Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure

Author(s): M. Hashem Pesaran

Source: *Econometrica*, Jul., 2006, Vol. 74, No. 4 (Jul., 2006), pp. 967-1012

Published by: The Econometric Society

Stable URL: https://www.jstor.org/stable/3805914

*Econometrica*, Vol. 74, No. 4 (July, 2006), 967–1012

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

BY M. HASHEM PESARAN[1]

This paper presents a new approach to estimation and inference in panel data models with a general multifactor error structure. The unobserved factors and the individual-specific errors are allowed to follow arbitrary stationary processes, and the number of unobserved factors need not be estimated. The basic idea is to filter the individual-specific regressors by means of cross-section averages 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 least squares applied to auxiliary regressions where the observed regressors are augmented with cross-sectional averages of the dependent variable and the individual-specific regressors. A number of estimators (referred to as common correlated effects (CCE) estimators) are proposed and their asymptotic distributions are derived. The small sample properties of mean group and pooled CCE estimators are investigated by Monte Carlo experiments, showing that the CCE estimators have satisfactory small sample properties even under a substantial degree of heterogeneity and dynamics, and for relatively small values of $N$ and $T$.

KEYWORDS: Cross-section dependence, large panels, common correlated effects, heterogeneity, estimation and inference.

## 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), Pesaran, Schuermann, and Weiner (2004), and Dees, di Mauro, Pesaran, and Smith (2005), as well as the literature on spatial econometrics recently surveyed by Anselin (2001). In the case of panel data models where the cross-section dimension ($N$) is small (typically $N < 10$) and the time series dimension ($T$) is large, the standard approach is to treat the equations from the different cross-section units as a system of seemingly unrelated regression equations (SURE) and then estimate the system by generalized least squares (GLS) techniques. This approach allows for

[1]I am most grateful to a co-editor and three anonymous referees for their helpful suggestions and constructive comments on earlier versions of this paper. I would also like to thank George Kapetanios, Yongcheol Shin, Ron Smith, Til Schuermann, Elisa Tosetti, 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 RES-000-23-0135) is gratefully acknowledged.

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 that depend on whether first- as well as second-order moment restrictions are utilized. In the case where idiosyncratic errors are homoscedastic 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 is 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.[2] 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 point out that, in general, the estimation procedure proposed by Coakley, Fuertes, and Smith will not be consistent if the unobserved factors and the included regressors are correlated. We propose a new

[2]Similar issues are also discussed in the analysis of (dynamic) factor models by Forni and Lippi (1997), Forni and Reichlin (1998), Stock and Watson (2002), Bai and Ng (2002), and Bai (2003), among others.

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 when $(N, T) \to \infty$, 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 among themselves and with the individual-specific regressors. We also allow for individual specific errors to be serially correlated and heteroscedastic, and we 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 to the various approaches adopted in the literature that focus on estimation of factor loadings as an input to 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 focuses 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.

We show that the CCE estimator of the individual-specific coefficients are consistent as $N, T \to \infty$, jointly, as long as a certain rank condition concerning the factor loadings is satisfied. In this case, the asymptotic distribution of the CCE estimator is derived 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 $N \to \infty$ for both $T$ fixed and $T \to \infty$, and we 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 factors, 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).

These results are confirmed by a number of Monte Carlo experiments, some of which are summarized in Section 7. Tests based on CCEMG and CCEP estimators are 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 is shown to perform slightly better than the CCEMG estimator in small samples. 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 result also illustrate the substantial bias and size distortions that result if error cross-section dependence is ignored.[3]

The rest of the paper is organized as follows: Section 2 sets out the multifactor residual model and its assumptions. Section 3 motivates the idea of approximating the 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 4 and their pooled counterpart is presented in Section 5. The mean group estimator based on the individual CCE estimators (i.e., CCEMG) is discussed in Section 5.1, and the pooled version (i.e., CCEP) is discussed in Section 5.2. 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 6. Section 7 reports the results of the Monte Carlo experiments. Section 8 concludes by identifying important areas for extensions and further developments.

NOTATION: The letter $K$ stands for a finite positive constant, $\|\mathbf{A}\| = [\mathrm{Tr}(\mathbf{A}\mathbf{A}')]^{1/2}$ is the Euclidean norm of the $m \times n$ matrix $\mathbf{A}$, and $\mathbf{A}^-$ denotes a generalized inverse of $\mathbf{A}$. The equation $a_n = O(b_n)$ states that the deterministic sequence $\{a_n\}$ is at most of order $b_n$, $\mathbf{x}_n = O_p(\mathbf{y}_n)$ states that 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 $\mathbf{y}_n$. The operator $\xrightarrow{\text{q.m.}}$ denotes convergence in quadratic mean (or mean squared error), $\xrightarrow{p}$ denotes convergence in probability, $\xrightarrow{d}$ denotes convergence in distribution, and $\overset{d}{\sim}$ denotes asymptotic equivalence of probability distributions. All asymptotics are carried out under $N \to \infty$, with either 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.

[3]The simulation results also highlight the importance of testing for error cross-section dependence in panel data models. General tests of error cross-section dependence are discussed in Pesaran (2004a).

## 2. A MULTIFACTOR RESIDUAL MODEL

Let $y_{it}$ be the observation on the $i$th cross-section unit at time $t$ for $i = 1, 2, \ldots, N$, $t = 1, 2, \ldots, T$, and suppose that it is generated according to the linear heterogeneous panel data model

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

where $\mathbf{d}_t$ is a $n \times 1$ vector of observed common effects (including deterministics such as intercepts or seasonal 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

$$e_{it} = \boldsymbol{\gamma}_i'\mathbf{f}_t + \varepsilon_{it}, \tag{2}$$

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 fairly general model for the individual specific regressors,

$$\mathbf{x}_{it} = \mathbf{A}_i'\mathbf{d}_t + \boldsymbol{\Gamma}_i'\mathbf{f}_t + \mathbf{v}_{it}, \tag{3}$$

where $\mathbf{A}_i$ and $\boldsymbol{\Gamma}_i$ are $n \times k$ and $m \times k$, factor loading matrices with fixed components, and $\mathbf{v}_{it}$ 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 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 (1)–(3), we now have the system of equations

$$\underset{(k+1)\times 1}{\mathbf{z}_{it}} = \begin{pmatrix} y_{it} \\ \mathbf{x}_{it} \end{pmatrix} = \underset{(k+1)\times n}{\mathbf{B}_i'} \underset{n\times 1}{\mathbf{d}_t} + \underset{(k+1)\times m}{\mathbf{C}_i'} \underset{m\times 1}{\mathbf{f}_t} + \underset{(k+1)\times 1}{\mathbf{u}_{it}}, \tag{4}$$

where

$$\mathbf{u}_{it} = \begin{pmatrix} \varepsilon_{it} + \boldsymbol{\beta}_i'\mathbf{v}_{it} \\ \mathbf{v}_{it} \end{pmatrix}, \tag{5}$$

$$\mathbf{B}_i = (\boldsymbol{\alpha}_i \quad \mathbf{A}_i) \begin{pmatrix} 1 & \mathbf{0} \\ \boldsymbol{\beta}_i & \mathbf{I}_k \end{pmatrix}, \quad \mathbf{C}_i = (\boldsymbol{\gamma}_i \quad \boldsymbol{\Gamma}_i) \begin{pmatrix} 1 & \mathbf{0} \\ \boldsymbol{\beta}_i & \mathbf{I}_k \end{pmatrix}, \tag{6}$$

$\mathbf{I}_k$ 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

$$\tilde{\boldsymbol{\Gamma}}_i = (\boldsymbol{\gamma}_i \quad \boldsymbol{\Gamma}_i). \tag{7}$$

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

The above setup 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 $\boldsymbol{\Gamma}_i = \mathbf{0}$. In most applications of interest, however, the individual-specific regressors are likely to be cross-sectionally dependent and a formulation such as (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 (1) and (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 demeaned observations. (v) The large $N$ and $T$ factor models recently analyzed by Stock and Watson (2002) 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 $\boldsymbol{\beta}_i$.[4] In our analysis, $\mathbf{f}_t$ are treated as nuisance "parameters" and no assumption is made about $m$, except that it is fixed as $N$ and $T$ tend to infinity.

In the panel literature with $T$ small and $N$ large, the primary parameters of interest are the means of the individual-specific slope coefficients $\boldsymbol{\beta}_i$, $i = 1, 2, \dots, N$. 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 relative to $E(\boldsymbol{\beta}_i) = \boldsymbol{\beta}$ and we 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 $i$, $j$, $t$, and $t'$.

[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 $\varepsilon_{it}$ in (5) are uncorrelated provides the $k$ restrictions needed for the exact identification of $\boldsymbol{\beta}_i$.

(a) For each $i$, $\varepsilon_{it}$ and $\mathbf{v}_{it}$ follow linear stationary processes with absolute summable autocovariances,

$$\varepsilon_{it} = \sum_{\ell=0}^{\infty} a_{i\ell}\zeta_{i,t-\ell} \tag{8}$$

and

$$\mathbf{v}_{it} = \sum_{\ell=0}^{\infty} \mathbf{S}_{i\ell}\boldsymbol{\nu}_{i,t-\ell}, \tag{9}$$

where $(\zeta_{it}, \boldsymbol{\nu}'_{it})'$ are $(k+1)\times 1$ vectors of identically, independently distributed (IID) random variables with mean zero, variance matrix $\mathbf{I}_{k+1}$, and finite fourth-order cumulants. In particular,

$$\mathrm{Var}(\varepsilon_{it}) = \sum_{\ell=0}^{\infty} a_{i\ell}^2 = \sigma_i^2 \le \bar{\sigma}^2 < \infty, \tag{10}$$

$$\mathrm{Var}(\mathbf{v}_{it}) = \sum_{\ell=0}^{\infty} \mathbf{S}_{i\ell}\mathbf{S}'_{i\ell} = \boldsymbol{\Sigma}_i \le \bar{\boldsymbol{\Sigma}} < \infty \tag{11}$$

for all $i$ and some constants $\bar{\sigma}^2$ and $\bar{\boldsymbol{\Sigma}}$, where $\sigma_i^2 > 0$ and $\boldsymbol{\Sigma}_i$ is a positive definite matrix.

ASSUMPTION 3—Factor Loadings: The unobserved factor loadings $\boldsymbol{\gamma}_i$ and $\boldsymbol{\Gamma}_i$ 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 $\boldsymbol{\Gamma}$, respectively, and finite variances. In particular,

$$\boldsymbol{\gamma}_i = \boldsymbol{\gamma} + \boldsymbol{\eta}_i, \quad \boldsymbol{\eta}_i \backsim \mathrm{IID}(\mathbf{0}, \boldsymbol{\Omega}_{\boldsymbol{\eta}}) \quad \text{for} \quad i = 1, 2, \dots, N, \tag{12}$$

where $\boldsymbol{\Omega}_\eta$ is a $m \times m$ symmetric nonnegative definite matrix, and $\|\boldsymbol{\gamma}\| < K$, $\|\boldsymbol{\Gamma}\| < K$, and $\|\boldsymbol{\Omega}_\eta\| < K$ for some positive constant $K < \infty$.

ASSUMPTION 4—Random Slope Coefficients: The slope coefficients $\boldsymbol{\beta}_i$ follow the random coefficient model

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{v}_i, \quad \boldsymbol{v}_i \backsim \mathrm{IID}(\mathbf{0}, \boldsymbol{\Omega}_{\boldsymbol{v}}), \quad \text{for} \quad i = 1, 2, \dots, N, \tag{13}$$

where $\|\boldsymbol{\beta}\| < K$, $\|\boldsymbol{\Omega}_{\boldsymbol{v}}\| < K$, $\boldsymbol{\Omega}_{\boldsymbol{v}}$ is a $k \times k$ symmetric nonnegative definite matrix, and the random deviations $\boldsymbol{v}_i$ are distributed independently of $\boldsymbol{\gamma}_j$, $\boldsymbol{\Gamma}_j$, $\varepsilon_{jt}$, $\mathbf{v}_{jt}$, 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 $\{w_j\}$ that satisfy the conditions[5]

$$(14) \qquad \text{(i)} \quad w_i = O\left(\frac{1}{N}\right), \quad \text{(ii)} \quad \sum_{i=1}^{N} w_i = 1, \quad \text{(iii)} \quad \sum_{i=1}^{N} |w_i| < K,$$

and let

$$(15) \qquad \bar{\mathbf{M}}_w = \mathbf{I}_T - \bar{\mathbf{H}}_w (\bar{\mathbf{H}}_w' \bar{\mathbf{H}}_w)^{-} \bar{\mathbf{H}}_w'$$

and

$$(16) \qquad \mathbf{M}_g = \mathbf{I}_T - \mathbf{G}(\mathbf{G}'\mathbf{G})^{-}\mathbf{G}',$$

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

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

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

$$(17) \qquad \hat{\boldsymbol{\Psi}}_{NT} = \sum_{i=1}^{N} \theta_i \left( \frac{\mathbf{X}_i' \bar{\mathbf{M}}_w \mathbf{X}_i}{T} \right)$$

is nonsingular for the scalar weights $\theta_i$ that satisfy the conditions

$$(18) \qquad \text{(i)} \quad \theta_i = O\left(\frac{1}{N}\right), \quad \text{(ii)} \quad \sum_{i=1}^{N} \theta_i = 1, \quad \text{(iii)} \quad \sum_{i=1}^{N} |\theta_i| < K.$$

REMARK 1: The residual factor model specified by (1)–(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.

[5]Note that the conditions in (14) also imply that $\sum_{i=1}^{N} w_i^2 = O(N^{-1})$.

REMARK 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 adjustments to the rate of convergence of the CCE estimator of the associated trend coefficient. The main results of this paper also hold if there are unit root processes among 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 (2006).

REMARK 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 that 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 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 5: It is worth noting that the common feature dynamics across $i$ are captured through the serial correlation structure of the common effects, and individual-specific dynamics are allowed through serial correlation in $\varepsilon_{it}$.

## 3. A GENERAL APPROACH TO ESTIMATION OF PANELS WITH COMMON EFFECTS

To deal with the residual cross section dependence, Coakley, Fuertes, and Smith (2002; CFS) propose a principal components estimator by augmenting the regression of $y_{it}$ on $\mathbf{x}_{it}$ with one or more principal component of the estimated OLS residuals $\hat{e}_{it}$, $i = 1, 2, \dots, N$, $t = 1, 2, \dots, T$, obtained from the first stage regression of $y_{it}$ on $\mathbf{x}_{it}$ for each $i$. However, it is easily established that, in general, the CFS estimator will not be consistent, because 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.[6] One way to overcome 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

[6]For a proof and further discussions, see Section 3 in Pesaran (2004b).

can be done consider the cross-section averages of the equations in (4), using the weights $w_j$,[7]

$$\bar{\mathbf{z}}_{wt} = \bar{\mathbf{B}}_w'\mathbf{d}_t + \bar{\mathbf{C}}_w'\mathbf{f}_t + \bar{\mathbf{u}}_{wt}, \tag{19}$$

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

$$\bar{\mathbf{B}}_w = \sum_{i=1}^N w_i \mathbf{B}_i, \quad \bar{\mathbf{C}}_w = \sum_{i=1}^N w_i \mathbf{C}_i, \quad \bar{\mathbf{u}}_{wt} = \sum_{i=1}^N w_i \mathbf{u}_{it}, \tag{20}$$

and suppose that

$$\operatorname{Rank}(\bar{\mathbf{C}}_w) = m \leq k+1 \quad \text{for all } N. \tag{21}$$

Then we have

$$\mathbf{f}_t = (\bar{\mathbf{C}}_w \bar{\mathbf{C}}_w')^{-1}\bar{\mathbf{C}}_w(\bar{\mathbf{z}}_{wt} - \bar{\mathbf{B}}_w'\mathbf{d}_t - \bar{\mathbf{u}}_{wt}). \tag{22}$$

However, using Lemma 1 in Appendix A, we have

$$\bar{\mathbf{u}}_{wt} \stackrel{\text{q.m.}}{\to} \mathbf{0}, \quad \text{as} \quad N \to \infty, \quad \text{for each } t, \tag{23}$$

and

$$\bar{\mathbf{C}}_w \stackrel{p}{\to} \mathbf{C} = \tilde{\boldsymbol{\Gamma}} \begin{pmatrix} 1 & \mathbf{0} \\ \boldsymbol{\beta} & \mathbf{I}_k \end{pmatrix}, \quad \text{as} \quad N \to \infty, \tag{24}$$

where

$$\tilde{\boldsymbol{\Gamma}} = (E(\boldsymbol{\gamma}_i), E(\boldsymbol{\Gamma}_i)) = (\boldsymbol{\gamma}, \boldsymbol{\Gamma}). \tag{25}$$

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

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

This suggests using $\bar{\mathbf{h}}_{wt} = (\mathbf{d}_t', \bar{\mathbf{z}}_{wt}')'$ as observable proxies for $\mathbf{f}_t$. Although 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

[7] 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 $\sigma_j^2$, $j = 1, 2, \ldots, N$. But for consistent estimation it is only required that the chosen weights satisfy the conditions in (14), in particular that for each $i$, $\sum_{j=1}^N w_{ij}^2 \to 0$ as $N \to \infty$.

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 (21) is not satisfied. Rank deficiency in $\mathbf{C}$ induces exact linear dependencies among the elements of $\bar{\mathbf{h}}_{wt}$ as $N \to \infty$. For example, in the extreme case where $\mathbf{C} = \mathbf{0}$, using (19), we have

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

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. However, 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 $N \to \infty$. As we shall show, the CCE estimators of $\boldsymbol{\beta}$ are not affected by the rank deficiency problem and continue to be asymptotically invariant to the factor loadings $\boldsymbol{\gamma}_i$, for any fixed $m$.

## 4. COMMON CORRELATED EFFECTS ESTIMATORS: INDIVIDUAL SPECIFIC COEFFICIENTS

For the individual slope coefficients the CCE is given by

$$\hat{\mathbf{b}}_i = (\mathbf{X}_i' \bar{\mathbf{M}}_w \mathbf{X}_i)^{-1} \mathbf{X}_i' \bar{\mathbf{M}}_w \mathbf{y}_i, \tag{26}$$

where $\mathbf{X}_i$ is defined above, $\mathbf{y}_i = (y_{i1}, y_{i2}, \ldots, y_{iT})'$, $\bar{\mathbf{M}}_w$ is given by

$$\bar{\mathbf{M}}_w = \mathbf{I}_T - \bar{\mathbf{H}}_w (\bar{\mathbf{H}}_w' \bar{\mathbf{H}}_w)^{-1} \bar{\mathbf{H}}_w', \tag{27}$$

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 $\text{Rank}(\tilde{\boldsymbol{\Gamma}}) = m$ ensures that under Assumptions 1–4, $T^{-1}(\bar{\mathbf{H}}_w' \bar{\mathbf{H}}_w)$ converges to a positive definite matrix for a fixed $T$ as $N \to \infty$, as well as when $(N, T) \xrightarrow{j} \infty$. However, $T^{-1}(\mathbf{X}_i' \bar{\mathbf{M}}_w \mathbf{X}_i)$ and its limit as $(N, T) \xrightarrow{j} \infty$ exist 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 as long as $\boldsymbol{\Sigma}_i$ is positive definite.

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

$$\mathbf{y}_i = \mathbf{D}\boldsymbol{\alpha}_i + \mathbf{X}_i \boldsymbol{\beta}_i + \mathbf{F}\boldsymbol{\gamma}_i + \boldsymbol{\varepsilon}_i, \tag{28}$$

where $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{iT})'$ and, as set out in Assumption 5, $\mathbf{D} = (\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_T)'$ and $\mathbf{F} = (\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_T)'$. Using (26) and (28) we have

$$\hat{\mathbf{b}}_i - \boldsymbol{\beta}_i = \left(\frac{\mathbf{X}_i' \bar{\mathbf{M}}_w \mathbf{X}_i}{T}\right)^{-1} \left(\frac{\mathbf{X}_i' \bar{\mathbf{M}}_w \mathbf{F}}{T}\right) \boldsymbol{\gamma}_i + \left(\frac{\mathbf{X}_i' \bar{\mathbf{M}}_w \mathbf{X}_i}{T}\right)^{-1} \left(\frac{\mathbf{X}_i' \bar{\mathbf{M}}_w \boldsymbol{\varepsilon}_i}{T}\right), \tag{29}$$

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 (3) and (19) in matrix notation, we first note that

$$\mathbf{X}_i = \mathbf{G}\mathbf{\Pi}_i + \mathbf{V}_i \tag{30}$$

and

$$\bar{\mathbf{H}}_w = \mathbf{G}\bar{\mathbf{P}}_w + \bar{\mathbf{U}}_w^*, \tag{31}$$

where $\mathbf{G} = (\mathbf{D}, \mathbf{F})$, $\mathbf{\Pi}_i = (\mathbf{A}_i', \mathbf{\Gamma}_i')'$, $\mathbf{V}_i = (\mathbf{v}_{i1}, \mathbf{v}_{i2}, \dots, \mathbf{v}_{iT})'$,

$$\underset{(n+m)\times(n+k+1)}{\bar{\mathbf{P}}_w} = \begin{pmatrix} \mathbf{I}_n & \bar{\mathbf{B}}_w \\ \mathbf{0} & \bar{\mathbf{C}}_w \end{pmatrix}, \quad \bar{\mathbf{U}}_w^* = (\mathbf{0}, \bar{\mathbf{U}}_w), \tag{32}$$

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

$$\|\bar{\mathbf{B}}_w\| = \sum_{i=1}^{N} |w_i| \|\mathbf{B}_i\| < K \quad \text{and} \quad \|\bar{\mathbf{C}}_w\| = \sum_{i=1}^{N} |w_i| \|\mathbf{C}_i\| < K \tag{33}$$

under (14) 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

$$\begin{aligned} &\frac{\mathbf{X}_i'\mathbf{G}}{T} = \mathbf{\Pi}_i'\left(\frac{\mathbf{G}'\mathbf{G}}{T}\right) + \frac{\mathbf{V}_i'\mathbf{G}}{T} = O_p(1), \\ &\frac{\mathbf{G}'\mathbf{G}}{T} = O_p(1), \quad \frac{\mathbf{G}'\mathbf{F}}{T} = O_p(1). \end{aligned} \tag{34}$$

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

$$\frac{\mathbf{X}_i'\bar{\mathbf{H}}_w}{T} = \left(\frac{\mathbf{X}_i'\mathbf{G}}{T}\right)\bar{\mathbf{P}}_w + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right), \tag{35}$$

$$\frac{\bar{\mathbf{H}}_w'\bar{\mathbf{H}}_w}{T} = \bar{\mathbf{P}}_w'\left(\frac{\mathbf{G}'\mathbf{G}}{T}\right)\bar{\mathbf{P}}_w + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right), \tag{36}$$

$$\frac{\bar{\mathbf{H}}_w'\mathbf{F}}{T} = \bar{\mathbf{P}}_w'\left(\frac{\mathbf{G}'\mathbf{F}}{T}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right). \tag{37}$$

Hence, we obtain the following result, which is critical to many of the derivations in this paper and does not require the rank condition (21):

$$\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{F}}{T} = \frac{\mathbf{X}_i'\bar{\mathbf{M}}_q\mathbf{F}}{T} + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right), \tag{38}$$

where

$$\bar{\mathbf{M}}_q = \mathbf{I}_T - \bar{\mathbf{Q}}_w(\bar{\mathbf{Q}}_w'\bar{\mathbf{Q}}_w)^{-}\bar{\mathbf{Q}}_w' \quad \text{with} \quad \bar{\mathbf{Q}}_w = \mathbf{G}\bar{\mathbf{P}}_w. \tag{39}$$

When the rank condition (21) is satisfied, using familiar results on generalized inverse, we have

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

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

$$\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{F}}{T} = O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right). \tag{40}$$

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

$$\left(\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{F}}{T}\right)\bar{\mathbf{C}}_w = O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right). \tag{41}$$

Also, using (6) and (13) we have

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

where $\bar{\boldsymbol{\Gamma}}_w = \sum_{i=1}^{N} w_i\boldsymbol{\Gamma}_i$. Substituting this result in (41) now yields

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

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

which in turn leads to

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

However, under Assumption 4 and (14), $\sum_{i=1}^{N} w_i\boldsymbol{\Gamma}_i\boldsymbol{v}_i = O_p(N^{-1/2})$ and, therefore,

$$\frac{\sqrt{N}(\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{F})\bar{\boldsymbol{\gamma}}_w}{T} = O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right). \tag{42}$$

This result is clearly implied by (40), irrespective of whether the factor loadings are random or simply bounded. However, the reverse is not true; (42) does not imply (40) if the rank condition is not satisfied.

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

$$\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{X}_i}{T} = \frac{\mathbf{X}_i'\bar{\mathbf{M}}_q\mathbf{X}_i}{T} + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right) \tag{43}$$

and

$$\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\boldsymbol{\varepsilon}_i}{T} = \frac{\mathbf{X}_i'\bar{\mathbf{M}}_q\boldsymbol{\varepsilon}_i}{T} + O_p\left(\frac{1}{N}\right). \tag{44}$$

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 (29), noting that $T^{-1}\mathbf{X}_i'\bar{\mathbf{M}}_q\mathbf{X}_i = O_p(1)$, and assuming that the rank condition (21) is satisfied, we have[8]

$$\hat{\mathbf{b}}_i - \boldsymbol{\beta}_i = \left(\frac{\mathbf{X}_i'\mathbf{M}_g\mathbf{X}_i}{T}\right)^{-1}\left(\frac{\mathbf{X}_i'\mathbf{M}_g\boldsymbol{\varepsilon}_i}{T}\right) + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right). \tag{45}$$

Whereas $\boldsymbol{\varepsilon}_i$ is independently distributed of $\mathbf{X}_i$ and $\mathbf{G} = (\mathbf{D}, \mathbf{F})$, then for a fixed $T$ and as $N \to \infty$, $E(\hat{\mathbf{b}}_i - \boldsymbol{\beta}_i) = \mathbf{0}$. The finite-$T$ distribution of $\hat{\mathbf{b}}_i - \boldsymbol{\beta}_i$ will not depend on the factor loadings as $N \to \infty$, but will depend on the probability density of $\boldsymbol{\varepsilon}_i$. For $N$ and $T$ sufficiently large, the distribution of $\sqrt{T}(\hat{\mathbf{b}}_i - \boldsymbol{\beta}_i)$ will be asymptotically normal if the rank condition (21) is satisfied, and if $N$ and $T$ are of the same order of magnitude, namely, if $T/N \to \kappa$ as $N$ and $T \to \infty$, where $\kappa$ is a positive finite constant.

The following theorem provides a formal statement of these results and the associated asymptotic distribution in the case where the rank condition is satisfied and $(N, T) \xrightarrow{j} \infty$.

THEOREM 1: *Consider the panel data model* (1) *and* (2), *and suppose that* $\|\boldsymbol{\beta}_i\| < K$, $\|\boldsymbol{\Pi}_i\| < K$, *Assumptions* 1, 2, *and* 5(a) *hold*, $(N, T) \xrightarrow{j} \infty$ (*in no particular order*), *and the rank condition* (21) *is satisfied. Then* $\hat{\mathbf{b}}_i$ *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*

$$\sqrt{T}(\hat{\mathbf{b}}_i - \boldsymbol{\beta}_i) \xrightarrow{d} \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_{b_i}), \tag{46}$$

*where*

$$\boldsymbol{\Sigma}_{b_i} = \boldsymbol{\Sigma}_i^{-1}\mathbf{S}_{i\varepsilon}\boldsymbol{\Sigma}_i^{-1}, \tag{47}$$

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

$\boldsymbol{\Sigma}_i$ *is defined by* (11),

$$\mathbf{S}_{i\varepsilon} = p \lim_{T\to\infty} [T^{-1}(\mathbf{X}_i'\mathbf{M}_g\boldsymbol{\Omega}_{\boldsymbol{\varepsilon}_i}\mathbf{M}_g\mathbf{X}_i)], \tag{48}$$

$\mathbf{M}_g$ *is defined by* (16), *and* $\boldsymbol{\Omega}_{\boldsymbol{\varepsilon}_i} = E(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i')$.

Using (45), consistency of $\hat{\mathbf{b}}_i$ follows almost immediately by noting that under our assumptions $T^{-1}(\mathbf{X}_i'\mathbf{M}_g\mathbf{X}_i) \xrightarrow{p} \boldsymbol{\Sigma}_i$ (a fixed positive definite matrix) and $T^{-1}(\mathbf{X}_i'\mathbf{M}_g\boldsymbol{\varepsilon}_i) \xrightarrow{p} \mathbf{0}$. To derive the asymptotic distribution of $\hat{\mathbf{b}}_i$ we first note that

$$\sqrt{T}(\hat{\mathbf{b}}_i - \boldsymbol{\beta}_i) = \left(\frac{\mathbf{X}_i'\mathbf{M}_g\mathbf{X}_i}{T}\right)^{-1}\frac{\mathbf{X}_i'\mathbf{M}_g\boldsymbol{\varepsilon}_i}{\sqrt{T}} + O_p\left(\frac{\sqrt{T}}{N}\right) + O_p\left(\frac{1}{\sqrt{N}}\right), \tag{49}$$

and because by assumption $\sqrt{T}/N \to 0$ as $(N, T) \xrightarrow{j} \infty$, we have

$$\sqrt{T}(\hat{\mathbf{b}}_i - \boldsymbol{\beta}_i) \overset{d}{\sim} \boldsymbol{\Sigma}_i^{-1}\left(\frac{\mathbf{X}_i'\mathbf{M}_g\boldsymbol{\varepsilon}_i}{\sqrt{T}}\right),$$

which establishes (46), recalling that $\boldsymbol{\varepsilon}_i$ is a stationary process assumed to be distributed independently of the stationary processes $\mathbf{X}_i$ and $\mathbf{G}$.

A consistent estimator of $\boldsymbol{\Sigma}_{b_i}$ can be obtained, for example, using the Newey and West (1987) type procedure and is given by

$$\hat{\boldsymbol{\Sigma}}_{T,b_i} = \left(\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{X}_i}{T}\right)^{-1}\hat{\mathbf{S}}_{i\varepsilon}\left(\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{X}_i}{T}\right)^{-1}, \tag{50}$$

where

$$\hat{\mathbf{S}}_{i\varepsilon} = \hat{\boldsymbol{\Lambda}}_{i0} + \sum_{j=1}^{p}\left(1 - \frac{j}{p+1}\right)(\hat{\boldsymbol{\Lambda}}_{ij} + \hat{\boldsymbol{\Lambda}}_{ij}'), \tag{51}$$

$$\hat{\boldsymbol{\Lambda}}_{ij} = T^{-1}\sum_{t=j+1}^{p} \hat{e}_{it}\hat{e}_{i,t-j}\hat{\mathbf{x}}_{it}\hat{\mathbf{x}}_{i,t-j}', \tag{52}$$

$p$ is the window size, $\hat{e}_{it}$ is the $t$th element of $\hat{\mathbf{e}}_i = \bar{\mathbf{M}}_w(\mathbf{y}_i - \mathbf{X}_i\hat{\mathbf{b}}_i)$, and $\hat{\mathbf{x}}_{it}'$ is the $t$th row of $\hat{\mathbf{X}}_i = \bar{\mathbf{M}}_w\mathbf{X}_i$.

When the rank condition (21), 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.

## 5. 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 (13), 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.

### 5.1. *Common Correlated Effects Mean Group Estimator*

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

$$\hat{\mathbf{b}}_{\mathrm{MG}} = N^{-1}\sum_{i=1}^{N}\hat{\mathbf{b}}_i. \tag{53}$$

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)):

$$\hat{\mathbf{b}}_{\mathrm{RC}} = \sum_{i=1}^{N}\hat{\boldsymbol{\Theta}}_i\hat{\mathbf{b}}_i, \tag{54}$$

where

$$\hat{\boldsymbol{\Theta}}_i = \left\{\sum_{j=1}^{N}[\hat{\boldsymbol{\Sigma}}_{T,b_j} + \hat{\boldsymbol{\Omega}}_{\boldsymbol{v}}]^{-1}\right\}^{-1}[\hat{\boldsymbol{\Sigma}}_{T,b_i} + \hat{\boldsymbol{\Omega}}_{\boldsymbol{v}}]^{-1}, \tag{55}$$

$\hat{\boldsymbol{\Sigma}}_{T,b_j}$ is given by (50), and $\hat{\boldsymbol{\Omega}}_{\boldsymbol{v}}$ is a consistent estimator of $\boldsymbol{\Omega}_{\boldsymbol{v}}$, the variance of $\boldsymbol{v}_i$ defined by (13). 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 (29) we have

$$\begin{aligned}\sqrt{N}(\hat{\mathbf{b}}_{\mathrm{MG}} - \boldsymbol{\beta}) &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\boldsymbol{v}_i + \frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{\Psi}}_{iT}^{-1}\left(\frac{\sqrt{N}\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{F}}{T}\right)\boldsymbol{\gamma}_i \\ &\quad + \frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{\Psi}}_{iT}^{-1}\left(\frac{\sqrt{N}\mathbf{X}_i'\bar{\mathbf{M}}_w\boldsymbol{\varepsilon}_i}{T}\right),\end{aligned} \tag{56}$$

where, by assumption, $\hat{\boldsymbol{\Psi}}_{iT}^{-1} = (T^{-1}\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{X}_i)^{-1}$ has second-order moments. In the case where the rank condition (21) is satisfied, using (40) we have

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

and it is easily seen that for all bounded values of the factor loadings,

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

Similarly, using (43) and (44),

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

where

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

However, whereas $\boldsymbol{\varepsilon}_i$ and $\mathbf{V}_i$ are distributed independently of each other as well as across $i$, it is easily seen that $E(\Delta_{NT}) = \mathbf{0}$, and

$$\begin{aligned}&\operatorname{Var}(\Delta_{NT}) \\ &= \frac{1}{NT}\sum_{i=1}^{N}E\left\{\left(\frac{\mathbf{V}_i'\mathbf{M}_g\mathbf{V}_i}{T}\right)^{-1}\left(\frac{\mathbf{V}_i'\mathbf{M}_g\boldsymbol{\Omega}_{\varepsilon_i}\mathbf{M}_g\mathbf{V}_i}{T}\right)\left(\frac{\mathbf{V}_i'\mathbf{M}_g\mathbf{V}_i}{T}\right)^{-1}\right\}\end{aligned}$$

and

$$\operatorname{Var}(\Delta_{NT}) = \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{\Sigma}_i^{-1}\mathbf{S}_{i\varepsilon}\boldsymbol{\Sigma}_i^{-1} + O\left(\frac{1}{T\sqrt{T}}\right) = O\left(\frac{1}{T}\right),$$

where $\mathbf{S}_{i\varepsilon}$ is given by (48). Hence

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

and

$$\sqrt{N}(\hat{\mathbf{b}}_{\mathrm{MG}} - \boldsymbol{\beta}) \xrightarrow{d} \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\mathrm{MG}}), \quad \text{as} \quad (N, T) \xrightarrow{j} \infty. \tag{57}$$

In the present case, $\boldsymbol{\Sigma}_{\mathrm{MG}} = \boldsymbol{\Omega}_{\boldsymbol{v}}$ and can be consistently estimated nonparametrically by

$$\hat{\boldsymbol{\Sigma}}_{\mathrm{MG}} = \frac{1}{N-1}\sum_{i=1}^{N}(\hat{\mathbf{b}}_i - \hat{\mathbf{b}}_{\mathrm{MG}})(\hat{\mathbf{b}}_i - \hat{\mathbf{b}}_{\mathrm{MG}})'. \tag{58}$$

Consider now the case where the rank condition is not satisfied, but the factor loadings satisfy the random coefficient model (12). In this case, using (12) we note that the second term in (56) can be written as

$$\boldsymbol{\chi}_{NT} = \frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{\Psi}}_{iT}^{-1}\left(\frac{\sqrt{N}\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{F}}{T}\right)(\bar{\boldsymbol{\gamma}}_w + \boldsymbol{\eta}_i - \bar{\boldsymbol{\eta}}_w), \tag{59}$$

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 (38), (42), and (43) we have

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

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

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

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.[9] In this case, the asymptotic variance of $\sqrt{N}(\hat{\mathbf{b}}_{\mathrm{MG}} - \boldsymbol{\beta})$ is given by

$$\boldsymbol{\Sigma}_{\mathrm{MG}} = \boldsymbol{\Omega}_{\boldsymbol{v}} + \lim_{N\to\infty}\left[\frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{\Sigma}_{iq}^{-1}\mathbf{Q}_{if}\boldsymbol{\Omega}_{\boldsymbol{\eta}}\mathbf{Q}_{if}'\boldsymbol{\Sigma}_{iq}^{-1})\right], \tag{60}$$

[9]The latter result follows using Lemma 4 and noting that as $T \to \infty$, $T^{-1}\mathbf{X}_i'\bar{\mathbf{M}}_q\mathbf{X}_i \xrightarrow{p} \boldsymbol{\Sigma}_{iq}$, which is a positive definite matrix.

where

$$\boldsymbol{\Sigma}_{iq} = p \lim_{T\to\infty} (T^{-1}\mathbf{X}_i'\bar{\mathbf{M}}_q\mathbf{X}_i) \quad \text{and} \quad \mathbf{Q}_{if} = p \lim_{T\to\infty} (T^{-1}\mathbf{X}_i'\bar{\mathbf{M}}_q\mathbf{F}), \tag{61}$$

and depends on the unobserved factors. Nevertheless, it can be consistently estimated nonparametrically using (58). To see this, first note that

$$\hat{\mathbf{b}}_i - \boldsymbol{\beta} = \boldsymbol{v}_i + \mathbf{h}_{iT} + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right), \tag{62}$$

where

$$\mathbf{h}_{iT} = \left(\frac{\mathbf{X}_i'\bar{\mathbf{M}}_q\mathbf{X}_i}{T}\right)^{-1} \frac{\mathbf{X}_i'\bar{\mathbf{M}}_q[\mathbf{F}(\boldsymbol{\eta}_i - \bar{\boldsymbol{\eta}}_w) + \boldsymbol{\varepsilon}_i]}{T} \tag{63}$$

and

$$\hat{\mathbf{b}}_i - \hat{\mathbf{b}}_{\mathrm{MG}} = (\boldsymbol{v}_i - \bar{\boldsymbol{v}}) + (\mathbf{h}_{iT} - \bar{\mathbf{h}}_T) + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right). \tag{64}$$

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

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

Hence, interestingly enough, the nonparametric variance estimator given by (58) is valid irrespective of whether the rank condition is satisfied.

The above results are summarized in the following general theorem:

THEOREM 2: *Consider the panel data model* (1) *and* (2), *and suppose that Assumptions* 1–4 *and* 5(a) *hold. Then the common correlated effects mean group estimator* $\mathbf{b}_{\mathrm{MG}}$ *defined by* (53), *is asymptotically* (*for a fixed* $T$ *and as* $N \to \infty$) *unbiased for* $\boldsymbol{\beta}$ *and as* $(N, T) \xrightarrow{j} \infty$,

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

*where* $\boldsymbol{\Sigma}_{\mathrm{MG}}$ *is given by* (60), *which is consistently estimated by* (58).

This theorem does not require that the rank condition (21) holds for any number $m$ of unobserved factors as long as $m$ is fixed and does not impose any restrictions on the relative rates of expansion of $N$ and $T$. However, 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.

### 5.2. *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

$$\hat{\mathbf{b}}_P = \left( \sum_{i=1}^{N} \theta_i \mathbf{X}_i' \bar{\mathbf{M}}_w \mathbf{X}_i \right)^{-1} \sum_{i=1}^{N} \theta_i \mathbf{X}_i' \bar{\mathbf{M}}_w \mathbf{y}_i. \tag{65}$$

Typically, the (pooling) weights $\theta_i$ are set equal to $1/N$, although in the general case where $\sigma_i^2$ differs across $i$, 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 (21) 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 $\hat{\mathbf{b}}_P$ are summarized in the following theorem; the proofs are provided in Appendix B.

THEOREM 3: *Consider the panel data model* (1) *and* (2), *suppose that Assumptions* 1–4 *and* 5(b) *hold, and* $\theta_i = w_i$. *Then the common correlated effects pooled estimator* $\hat{\mathbf{b}}_P$, *defined by* (65), *is asymptotically unbiased for* $\boldsymbol{\beta}$ *and as* $(N, T) \xrightarrow{j} \infty$ *we have*

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

*where*

$$\begin{aligned} \boldsymbol{\Sigma}_P^* &= \boldsymbol{\Psi}^{*-1} \mathbf{R}^* \boldsymbol{\Psi}^{*-1}, \\ \boldsymbol{\Psi}^* &= \lim_{N \to \infty} \left( \sum_{i=1}^{N} w_i \boldsymbol{\Sigma}_{iq} \right), \\ \mathbf{R}^* &= \lim_{N \to \infty} \left[ N^{-1} \sum_{i=1}^{N} \tilde{w}_i^2 (\boldsymbol{\Sigma}_{iq} \boldsymbol{\Omega}_{\boldsymbol{v}} \boldsymbol{\Sigma}_{iq} + \mathbf{Q}_{if} \boldsymbol{\Omega}_{\boldsymbol{\eta}} \mathbf{Q}_{if}') \right], \\ \tilde{w}_i &= \frac{w_i}{\sqrt{N^{-1} \sum_{i=1}^{N} w_i^2}}, \end{aligned} \tag{66}$$

*and* $\boldsymbol{\Sigma}_{iq}$ *and* $\mathbf{Q}_{if}$ *are defined by* (61).

Although the asymptotic variance matrix of $\hat{\mathbf{b}}_P$ depends on the unobserved factors and their loadings, it is nevertheless possible to estimate it consistently along lines similar to that followed in the case of CCEMG. Using (43) and (64), we first note that

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

and because $(\boldsymbol{v}_i - \bar{\boldsymbol{v}})$ and $(\mathbf{h}_{iT} - \bar{\mathbf{h}}_T)$ are independently distributed across $i$, we then have

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

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

$$\hat{\mathbf{R}}^* = \frac{1}{N-1}\sum_{i=1}^N \tilde{w}_i^2\left(\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{X}_i}{T}\right)(\hat{\mathbf{b}}_i - \hat{\mathbf{b}}_{\mathrm{MG}})(\hat{\mathbf{b}}_i - \hat{\mathbf{b}}_{\mathrm{MG}})'\left(\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{X}_i}{T}\right). \tag{67}$$

Using (43) we also note that $\boldsymbol{\Psi}^*$ can be consistently estimated by

$$\hat{\boldsymbol{\Psi}}^* = \sum_{i=1}^N w_i\left(\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{X}_i}{T}\right). \tag{68}$$

Hence,

$$\widehat{\mathrm{AVar}}(\hat{\mathbf{b}}_P) = \left(\sum_{i=1}^N w_i^2\right)\hat{\boldsymbol{\Psi}}^{*-1}\hat{\mathbf{R}}^*\hat{\boldsymbol{\Psi}}^{*-1}. \tag{69}$$

Finally, the case where $\boldsymbol{\beta}_i$'s are homogeneous, namely when $\boldsymbol{\Omega}_{\boldsymbol{v}} = \mathbf{0}$, 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 (66). Under $\boldsymbol{\beta}_i = \boldsymbol{\beta}$, and using (B.1) and (B.3) we have (noting that in this case $\boldsymbol{v}_i = \mathbf{0}$)

$$\left(\frac{\sum_{i=1}^N w_i^2}{T}\right)^{-1/2}(\hat{\mathbf{b}}_P - \boldsymbol{\beta}) \overset{d}{\sim} \boldsymbol{\Psi}^{*-1}\left[\frac{1}{\sqrt{TN}}\sum_{i=1}^N \tilde{w}_i\mathbf{X}_i'\bar{\mathbf{M}}_w(\mathbf{F}\boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_i)\right], \tag{70}$$

where we have also multiplied both sides of (B.1) by $\sqrt{T}$ to avoid a degenerate asymptotic distribution. It is easily seen that $\hat{\mathbf{b}}_P$ continues to be consistent for $\boldsymbol{\beta}$ as 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.[10]

THEOREM 4: *Consider the panel data model* (1) *and* (2), *and suppose that Assumptions* 1–4 *and* 5(b) *hold*, $m = 1$, *the rank condition* (21) *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*

$$\left(\frac{\sum_{i=1}^N w_i^2}{T}\right)^{-1/2} (\hat{\mathbf{b}}_P - \boldsymbol{\beta}) \xrightarrow{d} \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_{PH}), \tag{71}$$

*where*

$$\boldsymbol{\Sigma}_{PH} = \boldsymbol{\Psi}^{-1}\dot{\mathbf{R}}\boldsymbol{\Psi}^{-1}, \tag{72}$$

$$\boldsymbol{\Psi} = \lim_{N\to\infty}\left(\sum_{i=1}^N w_i \boldsymbol{\Sigma}_i\right), \quad \dot{\mathbf{R}} = \lim_{N\to\infty}\left(\frac{1}{N}\sum_{i=1}^N \tilde{w}_i^2 \mathbf{S}_{i\varepsilon}\right), \tag{73}$$

$\mathbf{S}_{i\varepsilon}$ *is defined by* (48), *and*

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

Under the assumptions of Theorem 4, the asymptotic variance matrix of $\hat{\mathbf{b}}_P$ can be consistently estimated by

$$\widehat{\mathrm{AVar}}(\hat{\mathbf{b}}_P) = \frac{1}{T}\left(\sum_{i=1}^N w_i \hat{\boldsymbol{\Psi}}_{iT}\right)^{-1}\left(\sum_{i=1}^N w_i^2 \hat{\mathbf{S}}_{i\varepsilon}\right)\left(\sum_{i=1}^N w_i \hat{\boldsymbol{\Psi}}_{iT}\right)^{-1}, \tag{74}$$

where $\hat{\mathbf{S}}_{i\varepsilon}$ is given by (51) with $\hat{\boldsymbol{\Lambda}}_{ij}$ defined by (52) computed using $\tilde{\mathbf{e}}_i = \bar{\mathbf{M}}_w(\mathbf{y}_i - \mathbf{X}_i\hat{\mathbf{b}}_P)$ in place of $\hat{\mathbf{e}}_i = \bar{\mathbf{M}}_w(\mathbf{y}_i - \mathbf{X}_i\hat{\mathbf{b}}_i)$. In general, however, where the conditions of Theorem 4 might not be satisfied, one could use the nonparametric variance estimator of $\hat{\mathbf{b}}_P$ given by (69). The Monte Carlo experiments to be reported in Section 7 support such a strategy.

## 6. DETERMINATION OF OPTIMAL WEIGHTS

Our asymptotic results hold for all weights $w_i$ that satisfy the conditions in (14). Clearly, these conditions do not uniquely determine these weights and

[10]See Appendix B for a proof.

the issue of an optimal choice for $w_i$ 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 (14). For the individual coefficients $\hat{\mathbf{b}}_i$, the asymptotic variance matrix is given by (47) and does not depend on $w_i$'s, and the asymptotic 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 (53).

Consider now the CCE pooled estimator $\hat{\mathbf{b}}_P$ under slope homogeneity and, to simplify the exposition, assume that the idiosyncratic errors $\varepsilon_{it}$ are serially uncorrelated. The asymptotic variance matrix of $\hat{\mathbf{b}}_P$ in this case is given by

$$\mathrm{AVar}(\hat{\mathbf{b}}_P) = \frac{1}{T}\left(\sum_{i=1}^{N} w_i \boldsymbol{\Sigma}_i\right)^{-1}\left(\sum_{i=1}^{N} w_i^2 \sigma_i^2 \boldsymbol{\Sigma}_i\right)\left(\sum_{i=1}^{N} w_i \boldsymbol{\Sigma}_i\right)^{-1},$$

and is minimized with $w_i$ set at

$$w_i^* = \frac{\sigma_i^{-2}}{\sum_{j=1}^{N} \sigma_j^{-2}}, \tag{75}$$

yielding

$$\mathrm{AVar}(\hat{\mathbf{b}}_P(w^*)) = \frac{1}{T}\left(\sum_{i=1}^{N} \sigma_i^{-2} \boldsymbol{\Sigma}_i\right)^{-1}. \tag{76}$$

Noting that $\boldsymbol{\Sigma}_i$ is a positive definite matrix, we can write

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

where

$$\mathcal{X}_i = \sigma_i^{-1} \boldsymbol{\Sigma}_i^{1/2} \quad \text{and} \quad \mathcal{Y}_i = w_i \sigma_i \boldsymbol{\Sigma}_i^{1/2}.$$

This now establishes that $[\mathrm{AVar}(\hat{\mathbf{b}}_P(w^*))^{-1} - \mathrm{AVar}(\hat{\mathbf{b}}_P)^{-1}]$ is a nonnegative definite matrix, with $\{w_i^*\}$ providing an optimal choice in the sense that $\mathrm{AVar}(\hat{\mathbf{b}}_P(w^*)) \leq \mathrm{AVar}(\hat{\mathbf{b}}_P)$. Not surprisingly, the pooled estimator computed using $w_i^*$ reduces to the generalized least squares estimator

$$\hat{\mathbf{b}}_P(w^*) = \left(\sum_{i=1}^{N} \sigma_i^{-2} \mathbf{X}_i' \bar{\mathbf{M}}_{w^*} \mathbf{X}_i\right)^{-1} \sum_{i=1}^{N} \sigma_i^{-2} \mathbf{X}_i' \bar{\mathbf{M}}_{w^*} \mathbf{y}_i. \tag{77}$$

Recall, however, that for the pooled estimator to remain asymptotically valid, the weights used for the construction of the aggregates must be the same as those used in the formation of the pooled estimator.

## 7. 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 (53) and (65), respectively, using the weights $w_i = \theta_i = 1/N$ and the data generating process (DGP)

$$y_{it} = \alpha_{i1}d_{1t} + \beta_{i1}x_{1it} + \beta_{i2}x_{2it} + \gamma_{i1}f_{1t} + \gamma_{i2}f_{2t} + \varepsilon_{it} \tag{78}$$

and

$$x_{ijt} = a_{ij1}d_{1t} + a_{ij2}d_{2t} + \gamma_{ij1}f_{1t} + \gamma_{ij3}f_{3t} + v_{ijt} \qquad (j = 1, 2), \tag{79}$$

for $i = 1, 2, \dots, N$, and $t = 1, 2, \dots, T$. This DGP is a restricted version of the general linear model considered in 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 (1) and (2), and

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

imposed on the (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

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

$$f_{jt} = \rho_{fj} f_{jt-1} + v_{fjt} \quad \text{for} \quad j = 1, 2, 3, \quad t = -49, \dots, 0, \dots, T,$$
$$v_{fjt} \sim \text{IIDN}(0, 1 - \rho_{fj}^2), \quad \rho_{fj} = 0.5, \quad f_{j,-50} = 0 \quad \text{for} \quad j = 1, 2, 3,$$

$$v_{ijt} = \rho_{vij} v_{ijt-1} + \boldsymbol{v}_{ijt}, \quad t = -49, \dots, 1, \dots, T,$$
$$\boldsymbol{v}_{ijt} \sim \text{IIDN}(0, 1 - \rho_{vij}^2), \quad v_{ji,-50} = 0,$$

and

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

The individual-specific errors of $y_{it}$ are generated as stationary AR(1) processes for half of the cross-section units

$$\varepsilon_{it} = \rho_{i\varepsilon}\varepsilon_{i,t-1} + \sigma_i(1-\rho_{i\varepsilon}^2)^{1/2}\zeta_{it} \qquad (i = 1, 2, \ldots, [N/2]), \tag{80}$$

and as MA(1) processes for the remaining cross-section units

$$\varepsilon_{it} = \sigma_i(1+\theta_{i\varepsilon}^2)^{-1/2}(\zeta_{it} + \theta_{i\varepsilon}\zeta_{i,t-1}) \qquad (i = [N/2]+1, \ldots, N)$$

with $\zeta_{it} \sim \text{IIDN}(0,1)$, $\sigma_i^2 \sim \text{IIDU}[0.5, 1.5]$, $\rho_{i\varepsilon} \sim \text{IIDU}[0.05, 0.95]$, and $\theta_{i\varepsilon} \sim \text{IIDU}[0, 1]$. Note that the CCE type estimators are robust to the serial correlation and error variance heterogeneity of $\varepsilon_{it}$ across $i$ and this mixed time series specification is intended to highlight the robustness of the CCE type estimators in small samples.

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

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

For the parameters of the unobserved common effects in the $y_{it}$ equation $\boldsymbol{\gamma}_i$, we considered two different sets that we denote by $\mathcal{A}$ and $\mathcal{B}$. Under set $\mathcal{A}$, $\boldsymbol{\gamma}_i$ are drawn such that the rank condition (21) is satisfied, namely

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

and

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

Under set $\mathcal{B}$,

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

so that

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

and the rank condition is *not* satisfied. For each set we conducted two different experiments:[11]

[11]We also carried out a number of experiments with $\gamma_{ij} \sim \text{IIDN}(0.5, 0.2)$ for $j = 1, 2$ that give a lower degree of error cross-section dependence as compared to $\gamma_{ij} \sim \text{IIDN}(1, 0.2)$, but obtained

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

The two versions of Experiment 1 will be denoted 1a and 1b, and those of Experiment 2 will be denoted 2a, and 2b.[12] 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 exclude these factors. The infeasible MG (pooled) estimator provides an upper bound to the efficiency of the CCEMG (CCEP) estimator under slope heterogeneity (homogeneity), whereas 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 2,000 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.

The simulations results are summarized in Tables I–IV. Each table provides estimates of bias, root mean squared errors (RMSE), size, and power. Tables I and II summarize the results for the full rank experiments 1a (homogeneous slope) and 2a (heterogeneous slope), while the results for the rank deficient experiments (1b and 2b) are summarized in Tables III and IV.

### 7.1. *Bias and RMSE*

The results for the naive estimator are reported only in the case of Experiment 1a. As can be seen from Table I, not surprisingly, this estimator is substantially biased, performs very poorly, and is subject to large size distortions—an outcome that continues to apply in the case of other experiments, which are not reported here to save space. In contrast, the bias and RMSE of the CCEMG and CCEP estimators are very small and comparable to the bias of the associated infeasible estimators. Comparison of the results in Tables I and II with those summarized in Tables III and IV also show that the bias of the CCE type estimators does not depend on whether the rank condition (21) is satisfied.

In the case of Experiment 1a (full rank + heterogeneous slopes), 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 28.5% higher than that of the infeasible MG, falls steadily with $N$ and $T$, and ends up being only

very similar results. We decided to report the outcomes of the experiments with the higher cross-section dependence, because they are likely to provide a more demanding check on the validity of the CCE estimators.

[12]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. Failure of the order or rank condition does not seem to play a significant role in outcomes.

TABLE I

SMALL SAMPLE PROPERTIES OF COMMON CORRELATED EFFECTS TYPE ESTIMATORS IN THE CASE OF EXPERIMENT 1A (FULL RANK + HETEROGENEOUS SLOPES)

| | Bias (×100) | | | | | RMSE (×100) | | | | | Size (5% level, $H_0:\beta_1=1.00$) | | | | | Power (5% level, $H_1:\beta_1=0.95$) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $(N,T)$ | 20 | 30 | 50 | 100 | 200 | 20 | 30 | 50 | 100 | 200 | 20 | 30 | 50 | 100 | 200 | 20 | 30 | 50 | 100 | 200 |
| CCE type estimators | | | | | | | | | | | | | | | | | | | | |
| *CCEMG* | | | | | | | | | | | | | | | | | | | | |
| 20 | 0.18 | −0.16 | −0.08 | 0.06 | −0.10 | 9.73 | 7.84 | 6.52 | 5.59 | 5.15 | 7.95 | 6.90 | 7.15 | 7.25 | 7.10 | 11.60 | 13.85 | 15.70 | 18.55 | 20.65 |
| 30 | −0.18 | 0.02 | −0.02 | −0.09 | 0.09 | 7.42 | 6.02 | 5.11 | 4.41 | 4.10 | 6.85 | 6.05 | 6.50 | 6.50 | 5.90 | 11.60 | 15.90 | 19.10 | 22.75 | 26.80 |
| 50 | −0.05 | 0.15 | −0.07 | 0.15 | −0.04 | 5.78 | 4.62 | 3.96 | 3.41 | 3.11 | 6.25 | 5.90 | 6.75 | 6.30 | 5.90 | 15.10 | 21.45 | 25.10 | 34.60 | 37.10 |
| 100 | 0.02 | 0.02 | 0.03 | 0.04 | 0.03 | 4.06 | 3.48 | 2.83 | 2.33 | 2.26 | 5.05 | 5.90 | 5.65 | 5.15 | 6.35 | 24.90 | 33.20 | 43.45 | 55.95 | 62.40 |
| 200 | −0.08 | −0.03 | −0.01 | 0.05 | 0.00 | 3.07 | 2.44 | 1.96 | 1.71 | 1.51 | 5.75 | 5.60 | 5.50 | 5.35 | 5.05 | 37.15 | 52.90 | 70.55 | 84.70 | 89.95 |
| *CCEP* | | | | | | | | | | | | | | | | | | | | |
| 20 | 0.26 | −0.13 | −0.03 | 0.02 | −0.12 | 8.70 | 7.42 | 6.44 | 5.70 | 5.36 | 8.00 | 7.75 | 7.45 | 7.65 | 7.10 | 12.45 | 14.05 | 16.00 | 18.15 | 20.20 |
| 30 | −0.23 | −0.04 | 0.01 | −0.09 | 0.11 | 6.99 | 5.91 | 5.21 | 4.52 | 4.24 | 6.45 | 6.35 | 7.20 | 6.45 | 6.80 | 13.15 | 15.80 | 19.15 | 21.75 | 27.75 |
| 50 | −0.05 | 0.16 | −0.04 | 0.13 | −0.01 | 5.27 | 4.52 | 3.98 | 3.43 | 3.19 | 6.25 | 6.15 | 6.30 | 6.00 | 6.25 | 17.00 | 21.90 | 26.25 | 33.50 | 37.15 |
| 100 | 0.08 | 0.03 | 0.01 | 0.01 | 0.02 | 3.73 | 3.31 | 2.84 | 2.35 | 2.28 | 4.90 | 6.00 | 5.30 | 5.20 | 6.15 | 28.70 | 35.45 | 44.00 | 54.10 | 61.60 |
| 200 | −0.05 | −0.05 | −0.04 | 0.04 | 0.01 | 2.69 | 2.34 | 1.95 | 1.70 | 1.53 | 4.80 | 5.55 | 4.85 | 5.10 | 4.60 | 45.20 | 56.55 | 70.85 | 83.80 | 89.35 |
| Infeasible estimators (including $f_{1t}$ and $f_{2t}$) | | | | | | | | | | | | | | | | | | | | |
| *Mean group* | | | | | | | | | | | | | | | | | | | | |
| 20 | −0.07 | −0.15 | −0.15 | 0.15 | −0.10 | 7.58 | 6.60 | 5.76 | 5.11 | 4.78 | 6.85 | 6.90 | 7.00 | 6.50 | 6.45 | 13.20 | 14.70 | 16.65 | 20.15 | 18.80 |
| 30 | −0.11 | −0.03 | 0.00 | −0.03 | 0.12 | 5.87 | 5.00 | 4.53 | 4.01 | 3.88 | 6.00 | 5.10 | 5.70 | 5.40 | 5.75 | 15.70 | 18.60 | 22.40 | 24.45 | 27.45 |
| 50 | 0.05 | 0.09 | −0.06 | 0.13 | −0.03 | 4.47 | 3.82 | 3.46 | 3.15 | 2.98 | 6.55 | 5.55 | 6.20 | 5.25 | 5.35 | 22.30 | 26.95 | 30.40 | 37.50 | 39.95 |
| 100 | 0.02 | 0.03 | 0.03 | 0.01 | 0.04 | 3.15 | 2.86 | 2.50 | 2.17 | 2.17 | 4.80 | 5.50 | 4.85 | 5.05 | 5.45 | 35.15 | 45.40 | 53.00 | 60.15 | 65.70 |
| 200 | −0.05 | 0.02 | −0.04 | 0.06 | 0.01 | 2.26 | 1.98 | 1.71 | 1.59 | 1.46 | 4.80 | 5.25 | 4.45 | 5.75 | 4.80 | 58.00 | 71.60 | 80.55 | 89.40 | 91.85 |
| *Pooled* | | | | | | | | | | | | | | | | | | | | |
| 20 | 0.15 | −0.19 | −0.20 | −0.05 | −0.09 | 7.22 | 6.71 | 6.44 | 5.98 | 5.76 | 6.60 | 7.25 | 7.20 | 7.40 | 7.20 | 13.30 | 14.40 | 16.60 | 18.55 | 17.45 |
| 30 | −0.13 | −0.10 | 0.07 | 0.03 | 0.13 | 6.02 | 5.39 | 5.16 | 4.66 | 4.56 | 6.90 | 5.10 | 6.80 | 5.50 | 5.65 | 15.65 | 16.80 | 19.70 | 20.60 | 22.80 |
| 50 | 0.16 | 0.15 | −0.05 | 0.14 | −0.01 | 4.50 | 4.11 | 3.81 | 3.62 | 3.50 | 5.95 | 6.25 | 6.05 | 5.60 | 5.85 | 23.05 | 25.80 | 27.70 | 31.55 | 31.60 |
| 100 | −0.06 | 0.03 | 0.03 | 0.01 | 0.05 | 3.15 | 3.06 | 2.78 | 2.57 | 2.56 | 5.15 | 6.00 | 4.95 | 4.85 | 5.35 | 34.75 | 41.25 | 44.00 | 48.85 | 53.25 |
| 200 | −0.08 | 0.00 | −0.06 | 0.06 | −0.01 | 2.29 | 2.12 | 1.90 | 1.86 | 1.72 | 5.00 | 5.45 | 4.65 | 5.05 | 4.60 | 58.55 | 66.55 | 72.70 | 77.80 | 81.05 |

*Continues*

TABLE I—*Continued*

| | Bias (×100) | | | | | RMSE (×100) | | | | | Size (5% level, $H_0:\beta_1=1.00$) | | | | | Power (5% level, $H_1:\beta_1=0.95$) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $(N,T)$ | 20 | 30 | 50 | 100 | 200 | 20 | 30 | 50 | 100 | 200 | 20 | 30 | 50 | 100 | 200 | 20 | 30 | 50 | 100 | 200 |
| Naïve estimators (excluding $f_{1t}$ and $f_{2t}$) | | | | | | | | | | | | | | | | | | | | |
| *Mean group* | | | | | | | | | | | | | | | | | | | | |
| 20 | 14.73 | 14.21 | 14.05 | 14.11 | 13.90 | 19.45 | 18.05 | 17.08 | 16.51 | 16.02 | 31.95 | 34.20 | 39.25 | 44.45 | 48.10 | 47.00 | 51.30 | 58.35 | 66.65 | 70.15 |
| 30 | 15.64 | 15.23 | 15.35 | 15.07 | 15.06 | 19.44 | 18.00 | 17.57 | 16.79 | 16.50 | 43.15 | 47.80 | 56.00 | 60.70 | 65.55 | 60.95 | 68.75 | 76.75 | 83.05 | 87.30 |
| 50 | 14.85 | 14.58 | 13.94 | 14.14 | 14.02 | 18.10 | 17.08 | 15.86 | 15.40 | 15.03 | 58.20 | 64.25 | 66.75 | 76.80 | 81.55 | 76.15 | 82.25 | 86.50 | 94.30 | 96.50 |
| 100 | 14.64 | 15.08 | 14.79 | 14.61 | 14.43 | 17.01 | 16.90 | 16.04 | 15.44 | 15.03 | 72.90 | 81.35 | 88.45 | 94.85 | 97.35 | 89.50 | 94.45 | 98.65 | 99.55 | 99.85 |
| 200 | 14.91 | 14.89 | 14.62 | 14.54 | 14.49 | 17.08 | 16.45 | 15.65 | 15.12 | 14.88 | 85.30 | 92.00 | 96.10 | 99.20 | 99.85 | 95.05 | 98.65 | 99.70 | 100.00 | 100.00 |
| *Pooled* | | | | | | | | | | | | | | | | | | | | |
| 20 | 14.93 | 14.55 | 14.75 | 14.79 | 14.77 | 19.74 | 18.49 | 17.88 | 17.21 | 16.93 | 38.80 | 40.10 | 45.35 | 47.45 | 50.50 | 52.85 | 58.00 | 63.50 | 68.25 | 71.05 |
| 30 | 16.81 | 16.64 | 17.05 | 17.06 | 16.97 | 20.83 | 19.65 | 19.29 | 18.88 | 18.41 | 51.05 | 55.95 | 62.45 | 68.70 | 72.00 | 66.70 | 73.60 | 80.50 | 85.70 | 89.60 |
| 50 | 16.47 | 16.36 | 15.83 | 16.30 | 16.25 | 20.20 | 19.19 | 17.95 | 17.64 | 17.29 | 65.95 | 70.75 | 73.75 | 82.80 | 87.60 | 79.95 | 85.05 | 89.45 | 95.80 | 97.15 |
| 100 | 15.81 | 16.67 | 16.56 | 16.52 | 16.48 | 18.82 | 18.89 | 18.02 | 17.48 | 17.15 | 77.15 | 84.75 | 91.15 | 95.85 | 98.10 | 90.50 | 94.50 | 98.30 | 99.45 | 99.85 |
| 200 | 16.08 | 16.44 | 16.37 | 16.51 | 16.57 | 18.89 | 18.53 | 17.75 | 17.24 | 17.04 | 85.50 | 91.75 | 95.70 | 99.20 | 99.90 | 95.05 | 98.20 | 99.65 | 100.00 | 100.00 |

*Notes*: The DGP is $y_{it}=\alpha_{i1}d_{1t}+\beta_{i1}x_{1it}+\beta_{i2}x_{2it}+\gamma_{i1}f_{1t}+\gamma_{i2}f_{2t}+\varepsilon_{it}$ with $\varepsilon_{it}=\rho_{i\varepsilon}\varepsilon_{i,t-1}+\sigma_i(1-\rho_{i\varepsilon}^2)^{1/2}\zeta_{it}$, $i=1,2,\ldots,[N/2]$, and $\varepsilon_{it}=\sigma_i(1+\theta_{i\varepsilon}^2)^{-1/2}(\zeta_{it}+\theta_{i\varepsilon}\zeta_{i,t-1})$, $i=[N/2]+1,\ldots,N$, $\zeta_{it}\sim \text{IIDN}(0,1)$, $\sigma_i^2\sim \text{IIDU}[0.5,1.5]$, $\rho_{i\varepsilon}\sim \text{IIDU}[0.05,0.95]$, and $\theta_{i\varepsilon}\sim \text{IIDU}[0,1]$. Regressors are generated by $x_{ijt}=a_{ij1}d_{1t}+a_{ij2}d_{2t}+\gamma_{ij1}f_{1t}+\gamma_{ij3}f_{3t}+\mathrm{v}_{ijt}$, $j=1,2$, for $i=1,2,\ldots,N$; $d_{1t}=1$, $d_{2t}=0.5d_{2,t-1}+v_{dt}$, $v_{dt}\sim \text{IIDN}(0,1-0.5^2)$, $d_{2,-50}=0$; $f_{jt}=0.5f_{jt-1}+v_{fjt}$, $v_{fjt}\sim \text{IIDN}(0,1-0.5^2)$, $f_{j,-50}=0$ for $j=1,2,3$; $\mathrm{v}_{ijt}=\rho_{\upsilon ij}\mathrm{v}_{ijt-1}+\upsilon_{ijt}$, $\upsilon_{ijt}\sim \text{IIDN}(0,1-\rho_{\upsilon ij}^2)$, $\mathrm{v}_{ij,-50}=0$, and $\rho_{\upsilon ij}\sim \text{IIDU}[0.05,0.95]$ for $j=1,2$, for $t=-49,\ldots,T$ with the first 50 observations discarded; $\alpha_{i1}\sim \text{IIDN}(1,1)$; $a_{ij\ell}\sim \text{IIDN}(0.5,0.5)$ for $j=1,2$, $\ell=1,2$; $\gamma_{i11}$ and $\gamma_{i23}\sim \text{IIDN}(0.5,0.50)$, $\gamma_{i13}$ and $\gamma_{i21}\sim \text{IIDN}(0,0.50)$; $\gamma_{i1}$ and $\gamma_{i2}\sim \text{IIDN}(1,0.2)$; $\beta_{ij}=1+\eta_{ij}$ with $\eta_{ij}\sim \text{IIDN}(0,0.04)$ for $j=1,2$; $\rho_{\upsilon ij}$, $\rho_{i\varepsilon}$, $\theta_{i\varepsilon}$, $\sigma_i^2$, $\alpha_{i1}$, $a_{ij\ell}$ for $j=1,2$, $\ell=1,2$ are fixed across replications. CCEMG and CCEP are defined by (53) and (65), using the weights $w_i=\theta_i=1/N$, and their standard errors are defined by (58) and (69), respectively. These notes also apply to Tables II–IV.

TABLE II

SMALL SAMPLE PROPERTIES OF COMMON CORRELATED EFFECTS TYPE ESTIMATORS IN THE CASE OF EXPERIMENT 2A (FULL RANK + HETEROGENEOUS SLOPES)[a]

| | Bias (×100) | | | | | RMSE (×100) | | | | | Size (5% level, $H_0:\beta_1=1.00$) | | | | | Power (5% level, $H_1:\beta_1=0.95$) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $(N,T)$ | 20 | 30 | 50 | 100 | 200 | 20 | 30 | 50 | 100 | 200 | 20 | 30 | 50 | 100 | 200 | 20 | 30 | 50 | 100 | 200 |
| CCE type estimators | | | | | | | | | | | | | | | | | | | | |
| *CCEMG* | | | | | | | | | | | | | | | | | | | | |
| 20 | 0.04 | −0.15 | 0.08 | −0.08 | 0.07 | 8.27 | 6.27 | 4.76 | 3.33 | 2.52 | 6.80 | 6.65 | 6.60 | 6.65 | 6.25 | 12.05 | 15.50 | 23.15 | 36.35 | 55.90 |
| 30 | −0.16 | 0.10 | 0.08 | 0.00 | 0.00 | 6.37 | 4.89 | 3.66 | 2.56 | 1.90 | 6.05 | 5.90 | 6.25 | 6.05 | 6.60 | 13.75 | 21.15 | 32.40 | 53.80 | 77.00 |
| 50 | −0.01 | −0.08 | 0.05 | 0.03 | 0.03 | 5.00 | 3.67 | 2.70 | 1.86 | 1.34 | 5.80 | 5.50 | 4.75 | 5.65 | 5.30 | 18.95 | 29.20 | 48.85 | 78.30 | 95.85 |
| 100 | −0.11 | −0.09 | 0.06 | −0.05 | −0.03 | 3.61 | 2.70 | 2.00 | 1.34 | 0.97 | 5.35 | 5.45 | 5.85 | 5.20 | 5.85 | 28.25 | 45.90 | 72.40 | 95.90 | 99.85 |
| 200 | 0.00 | 0.01 | 0.01 | 0.02 | 0.00 | 2.72 | 1.95 | 1.49 | 0.96 | 0.66 | 5.60 | 4.65 | 6.80 | 5.25 | 5.85 | 45.70 | 72.15 | 93.00 | 99.85 | 100.00 |
| *CCEP* | | | | | | | | | | | | | | | | | | | | |
| 20 | 0.14 | 0.09 | 0.14 | −0.05 | 0.06 | 6.82 | 5.47 | 4.37 | 3.19 | 2.62 | 6.85 | 6.20 | 6.60 | 6.60 | 6.55 | 15.10 | 18.20 | 26.45 | 37.70 | 54.70 |
| 30 | −0.18 | 0.04 | 0.07 | −0.02 | −0.01 | 5.31 | 4.42 | 3.49 | 2.52 | 1.94 | 5.75 | 6.85 | 6.60 | 6.40 | 7.10 | 16.20 | 24.05 | 36.80 | 55.20 | 75.40 |
| 50 | −0.03 | 0.07 | 0.00 | 0.01 | 0.03 | 4.19 | 3.31 | 2.53 | 1.79 | 1.34 | 6.10 | 5.45 | 5.05 | 5.60 | 5.25 | 23.90 | 36.30 | 51.90 | 79.10 | 96.00 |
| 100 | −0.04 | −0.06 | 0.03 | −0.06 | −0.04 | 2.90 | 2.38 | 1.81 | 1.25 | 0.94 | 5.90 | 6.15 | 6.45 | 5.35 | 5.70 | 39.15 | 57.05 | 80.75 | 97.50 | 99.85 |
| 200 | 0.04 | 0.01 | 0.00 | 0.02 | 0.00 | 2.16 | 1.65 | 1.33 | 0.92 | 0.65 | 4.80 | 4.20 | 6.25 | 5.85 | 5.60 | 65.05 | 84.70 | 96.75 | 99.90 | 100.00 |
| Infeasible estimators (including $f_{1t}$ and $f_{2t}$) | | | | | | | | | | | | | | | | | | | | |
| *Mean group* | | | | | | | | | | | | | | | | | | | | |
| 20 | −0.11 | 0.00 | 0.02 | 0.01 | −0.01 | 6.10 | 4.74 | 3.56 | 2.38 | 1.72 | 6.80 | 6.30 | 6.10 | 6.00 | 6.25 | 16.75 | 24.40 | 34.30 | 58.95 | 83.50 |
| 30 | −0.06 | 0.03 | 0.03 | 0.03 | −0.03 | 4.42 | 3.48 | 2.59 | 1.80 | 1.28 | 5.95 | 6.00 | 5.40 | 5.95 | 6.05 | 22.00 | 32.40 | 51.90 | 78.40 | 96.50 |
| 50 | 0.03 | −0.07 | 0.04 | 0.06 | 0.01 | 3.39 | 2.61 | 1.95 | 1.30 | 0.93 | 6.05 | 5.65 | 5.05 | 4.60 | 4.05 | 33.50 | 48.05 | 72.85 | 95.75 | 99.95 |
| 100 | −0.09 | −0.08 | 0.03 | −0.04 | −0.03 | 2.53 | 1.92 | 1.47 | 1.00 | 0.74 | 5.85 | 5.10 | 5.20 | 4.55 | 5.50 | 50.90 | 71.50 | 92.55 | 99.65 | 100.00 |
| 200 | 0.04 | −0.01 | 0.00 | −0.01 | 0.00 | 1.80 | 1.39 | 1.05 | 0.70 | 0.49 | 5.45 | 5.80 | 6.40 | 5.50 | 4.75 | 79.65 | 94.85 | 99.60 | 100.00 | 100.00 |
| *Pooled* | | | | | | | | | | | | | | | | | | | | |
| 20 | −0.08 | 0.03 | 0.06 | 0.02 | −0.01 | 4.07 | 3.29 | 2.45 | 1.69 | 1.27 | 6.15 | 6.90 | 6.15 | 5.90 | 6.65 | 23.00 | 35.10 | 51.55 | 80.50 | 96.00 |
| 30 | 0.00 | 0.02 | 0.04 | 0.00 | −0.01 | 3.50 | 2.75 | 2.16 | 1.50 | 1.09 | 5.75 | 5.25 | 5.70 | 5.90 | 5.70 | 31.65 | 45.30 | 64.80 | 89.60 | 99.20 |
| 50 | 0.10 | 0.01 | 0.01 | 0.03 | 0.01 | 2.68 | 2.16 | 1.61 | 1.13 | 0.80 | 5.45 | 5.25 | 5.20 | 5.15 | 4.65 | 48.50 | 64.70 | 86.20 | 98.60 | 100.00 |
| 100 | −0.05 | −0.03 | 0.01 | −0.05 | −0.02 | 1.83 | 1.49 | 1.14 | 0.82 | 0.60 | 5.25 | 5.35 | 4.90 | 4.55 | 5.80 | 76.35 | 90.50 | 99.05 | 100.00 | 100.00 |
| 200 | 0.05 | −0.03 | −0.01 | 0.00 | 0.00 | 1.36 | 1.09 | 0.84 | 0.58 | 0.41 | 5.80 | 5.00 | 5.25 | 5.60 | 4.90 | 96.00 | 99.40 | 99.95 | 100.00 | 100.00 |

[a]The notes to Table I also apply to this table.

TABLE III

SMALL SAMPLE PROPERTIES OF COMMON CORRELATED EFFECTS TYPE ESTIMATORS IN THE CASE OF EXPERIMENT 1B (RANK DEFICIENT + HETEROGENEOUS SLOPES)[a]

| | Bias (×100) | | | | | RMSE (×100) | | | | | Size (5% level, $H_0: \beta_1 = 1.00$) | | | | | Power (5% level, $H_1: \beta_1 = 0.95$) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $(N, T)$ | 20 | 30 | 50 | 100 | 200 | 20 | 30 | 50 | 100 | 200 | 20 | 30 | 50 | 100 | 200 | 20 | 30 | 50 | 100 | 200 |
| CCE type estimators | | | | | | | | | | | | | | | | | | | | |
| *CCEMG* | | | | | | | | | | | | | | | | | | | | |
| 20 | −0.26 | −0.50 | 0.03 | −0.12 | −0.21 | 12.00 | 9.53 | 7.79 | 6.28 | 5.56 | 6.25 | 6.75 | 7.40 | 7.25 | 7.20 | 8.70 | 9.05 | 13.45 | 15.55 | 17.40 |
| 30 | −0.36 | 0.08 | −0.07 | 0.12 | 0.04 | 9.60 | 7.71 | 6.32 | 5.07 | 4.49 | 5.25 | 5.95 | 6.85 | 6.40 | 6.70 | 8.45 | 11.55 | 15.25 | 19.20 | 22.70 |
| 50 | 0.25 | 0.08 | −0.03 | 0.14 | −0.02 | 7.34 | 5.97 | 4.77 | 4.00 | 3.44 | 5.55 | 5.80 | 5.75 | 6.75 | 6.10 | 11.90 | 14.60 | 20.80 | 28.30 | 32.35 |
| 100 | 0.04 | 0.00 | 0.07 | 0.05 | 0.02 | 5.65 | 4.39 | 3.56 | 2.73 | 2.46 | 6.15 | 6.40 | 6.10 | 5.10 | 5.90 | 17.15 | 24.50 | 33.05 | 45.70 | 55.50 |
| 200 | 0.03 | 0.07 | 0.03 | −0.01 | −0.02 | 4.04 | 3.08 | 2.41 | 1.97 | 1.68 | 5.20 | 5.50 | 5.05 | 5.40 | 4.75 | 25.20 | 38.65 | 54.30 | 72.00 | 83.45 |
| *CCEP* | | | | | | | | | | | | | | | | | | | | |
| 20 | 0.13 | −0.57 | 0.07 | −0.18 | −0.23 | 10.46 | 9.00 | 7.47 | 6.37 | 5.69 | 6.25 | 7.30 | 8.30 | 8.05 | 8.05 | 10.50 | 10.20 | 13.40 | 15.70 | 17.75 |
| 30 | −0.25 | 0.07 | 0.01 | 0.13 | 0.06 | 8.58 | 7.29 | 6.24 | 5.13 | 4.59 | 5.55 | 5.70 | 7.35 | 6.85 | 7.05 | 9.85 | 12.50 | 15.95 | 18.85 | 22.75 |
| 50 | 0.17 | 0.04 | 0.04 | 0.14 | −0.01 | 6.71 | 5.57 | 4.71 | 4.01 | 3.48 | 5.80 | 6.30 | 6.25 | 6.65 | 5.90 | 13.60 | 15.65 | 22.30 | 28.70 | 31.60 |
| 100 | 0.05 | −0.04 | 0.07 | 0.05 | 0.01 | 4.74 | 4.12 | 3.39 | 2.72 | 2.47 | 4.65 | 5.95 | 5.70 | 5.55 | 5.90 | 19.55 | 26.45 | 34.80 | 46.05 | 54.40 |
| 200 | −0.03 | 0.03 | −0.02 | −0.01 | −0.02 | 3.49 | 2.84 | 2.32 | 1.95 | 1.69 | 5.70 | 4.85 | 4.60 | 5.30 | 5.15 | 32.75 | 44.05 | 56.55 | 73.50 | 83.05 |
| Infeasible estimators (including $f_{1t}$ and $f_{2t}$) | | | | | | | | | | | | | | | | | | | | |
| *Mean group* | | | | | | | | | | | | | | | | | | | | |
| 20 | 0.13 | −0.28 | −0.10 | 0.05 | −0.14 | 7.54 | 6.42 | 5.74 | 5.07 | 4.84 | 6.25 | 5.95 | 6.55 | 7.05 | 6.85 | 13.65 | 12.85 | 15.80 | 19.05 | 20.15 |
| 30 | −0.15 | −0.10 | 0.02 | 0.07 | 0.09 | 5.69 | 4.93 | 4.57 | 4.00 | 3.87 | 4.40 | 5.60 | 6.75 | 5.35 | 5.80 | 14.60 | 16.65 | 21.95 | 24.95 | 27.50 |
| 50 | 0.14 | 0.03 | 0.00 | 0.06 | −0.02 | 4.33 | 3.84 | 3.50 | 3.18 | 2.96 | 4.95 | 5.30 | 6.05 | 6.05 | 5.40 | 22.65 | 26.85 | 32.15 | 36.50 | 38.85 |
| 100 | −0.01 | −0.04 | 0.05 | −0.01 | 0.01 | 3.21 | 2.91 | 2.47 | 2.19 | 2.16 | 4.90 | 6.55 | 4.95 | 4.95 | 5.60 | 35.65 | 44.50 | 54.00 | 60.45 | 64.80 |
| 200 | 0.11 | 0.01 | −0.03 | 0.05 | 0.00 | 2.27 | 1.99 | 1.72 | 1.61 | 1.46 | 4.75 | 4.65 | 5.00 | 5.30 | 4.30 | 61.35 | 70.35 | 81.40 | 88.30 | 92.00 |
| *Pooled* | | | | | | | | | | | | | | | | | | | | |
| 20 | 0.25 | −0.28 | −0.18 | −0.11 | −0.12 | 7.18 | 6.65 | 6.47 | 5.96 | 5.76 | 6.45 | 7.15 | 7.65 | 7.25 | 7.00 | 13.35 | 14.90 | 16.25 | 17.10 | 17.35 |
| 30 | −0.14 | −0.03 | 0.08 | 0.08 | 0.11 | 5.87 | 5.39 | 5.15 | 4.68 | 4.54 | 6.00 | 6.40 | 6.80 | 6.10 | 6.00 | 14.60 | 17.40 | 18.90 | 21.45 | 22.00 |
| 50 | 0.22 | 0.10 | 0.02 | 0.09 | −0.03 | 4.49 | 4.18 | 3.83 | 3.62 | 3.48 | 5.95 | 6.10 | 6.35 | 6.05 | 5.65 | 23.50 | 26.20 | 27.90 | 31.10 | 32.40 |
| 100 | −0.04 | 0.00 | 0.03 | 0.01 | 0.02 | 3.12 | 3.09 | 2.76 | 2.57 | 2.55 | 4.65 | 5.75 | 4.40 | 4.70 | 5.95 | 34.90 | 39.20 | 44.50 | 49.10 | 52.45 |
| 200 | 0.05 | 0.00 | −0.05 | 0.06 | −0.01 | 2.27 | 2.12 | 1.93 | 1.86 | 1.73 | 4.75 | 5.65 | 4.90 | 5.20 | 4.75 | 60.25 | 65.55 | 72.05 | 77.40 | 80.55 |

[a]The notes to Table I also apply to this table.

TABLE IV

SMALL SAMPLE PROPERTIES OF COMMON CORRELATED EFFECTS TYPE ESTIMATORS IN THE CASE OF EXPERIMENT 2B (RANK DEFICIENT + HETEROGENEOUS SLOPE)[a]

| | Bias (×100) | | | | | RMSE (×100) | | | | | Size (5% level, $H_0:\beta_1=1.00$) | | | | | Power (5% level, $H_1:\beta_1=0.95$) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $(N,T)$ | 20 | 30 | 50 | 100 | 200 | 20 | 30 | 50 | 100 | 200 | 20 | 30 | 50 | 100 | 200 | 20 | 30 | 50 | 100 | 200 |
| CCE type estimators | | | | | | | | | | | | | | | | | | | | |
| *CCEMG* | | | | | | | | | | | | | | | | | | | | |
| 20 | −0.23 | 0.12 | 0.26 | −0.04 | 0.05 | 11.36 | 8.55 | 6.38 | 4.55 | 3.39 | 6.60 | 6.45 | 6.20 | 6.80 | 6.80 | 9.80 | 12.15 | 17.15 | 25.55 | 39.80 |
| 30 | −0.05 | 0.07 | −0.17 | 0.03 | −0.06 | 8.97 | 6.95 | 5.17 | 3.59 | 2.62 | 5.50 | 5.80 | 5.95 | 5.75 | 5.65 | 10.80 | 13.60 | 19.05 | 33.00 | 51.10 |
| 50 | 0.00 | −0.04 | −0.03 | −0.03 | −0.03 | 6.99 | 5.24 | 3.85 | 2.71 | 1.84 | 5.60 | 5.95 | 5.95 | 5.85 | 4.25 | 13.05 | 17.85 | 28.50 | 48.80 | 75.75 |
| 100 | 0.00 | −0.13 | 0.08 | 0.04 | −0.01 | 4.94 | 3.78 | 2.80 | 1.95 | 1.33 | 5.00 | 4.85 | 5.30 | 5.35 | 5.55 | 19.25 | 27.45 | 46.85 | 76.00 | 95.65 |
| 200 | −0.04 | 0.17 | −0.01 | 0.00 | 0.00 | 3.90 | 2.77 | 2.06 | 1.35 | 0.94 | 5.60 | 5.05 | 5.95 | 5.30 | 5.35 | 27.75 | 47.50 | 70.05 | 95.25 | 99.90 |
| *CCEP* | | | | | | | | | | | | | | | | | | | | |
| 20 | −0.11 | 0.12 | 0.33 | −0.04 | 0.05 | 9.55 | 7.34 | 5.87 | 4.38 | 3.37 | 6.90 | 6.50 | 6.65 | 7.00 | 6.70 | 10.95 | 13.55 | 18.75 | 26.30 | 39.75 |
| 30 | 0.08 | −0.02 | −0.08 | −0.01 | −0.07 | 7.57 | 6.03 | 4.81 | 3.47 | 2.64 | 6.10 | 6.30 | 7.10 | 6.50 | 6.10 | 12.35 | 15.25 | 21.45 | 34.00 | 51.95 |
| 50 | 0.00 | −0.03 | −0.06 | −0.04 | −0.03 | 5.84 | 4.78 | 3.55 | 2.61 | 1.82 | 5.05 | 5.95 | 5.50 | 5.65 | 4.50 | 16.50 | 21.60 | 30.45 | 51.45 | 76.95 |
| 100 | −0.02 | −0.11 | 0.07 | 0.04 | −0.01 | 4.23 | 3.34 | 2.60 | 1.85 | 1.31 | 6.00 | 6.25 | 5.55 | 5.95 | 6.25 | 25.85 | 34.35 | 53.05 | 79.40 | 95.70 |
| 200 | −0.04 | 0.11 | −0.01 | 0.01 | 0.01 | 3.00 | 2.34 | 1.89 | 1.27 | 0.92 | 4.95 | 4.60 | 6.15 | 4.80 | 5.80 | 40.45 | 59.65 | 77.45 | 97.40 | 99.90 |
| Infeasible estimators (including $f_{1t}$ and $f_{2t}$) | | | | | | | | | | | | | | | | | | | | |
| *Mean group* | | | | | | | | | | | | | | | | | | | | |
| 20 | 0.22 | 0.12 | 0.06 | −0.07 | 0.05 | 5.90 | 4.73 | 3.48 | 2.43 | 1.76 | 5.95 | 6.30 | 6.05 | 6.25 | 7.30 | 18.30 | 23.70 | 36.00 | 58.20 | 83.20 |
| 30 | 0.10 | −0.03 | −0.05 | 0.00 | 0.00 | 4.45 | 3.41 | 2.51 | 1.83 | 1.28 | 6.20 | 6.20 | 4.90 | 6.60 | 6.35 | 22.85 | 32.40 | 50.90 | 79.20 | 97.40 |
| 50 | 0.09 | −0.12 | −0.04 | −0.02 | 0.03 | 3.39 | 2.65 | 1.95 | 1.36 | 0.96 | 5.40 | 5.80 | 5.50 | 5.35 | 5.40 | 34.55 | 47.60 | 72.20 | 95.05 | 100.00 |
| 100 | −0.05 | −0.05 | 0.01 | 0.01 | −0.01 | 2.43 | 1.91 | 1.45 | 1.01 | 0.73 | 4.95 | 4.70 | 5.00 | 4.50 | 5.70 | 51.80 | 72.85 | 92.45 | 99.75 | 100.00 |
| 200 | −0.01 | 0.02 | 0.00 | 0.00 | −0.01 | 1.81 | 1.37 | 1.04 | 0.71 | 0.50 | 4.85 | 4.85 | 5.55 | 4.50 | 6.05 | 79.00 | 94.95 | 99.75 | 100.00 | 100.00 |
| *Pooled* | | | | | | | | | | | | | | | | | | | | |
| 20 | 0.13 | 0.04 | 0.06 | 0.01 | 0.03 | 4.07 | 3.31 | 2.47 | 1.68 | 1.25 | 5.70 | 6.90 | 5.55 | 4.15 | 5.50 | 24.70 | 34.20 | 52.35 | 79.30 | 96.60 |
| 30 | 0.13 | −0.01 | −0.09 | 0.00 | 0.00 | 3.43 | 2.71 | 2.08 | 1.53 | 1.10 | 4.90 | 5.70 | 5.10 | 6.25 | 6.40 | 32.40 | 43.60 | 63.40 | 89.40 | 99.05 |
| 50 | 0.03 | −0.08 | −0.06 | −0.04 | 0.03 | 2.65 | 2.15 | 1.61 | 1.15 | 0.82 | 5.30 | 6.20 | 5.15 | 5.00 | 6.80 | 47.15 | 63.30 | 85.55 | 98.85 | 100.00 |
| 100 | 0.01 | −0.03 | 0.00 | 0.02 | −0.01 | 1.82 | 1.51 | 1.14 | 0.81 | 0.60 | 4.90 | 4.45 | 4.75 | 4.90 | 6.60 | 76.90 | 90.75 | 98.90 | 100.00 | 100.00 |
| 200 | −0.01 | 0.00 | 0.00 | −0.01 | −0.01 | 1.36 | 1.08 | 0.84 | 0.57 | 0.41 | 5.05 | 4.70 | 5.10 | 4.70 | 5.05 | 95.70 | 99.55 | 99.85 | 100.00 | 100.00 |

[a]The notes to Table I also apply to this table.

3.4% higher for $T = N = 200$. The Monte Carlo results also confirm the asymptotic efficiency of the MG type estimators relative to the pooled estimator under slope heterogeneity. This seems to occur for $(T, N) \geq 30$, although for relatively large samples, the differences between the two estimators are rather slight. It is also interesting to note that, in fact, the CCEP estimator 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 5% 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 III. In the rank deficient case, however, the efficiency loss of the CCEMG relative to the infeasible MG is higher, being 59.2% (compared to 28.5% under full rank) at $N = T = 20$ and 15.1% (compared to 3.4% under full rank) at $N = T = 200$.

The RMSE results for the homogeneous slope experiments, Experiments 2a and 2b, are summarized in Tables II and IV. 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 heterogeneous 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 IV.

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 setup. Therefore, within the realm of feasible estimators the choice is between CCEMG and CCEP. The simulation results favor the CCEP estimator for small to moderate sample sizes and slightly favor CCEMG when $N$ and $T$ are relatively large. This conclusion seems to be robust and 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. This issue is taken up in Kapetanios and Pesaran (2006).

### 7.2. *Size and Power*

For the full rank and heterogeneous experiments, Experiment 1a, size and power of a two-sided test of $\beta_1 = 1$ are reported in Table I. The variance of the CCEMG estimator is computed using (58), 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 oversized. As can be seen from the other tables, 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 2.

There are two alternative variance estimators for carrying out tests based on the CCEP estimator, namely the robust heterogeneous version given by (69) or the homogeneous version given by (74). It turns out that the heterogeneous version works well in both cases and has the added advantage that it does not depend on the choice of $p$, the size of the Bartlett window used in the computation of the Newey–West type estimator of $\mathbf{S}_{i\varepsilon}$. In view of this, all the test results for the pooled estimators (the CCEP or the pooled infeasible estimator) are based on the robust heterogeneous estimator (69).[13] The test results for the pooled estimators all have the correct size for $N, T \geq 20$ and the outcomes do not depend on whether the rank condition is satisfied.

The power of the various tests is computed under the alternative $\beta_1 = 0.95$ and reported in the last panels of Tables I–IV. Perhaps not surprisingly, given the homogeneous nature of the alternative, the pooled estimators tend to be more powerful than the MG type estimators, particularly for relatively small $N$ and $T$, although the performances of the two types of estimators converge as $N$ and $T$ are increased.

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 tends to be symmetric and have the familiar inverted bell shape. As an illustration, Figure 1 shows the power function of CCEMG and CCEP tests, as well as the associated infeasible tests, in the case of Experiment 1b for $N = 50$ and $T = 30$. The figure clearly shows that for this sample size, the CCEP test performs slightly better than the CCEMG test and compared with the tests based on the infeasible estimators, the two CCE tests seem to perform reasonably well.

## 8. 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

[13]Note that similar considerations also apply to the infeasible pooled estimator.

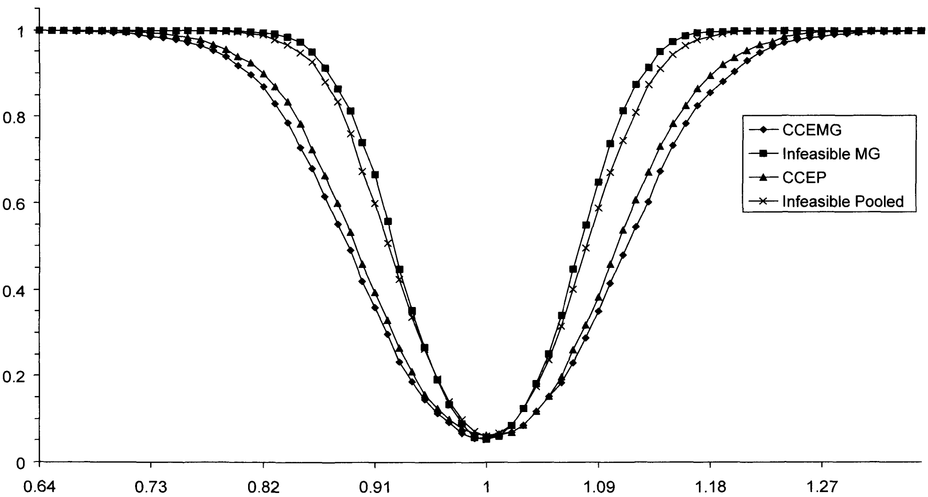


FIGURE 1.—Power function for Experiment 1b: $N = 50$, $T = 30$.

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

$$\hat{\mathbf{e}}_i = \bar{\mathbf{M}}_w(\mathbf{y}_i - \mathbf{X}_i\hat{\mathbf{b}}_i) \quad \text{or} \quad \tilde{\mathbf{e}}_i = \bar{\mathbf{M}}_w(\mathbf{y}_i - \mathbf{X}_i\hat{\mathbf{b}}_P).$$

Under our assumptions, for any fixed $m$ these residuals provide consistent estimates of $e_{it}$ in the multifactor model (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.

It is also of interest to compare the approach proposed in this paper with the alternative procedure that proxies the unobserved common factors by principal components (PC) of $y_{it}$ and $x_{it}$. This alternative is considered in a series of Monte Carlo experiments in Kapetanios and Pesaran (2006) and Coakley, Fuertes, and Smith (2006). Kapetanios and Pesaran's experiments allow for up to four regressors and factors, and find that the PC procedure does not perform as well as the CCE approach and leads to substantial size distortions even if, when using the PC procedure, the true number of unobserved factors is assumed to be known.

The regression model (1) allows for dynamics through the general dynamics of the common effects as well as the individual-specific dynamics in $e_{it}$. An alternative dynamic specification would be to allow for lagged values of $y_{it}$ to be included among the regressors. This issue is taken up in Pesaran (2005), who provides an application of the CCE approach to testing for unit roots in the presence of error cross-section dependence.

Another important extension is to multivariate panel data models such as panel vector autoregressions of the type discussed, for example, in Binder, Hsiao, and Pesaran (2005).

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

*Faculty of Economics, University of Cambridge, Sidgwick Avenue, Cambridge CB3 9DD, U.K., and USC; hashem.pesaran@econ.cam.ac.uk.*

*Manuscript received October, 2002; final revision received February, 2006.*

## APPENDIX A: STATEMENTS AND PROOFS OF LEMMAS

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

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

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

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

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

*where* $\bar{\mathbf{u}}_{wt} = \sum_{i=1}^{N} w_i \mathbf{u}_{it}$, $\mathbf{u}_{it}$ *is defined by* (5), *and the weights* $w_i$ *satisfy the conditions in* (14).

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 $\bar{\mathbf{v}}_{wt} = \sum_{i=1}^{N} \sum_{\ell=0}^{\infty} w_i \mathbf{S}_{i\ell} \boldsymbol{\nu}_{i,t-\ell}$. Whereas $\boldsymbol{\nu}_{it} \sim \text{IID}(\mathbf{0}, \mathbf{I}_k)$, then conditional on $w_i$ and $\mathbf{S}_{i\ell}$, $\text{Var}(\bar{\mathbf{v}}_{wt}) = \sum_{i=1}^{N} w_i^2 (\sum_{\ell=0}^{\infty} \mathbf{S}_{i\ell} \mathbf{S}_{i\ell}')$, and using (11) and (14) we have (unconditionally)

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

Similarly,

$$\text{Var}(\bar{\varepsilon}_{wt}) \leq \bar{\sigma}^2 \left( \sum_{i=1}^{N} w_i^2 \right) = O\left( \frac{1}{N} \right) \tag{A.7}$$

and

$$\text{Var}\left( \sum_{i=1}^{N} w_i \boldsymbol{\beta}_i' \mathbf{v}_{it} \right) = \sum_{i=1}^{N} w_i^2 E(\boldsymbol{\beta}_i' \boldsymbol{\Sigma}_i \boldsymbol{\beta}_i) \leq \sum_{i=1}^{N} w_i^2 E(\boldsymbol{\beta}_i' \boldsymbol{\beta}_i) \lambda_{\max}(\boldsymbol{\Sigma}_i),$$

where $\lambda_{\max}(\boldsymbol{\Sigma}_i)$ is the maximum eigenvalue of $\boldsymbol{\Sigma}_i$ that is bounded by Assumption 2. Also, either $\boldsymbol{\beta}_i' \boldsymbol{\beta}_i = \|\boldsymbol{\beta}_i\|^2 < K$ for each $i$ when $\boldsymbol{\beta}_i$ are treated as fixed or we have $E(\boldsymbol{\beta}_i' \boldsymbol{\beta}_i) = \boldsymbol{\beta}' \boldsymbol{\beta} + \text{Tr}(\boldsymbol{\Omega}_v) < K$ under the random coefficient assumption, Assumption 4. Therefore,

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

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

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

it also readily follows that

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

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

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

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

Furthermore,

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

*Q.E.D.*

LEMMA 2: *Suppose that either* $\|\boldsymbol{\beta}_i\| < K$ *for each* $i$ *or that the random coefficient assumption, Assumption* 4, *holds. Then under Assumptions* 1 *and* 2,

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

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

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

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

*where* $\bar{\mathbf{U}}_w = (\bar{\mathbf{u}}_{w1}, \bar{\mathbf{u}}_{w2}, \ldots, \bar{\mathbf{u}}_{wT})'$, $\bar{\mathbf{u}}_{wt}$ *is defined by* (A.5), *the weights* $w_i$ *satisfy the conditions in* (14), $\mathbf{V}_i = (\mathbf{v}_{i1}, \mathbf{v}_{i2}, \ldots, \mathbf{v}_{iT})'$, *and* $\mathbf{D}$ *and* $\mathbf{F}$ *are* $T \times n$ *and* $T \times m$ *data matrices on observed and unobserved common factors.*

PROOF: Note that $T^{-1}\bar{\mathbf{U}}'_w\bar{\mathbf{U}}_w = T^{-1}(\sum_{t=1}^{T}\bar{\mathbf{u}}_{wt}\bar{\mathbf{u}}'_{wt})$, 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, $E\|T^{-1}\bar{\mathbf{U}}'_w\bar{\mathbf{U}}_w\| \leq T^{-1}\sum_{t=1}^{T}E\|\bar{\mathbf{u}}_{wt}\|^2$ and, by (A.4), $E\|T^{-1}\bar{\mathbf{U}}'_w\bar{\mathbf{U}}_w\| = O(N^{-1})$, which establishes (A.10).

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

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

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

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

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

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

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

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

Also

$$E\left\|\frac{\sum_{t=1}^{T} f_{\ell t}\bar{\mathbf{u}}_{wt}}{T}\right\| \le \frac{\sum_{t=1}^{T} E\|f_{\ell t}\bar{\mathbf{u}}_{wt}\|}{T},$$

and using (A.4)

$$= O_p\left(\frac{1}{\sqrt{N}}\right) \quad \text{for any fixed } T.$$

Furthermore, because 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,

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

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 $\bar{\mathbf{V}}_w = \sum_{j=1}^N w_j\mathbf{V}_j$. Whereas, 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 $\bar{\mathbf{V}}_{w,-i}^* = \sum_{j=1,j\neq i}^N w_j\mathbf{V}_j\boldsymbol{\beta}_j$. Because $w_i = O(N^{-1})$, $\boldsymbol{\beta}_i$ is either bounded or satisfies 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 because 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, because 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. *Q.E.D.*

LEMMA 3: *Suppose that the conditions of Lemma* 2 *hold and* $\|\boldsymbol{\Pi}_i\| \leq K$, *where* $\boldsymbol{\Pi}_i = (\mathbf{A}_i', \boldsymbol{\Gamma}_i')'$ *and* $\mathbf{A}_i$ *and* $\boldsymbol{\Gamma}_i$ *are the parameters of the* $\mathbf{x}_{it}$ *process defined by* (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 (30) we have

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

and (A.22) follows from (A.11), (A.13), and the assumption that the elements of $\boldsymbol{\Pi}_i$ are bounded. *Q.E.D.*

LEMMA 4: *Suppose that Assumption* 3, *and conditions* (14) *and* (18) *hold, and that* $\mathbf{Q}_{iT}$ *is a* $k \times m$ *matrix, distributed independently of* $\boldsymbol{\eta}_i \sim \mathrm{IID}(\mathbf{0}, \boldsymbol{\Omega}_{\boldsymbol{\eta}})$, $\|\boldsymbol{\Omega}_{\boldsymbol{\eta}}\| < K$, *and* $E\|\mathbf{Q}_{iT}\| < K$. *Let*

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

*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* (12), (14), *and* (18), *respectively. Then*

$$\mathbf{q}_{NT} \xrightarrow{d} \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_{qT}), \quad as \quad N \to \infty,$$

*where*

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

*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 by observing that

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

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

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

and because by assumption,

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

*Q.E.D.*

## APPENDIX B: MATHEMATICAL PROOFS

PROOF OF THEOREM 3: Under (1) and (2), $\hat{\mathbf{b}}_P$ defined by (65) can be written as

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

where

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

and

$$\text{(B.3)}\quad\mathbf{q}_{NT}=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\tilde{\theta}_i\frac{(\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{F})\boldsymbol{\gamma}_i}{T}.$$

Using (12), we first note that $\boldsymbol{\gamma}_i = \bar{\boldsymbol{\gamma}}_w + \boldsymbol{\eta}_i - \bar{\boldsymbol{\eta}}_w$, where $\bar{\boldsymbol{\eta}}_w = \sum_{i=1}^N w_i \boldsymbol{\eta}_i$. Hence

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

In general, when the rank condition is not satisfied, $T^{-1}(\mathbf{X}_i' \bar{\mathbf{M}}_w \mathbf{F}) = O_p(1)$. (See (43).) Hence, the first term of $\mathbf{q}_{NT}$ will be unbounded unless $\theta_i = w_i$. However, when this condition is satisfied, because $\bar{\mathbf{X}}_w' \bar{\mathbf{M}}_w = \mathbf{0}$, we have

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

and using (43), it follows that

$$\text{(B.4)} \qquad \mathbf{q}_{NT} = \frac{1}{\sqrt{N}} \sum_{i=1}^N \tilde{w}_i \left( \frac{\mathbf{X}_i' \bar{\mathbf{M}}_w \mathbf{F}}{T} \right) \boldsymbol{\eta}_i = \frac{1}{\sqrt{N}} \sum_{i=1}^N \tilde{w}_i \left( \frac{\mathbf{X}_i' \bar{\mathbf{M}}_q \mathbf{F}}{T} \right) \boldsymbol{\eta}_i + O_p\left( \frac{1}{\sqrt{N}} \right) + O_p\left( \frac{1}{\sqrt{T}} \right),$$

where $\tilde{w}_i = w_i / (N^{-1} \sum_{i=1}^N w_i^2)^{1/2}$. Substituting this result in (B.1) and making use of (43) and (44), we have

$$\text{(B.5)} \qquad \left( \sum_{i=1}^N w_i^2 \right)^{-1/2} (\hat{\mathbf{b}}_P - \boldsymbol{\beta}) = \left( \sum_{i=1}^N w_i \frac{\mathbf{X}_i' \bar{\mathbf{M}}_q \mathbf{X}_i}{T} \right)^{-1} \left[ \frac{1}{\sqrt{N}} \sum_{i=1}^N \tilde{w}_i \frac{\mathbf{X}_i' \bar{\mathbf{M}}_q (\mathbf{X}_i \boldsymbol{v}_i + \boldsymbol{\varepsilon}_i + \mathbf{F} \boldsymbol{\eta}_i)}{T} \right] + O_p\left( \frac{1}{\sqrt{N}} \right) + O_p\left( \frac{1}{\sqrt{T}} \right).$$

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

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

where

$$\text{(B.6)} \qquad \mathbf{\Sigma}_P^* = \mathbf{\Psi}^{*-1}\mathbf{R}^*\mathbf{\Psi}^{*-1},$$

$$\text{(B.7)} \qquad \mathbf{\Psi}^* = \lim_{N\to\infty}\left(\sum_{i=1}^{N} w_i \mathbf{\Sigma}_{iq}\right),$$

$$\mathbf{R}^* = \lim_{N\to\infty}\left[N^{-1}\sum_{i=1}^{N}\tilde{w}_i^2(\mathbf{\Sigma}_{iq}\mathbf{\Omega}_{\boldsymbol{v}}\mathbf{\Sigma}_{iq} + \mathbf{Q}_{if}\mathbf{\Omega}_{\boldsymbol{\eta}}\mathbf{Q}'_{if})\right],$$

and $\mathbf{\Sigma}_{iq}$ and $\mathbf{Q}_{if}$ are defined by (61). *Q.E.D.*

PROOF OF THEOREM 4—Pooled Homogeneous Slope: As in the proof of Theorem 3, we first note that under $\theta_i = w_i$,

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

Also because the rank condition (21) is satisfied, using (22) we have

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

where $\bar{\mathbf{C}}_w'(\bar{\mathbf{C}}_w\bar{\mathbf{C}}_w')^{-1}$ is bounded for all $N$. Hence (noting that here $\eta_i$ is a scalar)

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

However,

$$\text{(B.8)} \qquad \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\tilde{w}_i\eta_i(\mathbf{X}_i'\bar{\mathbf{M}}_w\bar{\mathbf{U}}_w)$$
$$= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\tilde{w}_i\eta_i(\mathbf{X}_i'\bar{\mathbf{U}}_w)$$
$$- \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\tilde{w}_i\eta_i\left(\frac{\mathbf{X}_i'\bar{\mathbf{H}}_w}{T}\right)\left(\frac{\bar{\mathbf{H}}_w'\bar{\mathbf{H}}_w}{T}\right)^{-1}\bar{\mathbf{H}}_w'\bar{\mathbf{U}}_w,$$

where $\bar{\mathbf{H}}_w$ is defined by (31). Writing the first term as

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

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

$$\text{(B.9)}\qquad \frac{1}{N}\sum_{i=1}^{N}\tilde{w}_i\eta_i\left(\frac{\sqrt{N}\mathbf{X}_i'\bar{\mathbf{U}}_w}{\sqrt{T}}\right)\xrightarrow{p}0,\quad \text{as}\quad N\to\infty\quad \text{for all } T/N\to 0,$$

because $\tilde{w}_i = O(1)$, and $\eta_i$ are identically and independently distributed and distributed independently of $\sqrt{N}\mathbf{X}_i'\bar{\mathbf{U}}_w/\sqrt{T}$, with the terms $\eta_i(\sqrt{N}\mathbf{X}_i'\bar{\mathbf{U}}_w/\sqrt{T})$ having finite second-order moments. Consider the second term of (B.8) 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 (35) and (36) yields

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

Therefore,

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

where $\sqrt{N/T}\mathbf{G}'\bar{\mathbf{U}}_w = O_p(1)$, and $\eta_i$ are identically and independently distributed 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.9) in (B.8) now yields

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

and

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

but because the rank condition (21) is satisfied, using (44) 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

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

which establishes the validity of (71). *Q.E.D.*

## REFERENCES

AHN, S. G., Y.-H. LEE, AND P. SCHMIDT (2001): "GMM Estimation of Linear Panel Data Models with Time-Varying Individual Effects," *Journal of Econometrics*, 102, 219–255. [968,972]

ANSELIN, L. (2001): "Spatial Econometrics," in *A Companion to Theoretical Econometrics*, ed. by B. Baltagi. Oxford: Blackwell, 310–330. [967]

BAI, J. (2003): "Inferential Theory for Factor Models of Large Dimensions," *Econometrica*, 71, 135–171. [968]

BAI, J., AND S. NG (2002): "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70, 191–221. [968,972,1000]

BINDER, M., C. HSIAO, AND M. H. PESARAN (2005): "Estimation and Inference in Short Panel Vector Autoregressions with Unit Roots and Cointegration," *Econometric Theory*, 21, 795–837. [1001]

COAKLEY, J., A. FUERTES, AND R. P. SMITH (2002): "A Principal Components Approach to Cross-Section Dependence in Panels," Unpublished Manuscript, Birkbeck College, University of London.[968,975]

——— (2006): "Unobserved Heterogeneity in Panel Time Series Models," *Computational Statistics and Data Analysis*, 50, 2361–2380.[1001]

CONLEY, T. G., AND B. DUPOR (2003): "A Spatial Analysis of Sectoral Complementarity," *Journal of Political Economy*, 111, 311–352.[967]

CONLEY, T. G., AND G. TOPA (2002): "Socio-Economic Distance and Spatial Patterns in Unemployment," *Journal of Applied Econometrics*, 17, 303–327.[967]

DEES, S., F. DI MAURO, M. H. PESARAN, AND L. V. SMITH (2005): "Exploring the International Linkages of the Euro Area: A Global VAR Analysis," CESifo Working Paper 1425; *Journal of Applied Econometrics*, forthcoming.[967]

FORNI, M., AND M. LIPPI (1997): *Aggregation and the Microfoundations of Dynamic Macroeconomics*. Oxford: Clarendon Press.[968]

FORNI, M., AND L. REICHLIN (1998): "Let's Get Real: A Factor Analytical Approach to Disaggregated Business Cycle Dynamics," *Review of Economic Studies*, 65, 453–473.[968]

HOLTZ-EAKIN, D., W. K. NEWEY, AND H. ROSEN (1988): "Estimating Vector Autoregressions with Panel Data," *Econometrica*, 56, 1371–1395.[968]

HSIAO, C., M. H. PESARAN, AND A. K. TAHMISCIOGLU (1999): "Bayes Estimation of Short-Run Coefficients in Dynamic Panel Data Models," in *Analysis of Panels and Limited Dependent Variables: A Volume in Honour of G. S. Maddala*, ed. by C. Hsiao, K. Lahiri, L.-F. Lee, and M. H. Pesaran. Cambridge, U.K.: Cambridge University Press, 268–296. [982]

KAPETANIOS, G., AND M. H. PESARAN (2006): "Alternative Approaches to Estimation and Inference in Large Multifactor Panels: Small Sample Results with an Application to Modelling of Asset Returns," CESifo Working Paper 1416; in *The Refinement of Econometric Estimation and Test Procedures: Finite Sample and Asymptotic Analysis*, ed. by G. Phillips and E. Tzavalis. Cambridge, U.K.: Cambridge University Press, forthcoming.[998,1001]

KAPETANIOS, G., M. H. PESARAN, AND T. YAMAGATA (2006): "Analysis of Panel Data Models with Unit Roots and a Multifactor Error Structure," Unpublished Manuscript, Faculty of Economics, Cambridge University.[975]

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.[968,972]

LEE, Y. H. (1991): "Panel Data Models with Multiplicative Individual and Time Effects: Application to Compensation and Frontier Production Functions," Unpublished Ph.D. Dissertation, Michigan State University.[968,972]

LEE, K. C., AND M. H. PESARAN (1993): "The Role of Sectoral Interactions in Wage Determination in the UK Economy," *The Economic Journal*, 103, 21–55.[967]

NEWEY, W. K., AND K. D. WEST (1987): "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708. [981]

PESARAN, M. H. (2004a): "General Diagnostic Tests for Cross Section Dependence in Panels," CESifo Working Paper 1229; IZA Discussion Paper 1240.[970]

——— (2004b): "Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure," CESifo Working Paper 1331.[975]

——— (2005): "A Simple Panel Unit Root Test in the Presence of Cross Section Dependence," Revised Version of Cambridge Working Papers in Economics 0346.[1001]

PESARAN, M. H., T. SCHUERMANN, AND S. M. WEINER (2004): "Modeling Regional Interdependencies Using a Global Error-Correcting Macroeconometric Model" (with Discussions and Rejoinder), *Journal of Business Economics & Statistics*, 22, 129–181. [967,975]

PESARAN, M. H., AND R. P. SMITH (1995): "Estimating Long-Run Relationships from Dynamic Heterogeneous Panels," *Journal of Econometrics*, 68, 79–113. [982]

PHILLIPS, P. C. B., AND D. SUL (2003): "Dynamic Panel Estimation and Homogeneity Testing under Cross Section Dependence," *The Econometrics Journal*, 6, 217–259. [968]

ROBERTSON, D., AND J. SYMONS (2000): "Factor Residuals in SUR Regressions: Estimating Panels Allowing for Cross Sectional Correlation," Unpublished Manuscript, Faculty of Economics, University of Cambridge.[968]

STOCK, J., AND M. W. WATSON (2002): "Macroeconomic Forecasting Using Diffusion Indexes," *Journal of Business & Economic Statistics*, 20, 147–162. [968,972]

SWAMY, P. A. V. B. (1970): "Efficient Inference in Random Coefficient Regression Models," *Econometrica*, 38, 311–323. [969,972,982]