0.065 | 0.063 |
| N=50 | 0.048 | 0.065 | 0.064 | 0.056 | 0.056 |
| N=100 | 0.051 | 0.055 | 0.053 | 0.058 | 0.058 |
| N=200 | 0.050 | 0.058 | 0.050 | 0.059 | 0.047 |
| **CCEP(hetero)** | | | | | |
| N=20 | 0.074 | 0.077 | 0.072 | 0.084 | 0.077 |
| N=30 | 0.066 | 0.072 | 0.065 | 0.069 | 0.066 |
| N=50 | 0.060 | 0.062 | 0.064 | 0.053 | 0.056 |
| N=100 | 0.056 | 0.052 | 0.053 | 0.058 | 0.057 |
| N=200 | 0.044 | 0.059 | 0.054 | 0.057 | 0.048 |
| **Infeasible Estimators (including $f_{1t}$ and $f_{2t}$)** | | | | | |
| **Mean Group** | T=20 | T=30 | T=50 | T=100 | T=200 |
| N=20 | 0.071 | 0.069 | 0.062 | 0.062 | 0.061 |
| N=30 | 0.063 | 0.063 | 0.055 | 0.060 | 0.059 |
| N=50 | 0.055 | 0.064 | 0.055 | 0.055 | 0.053 |
| N=100 | 0.048 | 0.064 | 0.055 | 0.052 | 0.046 |
| N=200 | 0.060 | 0.049 | 0.054 | 0.055 | 0.049 |
| **Pooled(hetero)** | | | | | |
| N=20 | 0.072 | 0.066 | 0.070 | 0.067 | 0.068 |
| N=30 | 0.062 | 0.055 | 0.064 | 0.060 | 0.053 |
| N=50 | 0.053 | 0.053 | 0.066 | 0.054 | 0.055 |
| N=100 | 0.053 | 0.055 | 0.058 | 0.055 | 0.060 |
| N=200 | 0.058 | 0.048 | 0.048 | 0.052 | 0.053 |

**Table B1(iv): Power of the test ($H_0 : \beta_1 = 0.95$) at 0.05 level**
**Experiment B1: Rank Deficient + Heterogeneous Slope**

| **CCE Type Estimators** | | | | | |
|---|---|---|---|---|---|
| **CCEMG** | T=20 | T=30 | T=50 | T=100 | T=200 |
| N=20 | 0.093 | 0.106 | 0.143 | 0.176 | 0.224 |
| N=30 | 0.108 | 0.128 | 0.165 | 0.203 | 0.234 |
| N=50 | 0.114 | 0.172 | 0.225 | 0.284 | 0.332 |
| N=100 | 0.152 | 0.238 | 0.334 | 0.478 | 0.598 |
| N=200 | 0.254 | 0.382 | 0.560 | 0.739 | 0.866 |
| **CCEP(hetero)** | | | | | |
| N=20 | 0.114 | 0.111 | 0.146 | 0.167 | 0.212 |
| N=30 | 0.111 | 0.138 | 0.166 | 0.205 | 0.233 |
| N=50 | 0.137 | 0.187 | 0.233 | 0.283 | 0.323 |
| N=100 | 0.186 | 0.259 | 0.338 | 0.480 | 0.600 |
| N=200 | 0.315 | 0.453 | 0.586 | 0.739 | 0.859 |
| **Infeasible Estimators (including $f_{1t}$ and $f_{2t}$)** | | | | | |
| **Mean Group** | T=20 | T=30 | T=50 | T=100 | T=200 |
| N=20 | 0.133 | 0.144 | 0.165 | 0.203 | 0.227 |
| N=30 | 0.148 | 0.183 | 0.238 | 0.244 | 0.256 |
| N=50 | 0.211 | 0.293 | 0.342 | 0.376 | 0.397 |
| N=100 | 0.361 | 0.445 | 0.557 | 0.640 | 0.693 |
| N=200 | 0.574 | 0.721 | 0.835 | 0.897 | 0.936 |
| **Pooled(hetero)** | | | | | |
| N=20 | 0.135 | 0.145 | 0.165 | 0.174 | 0.197 |
| N=30 | 0.166 | 0.168 | 0.193 | 0.219 | 0.216 |
| N=50 | 0.213 | 0.271 | 0.284 | 0.294 | 0.322 |
| N=100 | 0.345 | 0.401 | 0.442 | 0.511 | 0.520 |
| N=200 | 0.574 | 0.658 | 0.721 | 0.770 | 0.808 |

**Table B2(i): Root Mean Squared Errors of Estimators of $\beta_1$**
**Experiment B2: Rank Deficient + Homogeneous Slope**

| **CCE Type Estimators** | | | | | |
|---|---|---|---|---|---|
| **CCEMG** | T=20 | T=30 | T=50 | T=100 | T=200 |
| N=20 | 0.1142 | 0.0832 | 0.0587 | 0.0410 | 0.0304 |
| N=30 | 0.0957 | 0.0691 | 0.0501 | 0.0335 | 0.0245 |
| N=50 | 0.0710 | 0.0524 | 0.0361 | 0.0251 | 0.0178 |
| N=100 | 0.0537 | 0.0382 | 0.0273 | 0.0172 | 0.0122 |
| N=200 | 0.0394 | 0.0276 | 0.0192 | 0.0127 | 0.0085 |
| **CCEP** | | | | | |
| N=20 | 0.0924 | 0.0746 | 0.0548 | 0.0401 | 0.0307 |
| N=30 | 0.0777 | 0.0603 | 0.0459 | 0.0323 | 0.0244 |
| N=50 | 0.0579 | 0.0454 | 0.0341 | 0.0244 | 0.0177 |
| N=100 | 0.0429 | 0.0328 | 0.0245 | 0.0164 | 0.0120 |
| N=200 | 0.0308 | 0.0237 | 0.0173 | 0.0120 | 0.0084 |
| **Infeasible Estimators (including $f_{1t}$ and $f_{2t}$)** | | | | | |
| **Mean Group** | T=20 | T=30 | T=50 | T=100 | T=200 |
| N=20 | 0.0563 | 0.0406 | 0.0290 | 0.0192 | 0.0138 |
| N=30 | 0.0472 | 0.0340 | 0.0241 | 0.0155 | 0.0107 |
| N=50 | 0.0335 | 0.0240 | 0.0166 | 0.0111 | 0.0077 |
| N=100 | 0.0261 | 0.0183 | 0.0134 | 0.0086 | 0.0058 |
| N=200 | 0.0185 | 0.0132 | 0.0095 | 0.0061 | 0.0042 |
| **Pooled** | | | | | |
| N=20 | 0.0424 | 0.0315 | 0.0242 | 0.0164 | 0.0115 |
| N=30 | 0.0341 | 0.0268 | 0.0193 | 0.0130 | 0.0090 |
| N=50 | 0.0235 | 0.0182 | 0.0133 | 0.0090 | 0.0063 |
| N=100 | 0.0184 | 0.0140 | 0.0105 | 0.0070 | 0.0048 |
| N=200 | 0.0128 | 0.0097 | 0.0074 | 0.0048 | 0.0034 |

**Table B2(ii): Size of the test ($H_0 : \beta_1 = 1$) at 0.05 level**
**Experiment B2: Rank Deficient + Homogeneous Slope**

| **CCE Type Estimators** | | | | | |
|---|---|---|---|---|---|
| **CCEMG** | T=20 | T=30 | T=50 | T=100 | T=200 |
| N=20 | 0.069 | 0.076 | 0.068 | 0.062 | 0.069 |
| N=30 | 0.059 | 0.063 | 0.069 | 0.063 | 0.061 |
| N=50 | 0.053 | 0.057 | 0.044 | 0.065 | 0.060 |
| N=100 | 0.055 | 0.060 | 0.061 | 0.055 | 0.051 |
| N=200 | 0.062 | 0.052 | 0.050 | 0.047 | 0.052 |
| **CCEP** | | | | | |
| N=20 | 0.076 | 0.096 | 0.093 | 0.131 | 0.163 |
| N=30 | 0.079 | 0.085 | 0.098 | 0.117 | 0.165 |
| N=50 | 0.068 | 0.078 | 0.093 | 0.111 | 0.132 |
| N=100 | 0.075 | 0.083 | 0.097 | 0.092 | 0.110 |
| N=200 | 0.073 | 0.080 | 0.081 | 0.084 | 0.099 |
| **CCEP(hetero)** | | | | | |
| N=20 | 0.066 | 0.072 | 0.065 | 0.062 | 0.070 |
| N=30 | 0.066 | 0.060 | 0.060 | 0.063 | 0.068 |
| N=50 | 0.056 | 0.057 | 0.054 | 0.066 | 0.055 |
| N=100 | 0.062 | 0.057 | 0.064 | 0.049 | 0.055 |
| N=200 | 0.061 | 0.054 | 0.053 | 0.049 | 0.054 |
| **Infeasible Estimators (including $f_{1t}$ and $f_{2t}$)** | | | | | |
| **Mean Group** | T=20 | T=30 | T=50 | T=100 | T=200 |
| N=20 | 0.061 | 0.064 | 0.068 | 0.066 | 0.074 |
| N=30 | 0.053 | 0.063 | 0.061 | 0.066 | 0.059 |
| N=50 | 0.051 | 0.049 | 0.045 | 0.049 | 0.053 |
| N=100 | 0.050 | 0.052 | 0.058 | 0.055 | 0.051 |
| N=200 | 0.047 | 0.048 | 0.054 | 0.050 | 0.046 |
| **Pooled** | | | | | |
| N=20 | 0.055 | 0.040 | 0.057 | 0.053 | 0.054 |
| N=30 | 0.053 | 0.051 | 0.057 | 0.055 | 0.050 |
| N=50 | 0.051 | 0.048 | 0.048 | 0.041 | 0.048 |
| N=100 | 0.050 | 0.049 | 0.052 | 0.052 | 0.051 |
| N=200 | 0.047 | 0.049 | 0.057 | 0.045 | 0.049 |

**Table B2(iii): Power of the test ($H_0 : \beta_1 = 0.95$) at 0.05 level**
**Experiment B2: Rank Deficient + Homogeneous Slope**

| **CCE Type Estimators** | | | | | |
|---|---|---|---|---|---|
| **CCEMG** | T=20 | T=30 | T=50 | T=100 | T=200 |
| N=20 | 0.104 | 0.134 | 0.192 | 0.302 | 0.454 |
| N=30 | 0.102 | 0.147 | 0.219 | 0.375 | 0.588 |
| N=50 | 0.121 | 0.186 | 0.312 | 0.564 | 0.820 |
| N=100 | 0.187 | 0.317 | 0.502 | 0.830 | 0.972 |
| N=200 | 0.284 | 0.471 | 0.784 | 0.977 | 0.999 |
| **CCEP(hetero)** | | | | | |
| N=20 | 0.114 | 0.151 | 0.209 | 0.309 | 0.465 |
| N=30 | 0.124 | 0.165 | 0.232 | 0.392 | 0.594 |
| N=50 | 0.171 | 0.227 | 0.350 | 0.596 | 0.824 |
| N=100 | 0.247 | 0.392 | 0.564 | 0.859 | 0.982 |
| N=200 | 0.418 | 0.582 | 0.847 | 0.984 | 0.999 |
| **Infeasible Estimators (including $f_{1t}$ and $f_{2t}$)** | | | | | |
| **Mean Group** | T=20 | T=30 | T=50 | T=100 | T=200 |
| N=20 | 0.180 | 0.265 | 0.445 | 0.740 | 0.955 |
| N=30 | 0.222 | 0.338 | 0.580 | 0.886 | 0.995 |
| N=50 | 0.330 | 0.558 | 0.831 | 0.991 | 1.000 |
| N=100 | 0.508 | 0.779 | 0.964 | 1.000 | 1.000 |
| N=200 | 0.772 | 0.959 | 1.000 | 1.000 | 1.000 |
| **Pooled** | | | | | |
| N=20 | 0.230 | 0.334 | 0.562 | 0.873 | 0.994 |
| N=30 | 0.325 | 0.483 | 0.763 | 0.970 | 1.000 |
| N=50 | 0.566 | 0.788 | 0.963 | 1.000 | 1.000 |
| N=100 | 0.778 | 0.950 | 0.998 | 1.000 | 1.000 |
| N=200 | 0.968 | 0.998 | 1.000 | 1.000 | 1.000 |

# CESifo Working Paper Series

(for full list see www.cesifo.de)

---

1269 Thomas Eichner and Rüdiger Pethig, Economic Land Use, Ecosystem Services and Microfounded Species Dynamics, September 2004

1270 Federico Revelli, Performance Rating and Yardstick Competition in Social Service Provision, September 2004

1271 Gerhard O. Orosel and Klaus G. Zauner, Vertical Product Differentiation When Quality is Unobservable to Buyers, September 2004

1272 Christoph Böhringer, Stefan Boeters, and Michael Feil, Taxation and Unemployment: An Applied General Equilibrium Approach, September 2004

1273 Assaf Razin and Efraim Sadka, Welfare Migration: Is the Net Fiscal Burden a Good Measure of its Economics Impact on the Welfare of the Native-Born Population?, September 2004

1274 Tomer Blumkin and Volker Grossmann, Ideological Polarization, Sticky Information, and Policy Reforms, September 2004

1275 Katherine Baicker and Nora Gordon, The Effect of Mandated State Education Spending on Total Local Resources, September 2004

1276 Gabriel J. Felbermayr and Wilhelm Kohler, Exploring the Intensive and Extensive Margins of World Trade, September 2004

1277 John Burbidge, Katherine Cuff and John Leach, Capital Tax Competition with Heterogeneous Firms and Agglomeration Effects, September 2004

1278 Joern-Steffen Pischke, Labor Market Institutions, Wages and Investment, September 2004

1279 Josef Falkinger and Volker Grossmann, Institutions and Development: The Interaction between Trade Regime and Political System, September 2004

1280 Paolo Surico, Inflation Targeting and Nonlinear Policy Rules: The Case of Asymmetric Preferences, September 2004

1281 Ayal Kimhi, Growth, Inequality and Labor Markets in LDCs: A Survey, September 2004

1282 Robert Dur and Amihai Glazer, Optimal Incentive Contracts for a Worker who Envies his Boss, September 2004

1283 Klaus Abberger, Nonparametric Regression and the Detection of Turning Points in the Ifo Business Climate, September 2004

1284 Werner Güth and Rupert Sausgruber, Tax Morale and Optimal Taxation, September 2004

1285 Luis H. R. Alvarez and Erkki Koskela, Does Risk Aversion Accelerate Optimal Forest Rotation under Uncertainty?, September 2004

1286 Giorgio Brunello and Maria De Paola, Market Failures and the Under-Provision of Training, September 2004

1287 Sanjeev Goyal, Marco van der Leij and José Luis Moraga-González, Economics: An Emerging Small World?, September 2004

1288 Sandro Maffei, Nikolai Raabe and Heinrich W. Ursprung, Political Repression and Child Labor: Theory and Empirical Evidence, September 2004

1289 Georg Götz and Klaus Gugler, Market Concentration and Product Variety under Spatial Competition: Evidence from Retail Gasoline, September 2004

1290 Jonathan Temple and Ludger Wößmann, Dualism and Cross-Country Growth Regressions, September 2004

1291 Ravi Kanbur, Jukka Pirttilä and Matti Tuomala, Non-Welfarist Optimal Taxation and Behavioral Public Economics, October 2004

1292 Maarten C. W. Janssen, José Luis Moraga-González and Matthijs R. Wildenbeest, Consumer Search and Oligopolistic Pricing: An Empirical Investigation, October 2004

1293 Kira Börner and Christa Hainz, The Political Economy of Corruption and the Role of Financial Institutions, October 2004

1294 Christoph A. Schaltegger and Lars P. Feld, Do Large Cabinets Favor Large Governments? Evidence from Swiss Sub-Federal Jurisdictions, October 2004

1295 Marc-Andreas Mündler, The Existence of Informationally Efficient Markets When Individuals Are Rational, October 2004

1296 Hendrik Jürges, Wolfram F. Richter and Kerstin Schneider, Teacher Quality and Incentives: Theoretical and Empirical Effects of Standards on Teacher Quality, October 2004

1297 David S. Evans and Michael Salinger, An Empirical Analysis of Bundling and Tying: Over-the-Counter Pain Relief and Cold Medicines, October 2004

1298 Gershon Ben-Shakhar, Gary Bornstein, Astrid Hopfensitz and Frans van Winden, Reciprocity and Emotions: Arousal, Self-Reports, and Expectations, October 2004

1299 B. Zorina Khan and Kenneth L. Sokoloff, Institutions and Technological Innovation During Early Economic Growth: Evidence from the Great Inventors of the United States, 1790 – 1930, October 2004

1300 Piero Gottardi and Roberto Serrano, Market Power and Information Revelation in Dynamic Trading, October 2004

1301 Alan V. Deardorff, Who Makes the Rules of Globalization?, October 2004

1302 Sheilagh Ogilvie, The Use and Abuse of Trust: Social Capital and its Deployment by Early Modern Guilds, October 2004

1303 Mario Jametti and Thomas von Ungern-Sternberg, Disaster Insurance or a Disastrous Insurance – Natural Disaster Insurance in France, October 2004

1304 Pieter A. Gautier and José Luis Moraga-González, Strategic Wage Setting and Coordination Frictions with Multiple Applications, October 2004

1305 Julia Darby, Anton Muscatelli and Graeme Roy, Fiscal Federalism, Fiscal Consolidations and Cuts in Central Government Grants: Evidence from an Event Study, October 2004

1306 Michael Waldman, Antitrust Perspectives for Durable-Goods Markets, October 2004

1307 Josef Honerkamp, Stefan Moog and Bernd Raffelhüschen, Earlier or Later: A General Equilibrium Analysis of Bringing Forward an Already Announced Tax Reform, October 2004

1308 M. Hashem Pesaran, A Pair-Wise Approach to Testing for Output and Growth Convergence, October 2004

1309 John Bishop and Ferran Mane, Educational Reform and Disadvantaged Students: Are They Better Off or Worse Off?, October 2004

1310 Alfredo Schclarek, Consumption and Keynesian Fiscal Policy, October 2004

1311 Wolfram F. Richter, Efficiency Effects of Tax Deductions for Work-Related Expenses, October 2004

1312 Franco Mariuzzo, Patrick Paul Walsh and Ciara Whelan, EU Merger Control in Differentiated Product Industries, October 2004

1313 Kurt Schmidheiny, Income Segregation and Local Progressive Taxation: Empirical Evidence from Switzerland, October 2004

1314 David S. Evans, Andrei Hagiu and Richard Schmalensee, A Survey of the Economic Role of Software Platforms in Computer-Based Industries, October 2004

1315 Frank Riedel and Elmar Wolfstetter, Immediate Demand Reduction in Simultaneous Ascending Bid Auctions, October 2004

1316 Patricia Crifo and Jean-Louis Rullière, Incentives and Anonymity Principle: Crowding Out Toward Users, October 2004

1317 Attila Ambrus and Rossella Argenziano, Network Markets and Consumers Coordination, October 2004

1318 Margarita Katsimi and Thomas Moutos, Monopoly, Inequality and Redistribution Via the Public Provision of Private Goods, October 2004

1319 Jens Josephson and Karl Wärneryd, Long-Run Selection and the Work Ethic, October 2004

1320 Jan K. Brueckner and Oleg Smirnov, Workings of the Melting Pot: Social Networks and the Evolution of Population Attributes, October 2004

1321 Thomas Fuchs and Ludger Wößmann, Computers and Student Learning: Bivariate and Multivariate Evidence on the Availability and Use of Computers at Home and at School, November 2004

1322 Alberto Bisin, Piero Gottardi and Adriano A. Rampini, Managerial Hedging and Portfolio Monitoring, November 2004

1323 Cecilia García-Peñalosa and Jean-François Wen, Redistribution and Occupational Choice in a Schumpeterian Growth Model, November 2004

1324 William Martin and Robert Rowthorn, Will Stability Last?, November 2004

1325 Jianpei Li and Elmar Wolfstetter, Partnership Dissolution, Complementarity, and Investment Incentives, November 2004

1326 Hans Fehr, Sabine Jokisch and Laurence J. Kotlikoff, Fertility, Mortality, and the Developed World's Demographic Transition, November 2004

1327 Adam Elbourne and Jakob de Haan, Asymmetric Monetary Transmission in EMU: The Robustness of VAR Conclusions and Cecchetti's Legal Family Theory, November 2004

1328 Karel-Jan Alsem, Steven Brakman, Lex Hoogduin and Gerard Kuper, The Impact of Newspapers on Consumer Confidence: Does Spin Bias Exist?, November 2004

1329 Chiona Balfoussia and Mike Wickens, Macroeconomic Sources of Risk in the Term Structure, November 2004

1330 Ludger Wößmann, The Effect Heterogeneity of Central Exams: Evidence from TIMSS, TIMSS-Repeat and PISA, November 2004

1331 M. Hashem Pesaran, Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure, November 2004