Econometrica, Vol. 67, No. 5 (September, 1999), 1057–1111

LINEAR REGRESSION LIMIT THEORY FOR NONSTATIONARY PANEL DATA[1]

BY PETER C. B. PHILLIPS AND HYUNGSIK R. MOON[2]

This paper develops a regression limit theory for nonstationary panel data with large numbers of cross section (n) and time series (T) observations. The limit theory allows for both sequential limits, wherein $T \to \infty$ followed by $n \to \infty$, and joint limits where $T, n \to \infty$ simultaneously; and the relationship between these multidimensional limits is explored. The panel structures considered allow for no time series cointegration, heterogeneous cointegration, homogeneous cointegration, and near-homogeneous cointegration. The paper explores the existence of long-run average relations between integrated panel vectors when there is no individual time series cointegration and when there is heterogeneous cointegration. These relations are parameterized in terms of the matrix regression coefficient of the long-run average covariance matrix. In the case of homogeneous and near homogeneous cointegrating panels, a panel fully modified regression estimator is developed and studied. The limit theory enables us to test hypotheses about the long run average parameters both within and between subgroups of the full population.

KEYWORDS: Nonstationary panel data, long-run average relations, multidimensional limits, panel cointegration regression, panel spurious regression.

1. INTRODUCTION

THERE HAS BEEN MUCH RECENT EMPIRICAL econometric work on economic models that uses panel data for which the time series component is nonstationary. Testing growth convergence theories in macroeconomics and estimating long-run relations between international financial series such as relative prices and exchange rates, and spot and future exchange rates are a few examples. This work has been facilitated by the construction and availability of a number of important panel data sets covering different individuals, regions, and countries over a relatively long time period, a notable example being the Penn World table. For such cases a new nonstationary panel data limit theory which allows for large n and large T asymptotics is useful. Much past panel data research has focused on identifying and estimating effects from stationary panels with a large cross section data dimension (n) but with few time series (T) observations. In

[1]An earlier version of this paper, Phillips and Moon (1997a), hereafter PM^a, was presented at the inaugural meeting of the New Zealand Econometric Society Group in Auckland, February 1997, while Phillips was visiting the University of Auckland. Some of the results were also presented in a training course on panel cointegration given by the first author at the EMBA meetings in Palm Cove, Australia, August 1996. The present paper is a shortened version of Phillips and Moon (1997b), hereafter PM^b, to which readers will be frequently referred for a full development of the algebra and limit theory given here.

[2]The authors thank a co-editor and three referees for comments on the paper, and Donald Andrews, David Pollard, and Oliver Linton for helpful discussions. Phillips thanks the NSF for research support under Grant No. SBR 94-22922, and Moon gratefully acknowledges financial support from a C.A. Anderson Prize Fellowship. The paper was typed by the authors in Scientific Word 2.5.

such cases a large n, fixed T limit theory is natural and Chamberlain (1984), Hsiao (1986), Matyas and Sevestre (1992), and Baltagi (1995) review much of this research.

The purpose of the present contribution is to investigate regressions with nonstationary panel data for which the time series component is an integrated process and where both T and n are large. In such cases, panel regressions can behave very differently from time series regressions. It has long been recognized by econometricians that panel data can distinguish effects that time series or cross section data alone cannot identify, and nonstationary panels provide a further instance of this phenomenon.

Suppose that we have two $I(1)$ random vectors, say $Y_{i,t}$ and $X_{i,t}$. When there is no cointegrating relation between $Y_{i,t}$ and $X_{i,t}$, and a time series regression for given i is performed, then the regression coefficient is well known to have a nondegenerate limit distribution and the regression is characterized as spurious (Granger and Newbold (1974) and Phillips (1986)). Now suppose that there are panel observations of $Y_{i,t}$ and $X_{i,t}$ with large cross sectional and time series components. In this case, even if the noise in the time series regression is strong, the noise can often be characterized as independent across individuals. Hence, by pooling the cross section and time series observations, we may attenuate the strong effect of the residuals in the regression while retaining the strength of the signal ($X_{i,t}$). In such a case, we can expect a panel-pooled regression to provide a consistent estimate of some long-run regression coefficient.

The present paper is concerned with developing a limit theory that is helpful in understanding and interpreting regressions of this type. In particular, we show the existence of an interesting long-run relation between panel vectors like $Y_{i,t}$ and $X_{i,t}$ that have no individual time series cointegrating relation. The new relation is a long-run *average* relationship over the cross section and it is parameterized in terms of a matrix regression coefficient derived from the cross section long-run average covariance matrix.

The following sections consider four possible panel structures for $Y_{i,t}$ and $X_{i,t}$: (i) no cointegrating relation, (ii) a heterogeneous cointegrating relation, (iii) a homogeneous cointegrating relation, and (iv) a near-homogeneous relation. Our analysis shows that in all four cases the pooled estimator is consistent and has a normal limit distribution. In the no cointegration and heterogeneous cointegration cases, we also study a limiting cross section estimator and prove that it is consistent and has a normal limit distribution, but that it is less efficient than the pooled estimator. In addition, in the case of homogeneous cointegration and near-homogeneous cointegration, we can construct a consistent estimator for the long-run regression coefficient, which we call a pooled FM (fully modified) estimator. This estimator has a faster coefficient convergence rate than the simple cross section and time series estimators.

Since the beginning of the 1990's there has been some ongoing research on nonstationary panel data that connects to our work here. Quah (1994), Levin and Lin (1993), and recently Im et al. (1996) considered unit root time series regressions with nonstationary panel data and proposed test statistics for unit roots. In addition, Pedroni (1995) studied some properties of cointegration statistics in pooled time series panels, and Robertson and Symons (1992) studied the biases that are likely to arise in practice with both stationary and nonstationary panel data. More closely related to our work is Pesaran and Smith (1995), who examined the impact of nonstationary variables on cross section regression estimates. They showed that spurious correlation between two $I(1)$ variables does not arise in the case of cross section regression with a finite number of time series observations under conditions such as exogenous regressors and iid disturbances. Our paper extends that result to a very general setting and provides a limit theory as $T \to \infty$ and $n \to \infty$ in panel regressions. The long-run relation defined in Pesaran and Smith (1995) is an average of randomly different cointegrating coefficients and they suggested cross section regression with time averaged data for consistent estimation. By contrast, our long run relation is the regression relation associated with the long run average covariance matrix and it is this regression that is the natural limit of a pooled panel regression. Further, we show that both pooled panel regression and limiting cross section regression estimators are consistent for this long-run average relation.

The limit theory developed here allows for both sequential limits, wherein $T \to \infty$ followed by $n \to \infty$, and joint limits where $T, n \to \infty$ simultaneously. A detailed discussion of multi-index asymptotic theory is provided and some general theorems for laws of large numbers and central limit theory are given. Sequential limit theory is easy to derive and generally leads to quick results for a variety of model configurations. Under some strengthening of the conditions, the results obtained under sequential limits are shown to apply also when $T, n \to \infty$ simultaneously. For a limit distribution theory we need the rate condition $n/T \to 0$. The latter condition indicates that the limit theory here is most likely to be useful in practice when n is moderate and T is large. We can expect such data configurations in multi-country macreconomic data, for example, when we restrict attention to groups of countries like OECD nations or developing countries. The limit theory enables us to test hypotheses about the long-run average parameters both within and between such subgroups of the full population.

The paper is organized as follows. Section 2 introduces the basic model, lays out assumptions, and gives some preliminary results including a multidimensional Beveridge Nelson (BN) decomposition. Section 3 develops a framework for asymptotics for double indexed processes that is used in the paper for both sequential and joint limit theories. Section 4 assumes that there is no cointegration among the $I(1)$ variable across all the individuals and gives asymptotic theories for a pooled regression estimator and a limiting cross section estimator. Section 5 assumes that there exists a cointegrating relation in the $I(1)$ variables across all the individuals and derives limit theories for the pooled estimators in three cases—heterogenous, homogeneous, and near-homogeneous cointegration. Section 6 indicates some extensions of this theory to allow for models with individual specific effects. Section 7 concludes the paper. Five appendices are included to develop the multidimensional limit theory, provide some technical background, relevant lemmas, and proofs of results in the paper.

Notation is fairly standard. The symbol "$\Rightarrow$" signifies weak convergence, "$:=$" is definitional equivalence, "$\equiv$" signifies equivalence in distribution, "$\xrightarrow{p}$" is convergence in probability, and "$\xrightarrow{a.s.}$" is convergence almost surely. The inequality ">0" signifies positive definiteness when applied to matrices. Stochastic processes such as Brownian motion $W(r)$ on $[0,1]$ are usually written as W, integrals such as $\int_0^1 W(r)\,dr$ as $\int W$, and stochastic integrals like $\int_0^1 W(r)\,dW(r)$ as $\int W\,dW$. Also $\mathrm{vec}(A)$ denotes vectorization of the matrix A by stacking columns, and $\|A\|$ is the Euclidean norm $(\mathrm{tr}(A'A))^{1/2}$.

2. ASSUMPTIONS, LARGE T ASYMPTOTICS, AND THE LONG-RUN AVERAGE COVARIANCE MATRIX

We start with a panel data model based on the vector integrated process

$$Z_{i,t} = Z_{i,t-1} + U_{i,t} \qquad (t=1,\dots,T;\ i=1,\dots,n) \tag{2.1}$$

with common initialization at $t=0$ satisfying

$$Z_{i,0} \text{ is iid across } i \text{ with } E\|Z_{i,0}\|^4 < \infty. \tag{2.2}$$

We partition the m-vectors $Z_{i,t}$ and $U_{i,t}$ in (2.1) into m_y and m_x components $(m = m_y + m_x)$ as $Z_{i,t} = (Y'_{i,t}, X'_{i,t})'$ and $U_{i,t} = (U'_{y_i,t}, U'_{x_i,t})'$. Condition (2.2) is made for convenience and could be generalized to allow for remote past initialization at the cost of some further complications (e.g., Phillips and Lee (1996)). The error $U_{i,t}$ is assumed to be generated by the random coefficient linear process

$$U_{i,t} = \sum_{s=0}^{\infty} C_{i,s} V_{i,t-s}, \tag{2.3}$$

where: (i) $\{C_{i,t}\}$ is a double sequence of $(m \times m)$ random matrices across i and over t; (ii) the m-vectors $V_{i,t}$ are iid across i and over t with $E(V_{i,t}) = 0$, $E(V_{i,t}V'_{i,t}) = I_m$, and, letting $V_{a,i,t}$ be the ath element of $V_{i,t}$, the $V_{a,i,t}$ are assumed to be independent across $a = 1,\dots,m$ with $E(V^4_{a,i,t}) = \upsilon^4$ for all i and t; (iii) $C_{i,s}$ and $V_{j,t}$ are independent for all i, j, t, and s.

We make two further assumptions about the random coefficients in (2.3). The first involves moment conditions and the second is a set of summability conditions on the moments of the random coefficients.

ASSUMPTION 1 (Random Coefficient Conditions):
(i) $\{C_{i,s}\}_i$ *is iid across i for all s.*
(ii) $E\|C_{i,s}\|^4 < \infty$ *for all s.*

Thus, $C_{i,s}$ is assumed to be iid across individuals and to have finite fourth moments that may vary over time. We allow $C_{i,s}$ to be dependent over s. This is important, because whenever $U_{i,t}$ is generated by a finite-parameter time series model like an autoregression, the coefficients in the Wold decomposition (2.3) will be nonlinear functions of these parameters that are lag (s) dependent and will therefore inevitably be dependent over s. Let $C_{a,i,s}$ be the ath element of $\text{vec}(C_{i,s})$. Also let $E(C_{a,i,s}^k) = \sigma_{k,a,s}$. Then we make the following assumption:

ASSUMPTION 2 (Summability Conditions): *The following hold for all* $a = 1, \ldots, m^2$:

(i) $\sum_{s=0}^{\infty} s^2 \sigma_{2,a,s} < \infty$.

(ii) $\sum_{s=0}^{\infty} s^4 (\sigma_{4,a,s})^{1/4} < \infty$.

Suppose the $U_{i,t}$ in (2.1) are generated by a random coefficient ARMA process whose characteristic equation has roots $\{\lambda_{ij}\colon\ j = 1, \ldots, J\}$. Then the coefficients $C_{i,s}$ in the Wold decomposition (2.3) are all linear combinations of powers of these characteristic roots. Under weak conditions on the distribution of the roots we can now verify Assumptions 1 and 2. Suppose, for instance, that the support of the distribution of the moduli of these roots is a compact set inside the stable region, so that $|\lambda_{ij}| \leq M_\lambda < 1$ a.s. Then all moments of $\|C_{i,s}\|$ are finite for all s, and series such as those in Assumption 2 are easily seen to be majorized by convergent series. For example, $\sum_{s=0}^{\infty} s^2 \sigma_{2,a,s} \leq M \sum_{s=0}^{\infty} s^2 M_\lambda^{2s} < \infty$ for some constant M. Similar conditions will ensure the validity of the alternative Assumptions 4 and 5 that are used later on in the paper.

The following lemma establishes the integrability of terms that appear frequently in our development.

LEMMA 1: *Let* $C_i(1) = \sum_{s=0}^{\infty} C_{i,s}$, $\tilde{C}_{i,s} = \sum_{t=s+1}^{\infty} C_{i,t}$, *and* $\tilde{U}_{i,t} = \sum_{s=0}^{\infty} \tilde{C}_{i,s} V_{i,t-s}$. *Under Assumptions* 1 *and* 2, *the following hold*:

(a) $E[\sum_{s=0}^{\infty} s^2 \|C_{i,s}\|^2] < \infty$.

(b) $E\|U_{i,t}\|^2 < M$ *for some* $M < \infty$.

(c) $E\|\tilde{U}_{i,t}\|^4 < M$ *for some* $M < \infty$.

(d) $E\|C_i(1)\|^4 < \infty$.

The temporal shift operator generating the iid sequence $\{V_{i,t}\}_t$ defines a measure preserving map on the product probability space induced by the independent sequences $\{V_{i,t}\}_t$ and $\{C_{i,t}\}_t$ and it generates the sequence $\{U_{i,t}\}_t$. Also the random coefficient sequence $\{C_{i,t}\}_t$ is square summable a.s. since $\sum_{t=0}^{\infty} \|C_{i,t}\|^2 < \infty$ a.s. by Lemma 1(a). Hence, the time series sequence $\{U_{i,t}\}_t$ is square integrable (Lemma 1(b)) and strictly stationary for all i. However, the sequence $\{U_{i,t}\}_t$ is not ergodic. This is because $F_{c_i} = \sigma(C_{i,0}, \ldots, C_{i,t}, \ldots)$, the sigma field generated by the sequence $\{C_{i,t}\}_{t=0}^{\infty}$, is an invariant sigma field with respect to the temporal shift operator and generates events with probability between zero and unity.

The following lemma shows that, for each i, $U_{i,t}$ satisfies a time series BN decomposition (see Phillips and Solo (1992)) almost surely.

LEMMA 2 (Panel BN decomposition): *Under Assumptions* 1 *and* 2 *the processes* $U_{i,t} = \sum_{s=0}^{\infty} C_{i,s} V_{i,t-s}$ *in* (2.3) *admit the following BN decomposition*:

$$U_{i,t} = C_i(1)V_{i,t} + \tilde{U}_{i,t-1} - \tilde{U}_{i,t} \qquad a.s. \tag{2.4}$$

Note that $C_i(1) = \sum_{s=0}^{\infty} C_{i,s} < \infty$ a.s. in view of Lemma 1(d), and $\tilde{U}_{i,t}$ are well defined square integrable random vectors by Lemma 1(c). Following Phillips and Solo (1992), partial sums of $U_{i,t}$ can be written as

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} U_{i,t} \overset{\text{a.s.}}{=} C_i(1) \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} V_{i,t} + \frac{1}{\sqrt{T}} \tilde{U}_{i,0} - \frac{1}{\sqrt{T}} \tilde{U}_{i,[Tr]}, \tag{2.5}$$

where $[Tr]$ denotes the integer part of Tr and it is a simple matter to establish that these partial sum processes satisfy functional laws. Indeed, we have the following large T result.

LEMMA 3 (Panel Functional CLT): *Under Assumptions* 1 *and* 2,

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} U_{i,t} \Rightarrow C_i(1) W_i(r) \qquad \text{as } T \to \infty \text{ for all } i.$$

Let $M_i(r) = (M_{y_i}(r)', M_{x_i}(r)')' = C_i(1)W_i(r) = (C_{y_i}(1)', C_{x_i}(1)')' W_i(r)$. Here, $M_i(r)$ is a randomly scaled (or mixed) Brownian Motion with conditional covariance matrix $C_i(1)C_i(1)'$, whose expectation is well defined because $\|EC_i(1)C_i(1)'\| < \infty$ in view of Lemma 1(d).

By the continuous mapping theorem and initial condition (2.2) we have

$$\frac{1}{T^2} \sum_{t=1}^{T} Z_{i,t} Z_{i,t}' \Rightarrow C_i(1) \int W_i W_i' C_i(1)' = \int M_i M_i',$$

as $T \to \infty$ for all i. Then, averaging over $i = 1, \ldots, n$, we have

$$\frac{1}{n} \sum_{i=1}^{n} \frac{1}{T^2} \sum_{t=1}^{T} Z_{i,t} Z_{i,t}' \Rightarrow \frac{1}{n} \sum_{i=1}^{n} C_i(1) \int W_i W_i' C_i(1)' = \frac{1}{n} \sum_{i=1}^{n} \int M_i M_i', \tag{2.6}$$

as $T \to \infty$ for any fixed n. Integrability of the summands in (2.6) follows readily under the given summability and moment conditions.

LEMMA 4: *Under Assumptions* 1 *and* 2, $E\|\int M_i M_i'\|^2 < \infty$.

In consequence, a strong law of large numbers applies to (2.6) as $n \to \infty$ and so

$$\frac{1}{n} \sum_{i=1}^{n} \int M_i M_i' \overset{\text{a.s.}}{\to} E\left(\int M_i M_i' \right). \tag{2.7}$$

The limit here depends not on the covariance matrix of $Z_{i,t}$, but on a parameter matrix that measures the long-run (over t) covariance of $Z_{i,t}$ averaged over i. This parameter matrix is constructed as follows.

Let Ω_i be the long-run conditional covariance matrix of $Z_{i,t} = (Y_{i,t}, X_{i,t})'$ conditioned on F_{c_i}, i.e.,

$$\Omega_i = \begin{pmatrix} \Omega_{y_i y_i} & \Omega_{y_i x_i} \\ \Omega_{x_i y_i} & \Omega_{x_i x_i} \end{pmatrix} = C_i(1)C_i(1)' = \begin{pmatrix} C_{y_i}(1)C_{y_i}(1)' & C_{y_i}(1)C_{x_i}(1)' \\ C_{x_i}(1)C_{y_i}(1)' & C_{x_i}(1)C_{x_i}(1)' \end{pmatrix},$$

where the partitions of Ω_i and $C_i(1)C_i(1)'$ are conformable. By Lemma 1(d), Ω_i is integrable and we denote

$$\Omega = \begin{pmatrix} \Omega_{yy} & \Omega_{yx} \\ \Omega_{xy} & \Omega_{xx} \end{pmatrix} = E(\Omega_i) = E(C_i(1)C_i(1)').$$

We call Ω the *long-run average covariance* matrix of $Z_{i,t}$.

It is now apparent from (2.6) that $E(\int M_i M_i') = E[C_i(1)E(\int W_i W_i')C_i(1)']$. A simple calculation reveals that $E(\int W_i W_i') = \frac{1}{2}I_m$, so that (2.7) becomes

$$\text{(2.8)} \qquad \frac{1}{n}\sum_{i=1}^{n}\int M_i M_i' \overset{\text{a.s.}}{\to} \frac{1}{2}\Omega,$$

showing that the limit of the second moment matrix of the data depends on the long-run average covariance matrix Ω. Taken together, (2.6) and (2.7) give us an instance of an asymptotic development in which $T \to \infty$, followed by $n \to \infty$, leading to the sequence of limits

$$\text{(2.9)} \qquad X_{n,T} := \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{T^2}\sum_{i=1}^{T} Z_{i,t}, Z_{i,t}'\right) \Rightarrow \frac{1}{n}\sum_{i=1}^{n}\int M_i M_i' \overset{\text{a.s.}}{\to} \frac{1}{2}\Omega$$

for the double indexed process $X_{n,T}$. The next section discusses this type of sequential asymptotic theory in relation to more general joint asymptotics.

3. LIMIT THEORY FOR MULTIDIMENSIONAL PROCESSES

Throughout this paper, attention is focused on deriving the limit behavior of a double indexed process $X_{n,T}$, such as that given in (2.9). In general, the limit of $X_{n,T}$ depends on the treatment of the two indices, n and T, and the properties that link the rows and columns of the process. Several approaches are possible. One approach is to fix one of the indexes, say n, and allow the other (T) to pass to infinity, giving an intermediate limit. By letting n pass to infinity subsequently, a *sequential* limit theory is obtained. We write this type of limit process in the form $(T, n \to \infty)_{\text{seq}}$, where the order of the indices is critical to the meaning. While they often lead to tractable deviations, sequential limits can give asymptotic results that are misleading in cases where both indexes pass to infinity simultaneously. A second approach is to pass to infinity along a specific diagonal path (in the two dimensional array) determined by a monotonically increasing function relation of the type $T = T(n)$ while the index $n \to \infty$. (This approach really requires specificity about the functional dependence only in the limit, so it includes cases such as those where we assume that $(T/n) \to c \neq 0$, in which case we simply have $T(n) = cn$.) We write this type of limit process in the form $(T(n), n \to \infty)_{\text{diag}}$. This approach also simplifies the asymptotic theory by replacing $X_{n,T}$ with the single indexed process $X_{n,T(n)}$. The drawback of *diagonal path* limit theory is that the assumed expansion path $(T(n), n) \to \infty$ is highly specific and may not provide an appropriate approximation for a given (T,n) situation. Moreover, the limit theory can depend on the specific functional relation $T = T(n)$ that is used in the asymptotic development. (A recent econometric example of this situation is analyzed in Phillips and Lee (1996).) A third approach is to allow both indexes to pass to infinity simultaneously without placing specific diagonal path restrictions on the divergence. We write this type of limit process in the form $(T, n \to \infty)$. Generally speaking, such *joint* limit theory requires stronger conditions (linking the rows and columns of the joint array, and on the moments of the component variates) to establish than sequential convergence or diagonal path convergence. But, by the same token, the results are also stronger and may be expected to be relevant to a wider range of circumstances, provided the conditions hold.

The asymptotic development in this paper will involve both sequential limit theory and joint limit theory arguments. The sequential limits are especially helpful in extracting quick asymptotics and they are useful because they bring into play all of the key elements in our final limit theory in a straightforward way. The joint limit theory is more difficult to derive and applies under stronger conditions. Fortunately, these conditions do not seem to exclude cases of major importance for the type of large T and moderate n empirical applications that we have in mind for our methods.

The following subsections define the convergence concepts that we need and give some conditions that assure joint convergence.

3.1. *Definitions and Some Relations between Sequential and Joint Limits*

A typical double index process of the type that occurs in this paper has the linear form

$$X_{n,T} = \frac{1}{k_n} \sum_{i=1}^{n} Y_{i,T}, \tag{3.1}$$

where $Y_{i,T}$ are independent m-component random vectors across i defined on a probability space (Ω, F, P). A typical $Y_{i,T}$ component in our case has the form of a standardized sum

$$(3.2) \qquad Y_{i,T} = \frac{1}{d_T} \sum_{t=1}^{T} f(Z_{i,t=[Tr]}),$$

where the $Z_{i,[Tr]}$ are random elements in the space[3] $D[0,1]^h$, for some integer h, within the space (Ω, F, P), d_T is a standardizing factor, and f is a continuous functional from $D[0,1]^h$ to $\mathbb{R}^m$. We alert the reader that the meaning of the notation $X_{n,T}$ and $Y_{i,T}$ in (3.1) and (3.2) above is different from that of the symbols $Y_{i,t}$ and $X_{i,t}$ which represent components of $Z_{i,t}$ given in (2.1) that appear in other sections of this paper. The differences in meaning should be obvious from the context.

DEFINITION 1: (a) A sequence of m-vectors $\{X_{n,T}\}$ on (Ω, F, P) is said to *converge in probability to X sequentially*, written $X_{n,T} \to_p X$ in sequential limit as $(T, n \to \infty)_{\text{seq}}$, if

$$\lim_{n\to\infty} \lim_{T\to\infty} P\{\| X_{n,T} - X \| > \varepsilon\} = 0 \qquad \forall \varepsilon > 0.$$

(b) $X_{n,T}$ converges in distribution sequentially to the m-vector X, written $X_{n,T} \Rightarrow X$ in sequential limit as $(T, n \to \infty)_{\text{seq}}$, if

$$\lim_{n\to\infty} \lim_{T\to\infty} | Ef(X_{n,T}) - Ef(X) | = 0 \qquad \forall f \in C,$$

where C is the class of all bounded, continuous, real functions on $\mathbb{R}^m$.

In practice, we can find the sequential limits of $X_{n,T}$ in (3.1) as follows. Using time series limit theory we find the limit behavior of $Y_{i,T}$. Suppose, for example, that as $T \to \infty$

$$(3.3) \qquad Y_{i,T} \Rightarrow Y_i$$

or

$$(3.4) \qquad Y_{i,T} \overset{p}{\to} Y_i \qquad \text{for all } i.$$

Then, by the independence of $Y_{i,T}$ across i for all T, we have $X_{n,T} \Rightarrow X_n$ or $X_{n,T} \to_p X_n$ as $T \to \infty$ for all n, where $X_n = (1/k_n)\sum_{i=1}^{n} Y_i$.

By enlarging the underlying probability space if necessary, we can take it in the case of (3.3) that all the Y_i's are defined on the same probability space.

[3] $D[0,1]^h$ is a product metric space of h independent copies of $D[0,1]$, the space of the all real valued functions on the interval $[0,1]$ that are right continuous and have finite left limits. $D[0,1]^h$ is endowed with the metric $\rho_h(f, g) = \max_i\{\rho(f_i, g_i)\colon i \in \{1, \ldots, h\}, f_i, g_i \in D[0,1]\}$, where ρ is the modified Skorohod metric (see Billingsley (1968, p. 112)) under which $D[0,1]$ is separable and complete.

Hence, the sum of the limit random variables $\sum_{i=1}^{n} Y_i$ is well defined on the same space. Next, we allow $n \to \infty$ and apply a limit theory to the standardized sum

$$X_n = \frac{1}{k_n} \sum_{i=1}^{n} Y_i. \tag{3.5}$$

Under some regularity conditions, we can now find the sequential limit, X, or X_n. For example, if $k_n = n$, we can apply a law of large numbers (LLN) to X_n and if $k_n = \sqrt{n}$ we can use an appropriate central limit theorem (CLT).

The requirement that the Y_i's in (3.3) are defined on the same probability space is important, especially when we apply an LLN to X_n in the second stage (3.5). The reason is as follows. The weak convergence $X_{n,T} \Rightarrow X_n$ as $T \to \infty$ involves only the implication that the distribution of the $X_{n,T}$ converges to the distribution of the X_n, not any properties of the probability space where the X_n are defined. Indeed, if the weak convergence is mixing (e.g., see Hall and Heyde (1980)), then X_n escapes from the underlying probability space when $T \to \infty$. However, to employ an LLN to the sequence of X_n, these variates need to be defined on the same probability space. The requirement that the Y_i's in (3.3) are defined on the same probability space can be accommodated by suitably enlarging the underlying space. The construction of such a probability space is provided in Appendix B.

Next, we define the concepts of joint convergence in probability and joint weak convergence.

DEFINITION 2: (a) Suppose that the m-vector random sequence $X_{n,T}$ and X are defined on a probability space (Ω, F, P). $X_{n,T}$ is said to *converge in probability jointly to* X, written $X_{n,T} \to_p X$ as $(T, n \to \infty)$, if

$$\lim_{T, n \to \infty} P\{\| X_{n,T} - X \| > \varepsilon\} = 0 \qquad \forall \varepsilon > 0. \tag{3.6}$$

(b) $X_{n,T}$ is said to converge in distribution jointly to a $(m \times 1)$ random vector X, written $X_{n,T} \Rightarrow X$ as $(T, n \to \infty)$, if

$$\lim_{T, n \to \infty} | Ef(X_{n,T}) - Ef(X) | = 0 \qquad \forall f \in C, \tag{3.7}$$

where C is the class of all bounded, continuous real functions on $\mathbb{R}^m$.

REMARKS: (a) Evidently, joint convergence implies diagonal convergence on all monotonic diagonal paths. Moreover, a version of the converse is also true, namely that $X_{n,T} \to_p X$ (or $X_{n,T} \Rightarrow X$) as $(T, n \to \infty)$ if $X_{n,T(n)} \to_p X$ as $(T(n), n \to \infty)_{\text{diag}}$ for all $T(n) \to \infty$ monotonically as $n \to \infty$.

(b) In some of our results, we need to place a condition on the indexes in joint convergence of the form $n/T \to 0$. Joint convergence as $(T, n \to \infty)$ is then said to apply subject to this condition. The definitional limits given in (3.6) and (3.7) above are naturally subject to the same condition regarding the passage of the indexes to infinity in this case.

Sequential limits are by no means always the same as joint limits or diagonal path limits. Sometimes, different normalizations even are required to obtain nondegenerate limits. PM[b] gives several examples. Nevertheless, under some circumstances we can establish a relationship between sequential limits and joint limits. The following two lemmas give some elementary conditions.

LEMMA 5 (Conditions for Joint Convergence to Imply Sequential Convergence): (a) *Suppose there exist random vectors* X_n *on the same probability space as* $X_{n,T}$ *satisfying, for all* n, $X_{n,T} \to_p X_n$ *as* $T \to \infty$. *If* $X_{n,T} \to_p X$ *as* $(T, n \to \infty)$, *then* $X_{n,T} \to_p X$ *sequentially as* $(T, n \to \infty)_{\text{seq}}$.

(b) *Suppose there exist random vectors* X_n *such that, for any fixed* n, $X_{n,T} \Rightarrow X_n$ *as* $T \to \infty$. *If* $X_{n,T} \Rightarrow X$ *as* $n, T \to \infty$, *then* $X_{n,T} \Rightarrow X$ *sequentially as* $(T, n \to \infty)_{\text{seq}}$.

LEMMA 6 (Conditions for Sequential Convergence to Imply Joint Convergence): (a) *Suppose there exist random vectors* X_n *and* X *on the same probability space as* $X_{n,T}$ *satisfying, for all* n, $X_{n,T} \to_p X_n$ *as* $T \to \infty$ *and* $X_n \to_p X$ *as* $n \to \infty$. *Then,* $X_{n,T} \to_p X$ *as* $(n, T \to \infty)$ *if and only if,*

$$(3.8) \qquad \limsup_{n,T} P\{\| X_{n,T} - X_n \| > \varepsilon\} = 0 \qquad \forall \varepsilon > 0.$$

(b) *Suppose there exist random vectors* X_n *such that, for any fixed* n, $X_{n,T} \Rightarrow X_n$ *as* $T \to \infty$ *and* $X_n \Rightarrow X$ *as* $n \to \infty$. *Then,* $X_{n,T} \Rightarrow X$ *as* $(n, T \to \infty)$ *if and only if,*

$$(3.9) \qquad \limsup_{n,T} | E(f(X_{n,T})) - E(f(X_n)) | = 0 \qquad \forall f \in C.$$

3.2. *Joint Convergence in Probability*

Consider a double indexed process $X_{n,T}$ whose typical form is an average of $(m \times 1)$ random vectors $Y_{i,T}$,

$$(3.10) \qquad X_{n,T} = \frac{1}{n} \sum_{i=1}^{n} Y_{i,T},$$

where the $Y_{i,T}$ are independent across i for all T. The concern is to establish conditions under which a probability limit of $X_{n,T}$ in (3.10) exists and to develop methods of finding this probability limit.

Suppose the $X_{n,T}$ are integrable and let

$$(3.11) \qquad \mu_X = \lim_{n,T \to \infty} EX_{n,T} = \lim_{n,T \to \infty} \frac{1}{n} \sum_{i=1}^{n} EY_{i,T} \quad \text{be finite}.$$

By definition it is sufficient for $X_{n,T} \to_p \mu_X$ as $(n, T \to \infty)$ to show that

$$(3.12) \qquad \lim_{n,T \to \infty} P\left\{ \left\| \frac{1}{n} \sum_{i=1}^{n} (Y_{i,T} - EY_{i,T}) \right\| > \varepsilon \right\} = 0 \qquad \text{for all } \varepsilon > 0.$$

In some applications (3.12) can be verified by showing that

$$\lim_{n,T\to\infty} E\left\| \frac{1}{n} \sum_{i=1}^{n} (Y_{i,T} - EY_{i,T}) \right\| = 0 \tag{3.13}$$

using the Markov inequality. Or, if the $X_{n,T}$ are square integrable, (3.12) follows by Chebychev's inequality when

$$\lim_{n,T\to\infty} E\left\| \frac{1}{n} \sum_{i=1}^{n} (Y_{i,T} - EY_{i,T}) \right\|^2 = \lim_{n,T\to\infty} \frac{1}{n^2} \sum_{i=1}^{n} E\|Y_{i,T} - EY_{i,T}\|^2 = 0, \tag{3.14}$$

where the first equality holds because the $Y_{i,T}$ are independent across i for all T.

Sequential probability limits can also be derived. From time series limit theory we may obtain the limit behavior of $Y_{i,T}$ when $T \to \infty$. Suppose, for instance, that as $T \to \infty$

$$Y_{i,T} \Rightarrow Y_i \qquad \forall i \tag{3.15}$$

or

$$Y_{i,T} \xrightarrow{p} Y_i \qquad \forall i \tag{3.16}$$

so that, by the independence of $Y_{i,T}$ across i for all T, it follows $X_{n,T} \Rightarrow X_n$ or $X_{n,T} \to_p X_n$ as $T \to \infty$ for all n, where $X_n = (1/n)\sum_{i=1}^n Y_i$.

Suppose also, in the case of (3.15), that the Y_i are defined on the same probability space for all i so that the sum of the limit random variables $(1/n)\sum_{i=1}^n Y_i$ is meaningful. Appendix B(1) provides a construction for doing this and, hereafter, we assume that the random vectors Y_i in (3.15) exist on the same probability space whenever we use sequential limit arguments. By allowing $n \to \infty$ and applying a standard strong law for independent random variables to

$$X_n = \frac{1}{n} \sum_{i=1}^{n} Y_i, \tag{3.17}$$

under some regularity conditions,[4] we may find the sequential limit X. Let

$$\tilde{\mu}_X = \lim_{n} \frac{1}{n} \sum_{i=1}^{n} EY_i. \tag{3.18}$$

Then

$$X_n = \frac{1}{n} \sum_{i=1}^{n} Y_i \xrightarrow{\text{a.s.}} \tilde{\mu}_X = \lim_{n} \frac{1}{n} \sum_{i=1}^{n} EY_i.$$

A fundamental question is whether the joint probability limit μ_X in (3.11) is equivalent to the sequential probability limit $\tilde{\mu}_X$ in (3.18). Lemma 6 provides

[4] Simple sufficient conditions are that the Y_i are independent with $\sup_i E\|Y_i - EY_i\|^2 < \infty$.

one solution. According to Lemma 6, it is enough to verify condition (3.9) with $X_{n,T} = (1/n)\Sigma_{i=1}^{n} Y_{i,T}$ and $X_n = (1/n)\Sigma_{i=1}^{n}\Sigma_{i=1}^{n} Y_i$ to conclude that $X_{n,T} \to_p \tilde{\mu}_X$ as $n, T \to \infty$, where $\tilde{\mu}_X = \lim_n (1/n)\Sigma_{i=1}^{n} EY_i$. The following theorem gives a set of sufficient conditions under which condition (3.9) is satisfied, so that the probability limits μ_X in (3.11) and $\tilde{\mu}_X$ in (3.18) are equivalent.

THEOREM 1 (Joint Probability Limits): *Suppose the* $(m \times 1)$ *random vectors* $Y_{i,T}$ *are independent across i for all T and integrable. Assume that* $Y_{i,T} \Rightarrow Y_i$ *as* $T \to \infty$ *for all i. Let the following hold*:

(i) $\limsup_{n,T}(1/n)\Sigma_{i=1}^{n} E\|Y_{i,T}\| < \infty$;

(ii) $\limsup_{n,T}(1/n)\Sigma_{i=1}^{n} \|EY_{i,T} - EY_i\| = 0$;

(iii) $\limsup_{n,T}(1/n)\Sigma_{i=1}^{n} E\|Y_{i,T}\| 1\{\|Y_{i,T}\| > n\varepsilon\} = 0 \ \forall \varepsilon > 0$; *and*

(iv) $\limsup_{n}(1/n)\Sigma_{i=1}^{n} E\|Y_i\| 1\{\|Y_i\| > n\varepsilon\} = 0 \ \forall \varepsilon > 0$.

(a) *Then condition* (3.9) *holds*.

(b) *If* $\lim_{n\to\infty}(1/n)\Sigma_{i=1}^{n} EY_i (:= \tilde{\mu}_X)$ *exists and* $X_n \to_p \tilde{\mu}_X$ *as* $n \to \infty$, *then* $X_{n,T} \to_p \tilde{\mu}_X$ *as* $(n, T \to \infty)$.

In establishing the existence of a joint probability limit of $(1/n)\Sigma_{i=1}^{n} Y_{i,T}$, Theorem 1 requires only first moment assumptions on $Y_{i,T}$ and is, in this respect, less demanding than (3.14), which uses second moments of $Y_{i,T}$. Theorem 1 is particularly useful when the first moment condition (3.13) is not so easy to establish.

An important special case arises when the $Y_{i,T}$ are a scaled version of some iid random vectors $Q_{i,T}$.[5] Suppose that $Y_{i,T} = C_i Q_{i,T}$, where the $\{Q_{i,T}\}_i$ are iid for all T and the C_i are $(m \times m)$ nonrandom matrices for all i. Suppose that $Q_{i,T} \Rightarrow Q_i$ as $T \to \infty$ for all i, so that $Y_i = C_i Q_i$. In general, the $Y_{i,T}$ are heterogenous across i unless the C_i are the same for all i. The source of the heterogeneity of $Y_{i,T}$ is the scale effect C_i, and then the heterogeneity from C_i is smoothed by letting $n \to \infty$. We have the following result for this special case.

COROLLARY 1: *Suppose that* $Y_{i,T} = C_i Q_{i,T}$, *where the* $Q_{i,T}$ *are iid across i for all* T, *and the* C_i *are* $(m \times m)$ *nonrandom matrices for all i. Assume that the* $Q_{i,T}$ *are integrable for all T and* $Q_{i,T} \Rightarrow Q_i$ *as* $T \to \infty$. *Assume that* $C = \lim_n (1/n)\Sigma_{i=1}^{n} C_i$ *exists. If* $\|Q_{i,T}\|$ *is uniformly integrable in T for all i, and if* $\sup_i \|C_i\| < \infty$, *then* $(1/n)\Sigma_{i=1}^{n} Y_{i,T} \to_p CE(Q_i)$ *as* $(n, T \to \infty)$.

[5] In many applications, an $I(1)$ process $Z_{i,t}$ can be decomposed into a scaled random walk process plus an error term, that is, $Z_{i,t} = C_i(1)S_{i,t} + \tilde{U}_{i,0} - \tilde{U}_{i,t}$, where $S_{i,t} = S_{i,t-1} + U_{i,t}$ and $C_i(1)$ is the long-run moving average coefficient of $\Delta Z_{i,t}$ (see Phillips and Solo (1992)). Then, the scale factor C_i is $C_i(1)$ and $Q_{i,T}$ corresponds to $f(S_{i,t}/\sqrt{T})$, where f is a continuous functional on some metric space.

3.3. *Joint Central Limit Theory*

This section considers joint convergence in distribution of the $\sqrt{n}$-standardized double sequence $X_{n,T}$,

$$X_{n,T} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_{i,T}, \tag{3.19}$$

where the $Y_{i,T}$ are independent $(m \times 1)$ random vectors across i with $EY_{i,T} = 0$ and $EY_{i,T}Y'_{i,T} = \Omega_{i,T}$.

One approach to the limit distribution of $X_{n,T}$ is to attempt to use a multivariate CLT directly. This approach is particularly appropriate in the case of diagonal path limits where we have $X_{n,T(n)} = (1/\sqrt{n})\sum_{i=1}^{n} Y_{i,T(n)}$ and a suitable multivariate CLT for triangular arrays can be applied. This idea was employed by Quah (1994) and Levin and Lin (1993) in their work on panel unit root tests. But, in general, when n and T go to infinity and no specific expansion relation between n and T is assumed, we cannot use traditional CLT's in this way. In what follows, therefore, we develop a joint CLT for $(n, T \to \infty)$, using a Lindeberg condition for a double indexed process.

First, take the case where the $Y_{i,T}$ in (3.19) are scalar random variables. Let $s_{n,T}^2 = \sum_{i=1}^{n} \Omega_{i,T}$ and define $\xi_{i,n,T} = Y_{i,T}/s_{n,T}$. Then we have the following result.

THEOREM 2 (Joint Limit CLT): *Suppose that for* $\forall \varepsilon > 0$,

$$\lim_{n,T\to\infty} \sum_{i=1}^{n} E\left[\xi_{i,n,T}^2 1\{|\xi_{i,n,T}| > \varepsilon\}\right] = 0. \tag{3.20}$$

Then, as $(n, T \to \infty)$,

$$\sum_{i=1}^{n} \xi_{i,n,T} \Rightarrow N(0,1).$$

There are some interesting special cases of this joint CLT. The following result, which is related to a theorem of Eicker (1963), arises when the $(m \times 1)$ random vectors $Y_{i,T}$ are scaled versions of iid random vectors $Q_{i,T}$.

THEOREM 3 (Joint Limit CLT for Scaled Variates): *Suppose that* $Y_{i,T} = C_i Q_{i,T}$, *where the* $(m \times 1)$ *random vectors* $Q_{i,T}$ *are* iid$(0, \Sigma_T)$ *across i for all T and the* C_i *are* $(m \times m)$ *nonzero and nonrandom matrices. Assume the following conditions hold*:

(i) *Let* $\sigma_T^2 = \lambda_{\min}(\Sigma_T)$ *and* $\liminf_T \sigma_T^2 > 0$;
(ii) $\max_{i \le n} \|C_i\|^2 / \lambda_{\min}(\sum_{i=1}^{n} C_i C_i') = O(1/n)$ *as* $n \to \infty$;
(iii) $\|Q_{i,T}\|^2$ *are uniformly integrable in* T;
(iv) $\lim_{n,T} (1/n) \sum_{i=1}^{n} C_i \Sigma_T C_i' = \Omega > 0$.

Then,

$$X_{n,T} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_{i,T} \Rightarrow N(0, \Omega) \qquad \text{as } n, T \to \infty.$$

Sequential weak convergence of $(1/\sqrt{n})\Sigma_{i=1}^{n} Y_{i,T}$ can be derived in the same way as the sequential probability limit of $(1/n)\Sigma_{i=1}^{n} Y_{i,T}$ considered earlier. Suppose that, for each i as $T \to \infty$, the random variables $Y_{i,T}$ converge in distribution to Y_i, where the Y_i are independent with mean zero and variance Ω_i. Then, $(1/\sqrt{n})\Sigma_{i=1}^{n} Y_i \Rightarrow N(0, \Omega)$ if for all $\varepsilon > 0$ as $n \to \infty$

$$(3.21) \qquad E\frac{Y_i^2}{s_n^2} 1\left\{\left|\frac{Y_i^2}{s_n^2}\right| > \varepsilon\right\} \to 0,$$

and

$$(3.22) \qquad \frac{1}{n} s_n^2 = \frac{1}{n}\sum_{i=1}^{n} \Omega_i \to \Omega.$$

In many econometric applications Y_i is a Gaussian random variable or a function of the Gaussian process. So the Y_i usually possess higher moments. Second moment requirements then follow automatically, and the Lindeberg condition (3.21) for $(1/\sqrt{n})\Sigma_{i=1}^{n} Y_i$ may be verified directly using a Liapounov condition.

Additional Remarks

(a) Sequential weak convergence of $(1/\sqrt{n})\Sigma_{i=1}^{n} Y_{i,T}$ to $N(0, \Omega)$ under conditions (3.21) and (3.22) does not imply that $(1/\sqrt{n})\Sigma_{i=1}^{n} Y_{i,T}$ converges in distribution jointly to $N(0, \Omega)$ as $(n, T \to \infty)$. According to Lemma 6(ii), condition (3.9) is a necessary and sufficient condition for $(1/\sqrt{n})\Sigma_{i=1}^{n} Y_{i,T}$ to converge in distribution jointly to the sequential limit distribution $N(0, \Omega)$. In this case, therefore, condition (3.20) and the condition that $(1/n)\Sigma_{i=1}^{n} \Omega_{i,T} \to \Omega$ as $(n, T \to \infty)$ provide sufficient conditions for condition (3.9).

(b) When $Y_{i,T}$ in (3.19) does not have mean zero, but instead has mean zero asymptotically as $T \to \infty$ for each i, joint CLT's such as Theorem 2 or Corollary 3 cannot be applied. In this case, T needs to increase fast enough to make the $\sqrt{n}$-standardized sum of the biases small. That is, $(1/\sqrt{n})\Sigma_{i=1}^{n} EY_{i,T}$ should go to zero as $(n, T \to \infty)$. In this case, asymptotic normality of $(1/\sqrt{n})\Sigma_{i=1}^{n} Y_{i,T}$ will continue to hold provided the expansion rate between n and T allows the bias to go to zero. The next section gives an example where this problem arises (e.g., see Theorem 4).

4. SPURIOUS PANEL REGRESSION

This section considers the case where the two component random vectors $Y_{i,t}$ and $X_{i,t}$ of $Z_{i,t}$ in (2.1) have no cointegrating relation for any i. This case is covered by the following assumption.

ASSUMPTION 3 (Spurious Regression): *The random matrices Ω_i are positive definite almost surely.*

Suppose that we perform a time series regression of $Y_{i,t}$ on $X_{i,t}$:

$$Y_{i,t} = \hat{\beta}_i X_{i,t} + \hat{U}_{i,t}, \tag{4.1}$$

where $\hat{\beta}_i = \sum_{t=1}^{T} Y_{i,t} X'_{i,t} (\sum_{t=1}^{T} X_{i,t} X'_{i,t})^{-1}$. As is well known (e.g., Phillips (1986)), under Assumption 3 the regression coefficient estimator $\hat{\beta}_i$ has the following nondegenerate limit distribution, as $T \to \infty$:

$$\hat{\beta}_i \Rightarrow \int M_{y_i} M'_{x_i} \left(\int M_{x_i} M'_{x_i} \right)^{-1} \qquad \text{for all } i. \tag{4.2}$$

The weak convergence result (4.2) implies that regression (4.1) is spurious in the sense that the regression of $Y_{i,t}$ on $X_{i,t}$ does not identify any fixed long-run relation between $Y_{i,t}$ and $X_{i,t}$. By contrast, the main result in what follows is that, in a panel data set, such regressions are no longer spurious and do, in fact, distinguish a long-run average relation between $Y_{i,t}$ and $X_{i,t}$.

Consider the following linear least-squares regression of $Y_{i,t}$ on $X_{i,t}$ with pooled panel data:

$$Y_{i,t} = \hat{\beta}_{n,T} X_{i,t} + \hat{U}_{i,t}, \tag{4.3}$$

where

$$\hat{\beta}_{n,T} = \left(\sum_{i=1}^{n} \sum_{t=1}^{T} Y_{i,t} X'_{i,t} \right) \left(\sum_{i=1}^{n} \sum_{t=1}^{T} X_{i,t} X'_{i,t} \right)^{-1}. \tag{4.4}$$

We now proceed to develop an asymptotic theory for $\hat{\beta}_{n,T}$. The approach we adopt is to derive the limit under sequential convergence of the indices (T, n), and then show that the limit continues to hold under joint convergence $(T, n \to \infty)$ provided certain conditions hold. In many cases, this is the simplest way to proceed.

Indeed, for estimators like $\hat{\beta}_{n,T}$, asymptotic results are readily obtained using sequential asymptotics such as $(T, n \to \infty)_{\text{seq}}$. According to (2.6) in first stage asymptotics, the pooled estimator $\hat{\beta}_{n,T}$ has the following limit distribution:

$$\hat{\beta}_{n,T} \Rightarrow \left(\frac{1}{n} \sum_{i=1}^{n} \int M_{y_i} M'_{x_i} \right) \left(\frac{1}{n} \sum_{i=1}^{n} \int M_{x_i} M'_{x_i} \right)^{-1}$$

as $T \to \infty$ for any fixed n. From Lemma 4 we know that $\int M_{y_i} M'_{x_i}$ and $\int M_{x_i} M'_{x_i}$ have finite second moments. Also, as in (2.8) above, by direct calculation we get $E(\int M_i M'_i) = \frac{1}{2} E(\Omega_i) = \frac{1}{2}\Omega$. And then, applying the strong law of large numbers as in (2.7), we have

$$\frac{1}{n} \sum_{i=1}^{n} \int M_{y_i} M'_{x_i} \overset{\text{a.s.}}{\to} \frac{1}{2} \Omega_{yx} \qquad \text{and} \qquad \frac{1}{n} \sum_{i=1}^{n} \int M_{x_i} M'_{x_i} \overset{\text{a.s.}}{\to} \frac{1}{2} \Omega_{xx},$$

as $n \to \infty$. By Assumption 3, $\Omega_{x_i x_i}$ is positive definite a.s., and $c'\Omega_{x_i x_i} c > 0$ a.s. for any $c \neq 0$ in $\mathbb{R}^{m_x}$. Thus, $Ec'\Omega_{x_i x_i} c = c'\Omega_{xx} c > 0$, which implies that Ω_{xx} is positive definite. Hence Ω_{xx}^{-1} exists, and so we have as $(T, n \to \infty)_{\text{seq}}$:

$$\hat{\beta}_{n,T} \xrightarrow{p} \Omega_{yx} \Omega_{xx}^{-1}.$$

Let $\beta = \Omega_{yx}\Omega_{xx}^{-1}$. We will call the parameter β the *long-run average* regression coefficient. It is the matrix regression coefficient (of y on x) associated with the long-run average covariance matrix Ω. To find the limit distribution of $\hat{\beta}_{n,T}$ we rescale the centered estimator $(\hat{\beta}_{n,T} - \beta)$ by $\sqrt{n}$ and let $T \to \infty$ for fixed n. For all fixed n as $T \to \infty$ we have

$$\sqrt{n}\left(\hat{\beta}_{n,T} - \beta\right) \Rightarrow \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\int M_{y_i} M_{x_i}' - \beta \int M_{x_i} M_{x_i}'\right) \left(\frac{1}{n} \sum_{i=1}^{n} \int M_{x_i} M_{x_i}'\right)^{-1}. \tag{4.5}$$

Note that

$$E\left(\int M_{y_i} M_{x_i}' - \beta \int M_{x_i} M_{x_i}'\right) = E\left[E\left(\int M_{y_i} M_{x_i}' - \beta \int M_{x_i} M_{x_i}' | F_{c_i}\right)\right]$$
$$= \tfrac{1}{2} E\left(\Omega_{y_i x_i} - \Omega_{yx}\Omega_{xx}^{-1}\Omega_{x_i x_i}\right) = 0,$$

where the conditional expectation exists because $\int M_{y_i} M_{x_i}' - \beta \int M_{x_i} M_{x_i}'$ is square integrable by Lemma 4. Thus, the numerator of (4.5) has mean zero. Also, we know that the numerator has finite second moments from Lemma 4, and with a straightforward calculation the variance matrix is found[6] to be

$$\begin{aligned}
&E\left(\operatorname{vec}\left(\int M_{y_i} M_{x_i}' - \beta \int M_{x_i} M_{x_i}'\right) \operatorname{vec}\left(\int M_{y_i} M_{x_i}' - \beta \int M_{x_i} M_{x_i}'\right)'\right) \\
&= \tfrac{1}{6} E\left(\Omega_{x_i x_i} \otimes (\Omega_{y_i y_i} - \beta\Omega_{x_i y_i} - \Omega_{y_i x_i}\beta' + \beta\Omega_{x_i x_i}\beta')\right) \\
&\quad + \tfrac{1}{6} E\left((\Omega_{x_i y_i} - \Omega_{x_i x_i}\beta') \otimes (\Omega_{y_i x_i} - \beta\Omega_{x_i x_i}) K_{m_y m_x}\right) \\
&\quad + \tfrac{1}{4} E\left(\operatorname{vec}(\Omega_{y_i x_i} - \beta\Omega_{x_i x_i})\left(\operatorname{vec}(\Omega_{y_i x_i} - \beta\Omega_{x_i x_i})\right)'\right) \\
&= \Theta, \qquad \text{say}
\end{aligned} \tag{4.6}$$

where $K_{m_y m_x}$ is the $(m_y m_x \times m_y m_x)$ commutation matrix (e.g., see Magnus and Neudecker (1988)). The sequence of random matrices $\{\int M_{y_i} M_{x_i}' - \beta \int M_{x_i} M_{x_i}'\}_i$ in the numerator of the matrix quotient (4.5) is iid $(0, \Theta)$ across i. From the multivariate Lindeberg-Levy theorem, we then get as $n \to \infty$

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\int M_{y_i} M_{x_i}' - \beta \int M_{x_i} M_{x_i}'\right) \Rightarrow N(0, \Theta). \tag{4.7}$$

[6] The calculations are given in Appendix C of PM[b].

Combining (4.7) with the limit

$$\frac{1}{n}\sum_{i=1}^{n}\int M_{x_i}M_{x_i}' \overset{\text{a.s.}}{\to} \frac{1}{2}\Omega_{xx} \qquad \text{as} \qquad n\to\infty,$$

we have the following limit distribution of the pooled estimator $\hat{\beta}_{n,T}$ as $(T, n \to \infty)_{\text{seq}}$

$$\sqrt{n}\left(\hat{\beta}_{n,T}-\beta\right) \Rightarrow N\left(0, 4\left(\Omega_{xx}^{-1}\otimes I_{m_y}\right)\Theta\left(\Omega_{xx}^{-1}\otimes I_{m_y}\right)\right). \tag{4.8}$$

Theorem 4 below shows that these results continue to hold in joint asymptotics as $(T, n \to \infty)$. For the limit distribution (4.8) to hold in this case we need the additional requirement that $n/T \to 0$. No additional condition is required for consistency.

THEOREM 4: *Suppose Assumptions* 1, 2, *and* 3 *hold.*

(a) *Then, as* $(n, T \to \infty)$, *we have* $\hat{\beta}_{n,T} \to_p \beta$.

(b) *If* $(n, T \to \infty)$ *and* $n/T \to 0$, *then*

$$\sqrt{n}\left(\hat{\beta}_{n,T}-\beta\right) \Rightarrow N\left(0, 4\left(\Omega_{xx}^{-1}\otimes I_{m_y}\right)\Theta\left(\Omega_{xx}^{-1}\otimes I_{m_y}\right)\right).$$

REMARKS: (a) The restriction $(n/T) \to 0$ in Theorem 4 controls the effects of bias in the panel regression. Under the assumptions on the DGP given in Section 2, the expectation of the components in the numerator of $\sqrt{n}(\hat{\beta}_{n,T}-\beta)$ is generally nonzero, i.e.,

$$E\left(\frac{1}{T^2}\sum_{t=1}^{T}(Y_{i,t}X_{i,t}' - \beta X_{i,t}X_{i,t}')\right) \neq 0,$$

whereas

$$E\left(\frac{1}{T^2}\sum_{t=1}^{T}(Y_{i,t}X_{i,t}' - \beta X_{i,t}X_{i,t}')\right) \to 0,$$

as $T \to \infty$ for all i. In this case, the condition $(n/T) \to 0$ prevents the bias from having a dominating asymptotic effect on the standardized quantity $\sqrt{n}(\hat{\beta}_{n,T} - \beta)$. But, when $(n/T) \nrightarrow 0$, the bias can dominate and the asymptotic behavior can be very different. For example, suppose that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}E\left(\frac{1}{T^2}\sum_{t=1}^{T}(Y_{i,t}X_{i,t}' - \beta X_{i,t}X_{i,t}')\right) \to b$$

along some diagonal limit $(T(n), n \to \infty)_{\text{diag}}$. In this event, we can expect to have a limit distribution with an asymptotic bias b. Further, the required restriction on the expansion rate between n and T will change depending on the underlying assumptions about the DGP. For example, if the shocks $U_{i,t}(=\Delta Z_{i,t})$ are iid over t and $Z_{i,0}=0$ for all i, then our results hold as $(n, T \to \infty)$ without imposing any restriction on the expansion rate between n and T.

(b) Theorem 4 holds for any partition of $Z_{i,t}$. If the panel data form a vector unit root model such as (2.1), then we can estimate the average long-run relation between any two subvectors. In effect, therefore, there is an average long-run relationship between any two subvector components of an integrated process over a cross section population.

(c) A key factor in determining these results is that panel data provide iid cross section information that is unavailable in a simple time series context. In consequence, we may expect some version of these results to apply when the regression utilizes only a fraction of the time series data. Suppose we regress $Y_{i,t}$ on $X_{i,t}$ using the cross section observations at time period $t=[Tr]$ with $0<r\leq 1$. The cross section OLS estimator $\tilde{\beta}_{n,[Tr]}$ is then defined by

$$\tilde{\beta}_{n,[Tr]} = \left(\sum_{i=1}^{n} Y_{i,t} X'_{i,t} \right) \left(\sum_{i=1}^{n} X_{i,t} X'_{i,t} \right)^{-1}. \tag{4.9}$$

Using similar arguments to those employed above, we can show that, in sequential asymptotics as $(T, n \to \infty)_{\text{seq}}$, we have

$$\tilde{\beta}_{n,[Tr]} \xrightarrow{p} \beta, \qquad \text{and}$$

$$\sqrt{n}\left(\tilde{\beta}_{n,[Tr]} - \beta \right) \Rightarrow N\Big(0, \left(\Omega_{xx}^{-1} \otimes I_{m_y} \right) \tilde{\Theta} \left(\Omega_{xx}^{-1} \otimes I_{m_y} \right)\Big),$$

where

$$\begin{aligned}
\tilde{\Theta} = {} & E\big(\Omega_{x_i x_i} \otimes (\Omega_{y_i y_i} - \beta \Omega_{x_i x_i} - \Omega_{x_i x_i} \beta' + \beta \Omega_{x_i x_i} \beta') \big) \\
& + E\Big((\Omega_{x_i y_i} - \Omega_{x_i x_i} \beta') \otimes (\Omega_{y_i x_i} - \beta \Omega_{x_i x_i}) K_{m_y m_x} \Big) \\
& + E\Big(\operatorname{vec}(\Omega_{y_i x_i} - \beta \Omega_{x_i x_i}) \big(\operatorname{vec}(\Omega_{y_i x_i} - \beta \Omega_{x_i x_i}) \big)' \Big).
\end{aligned}$$

(d) Since $(\Omega_{xx}^{-1} \otimes I_{m_y}) \tilde{\Theta} (\Omega_{xx}^{-1} \otimes I_{m_y}) - 4(\Omega_{xx}^{-1} \otimes I_{m_y}) \Theta (\Omega_{xx}^{-1} \otimes I_{m_y}) > 0$, the cross section estimator $\tilde{\beta}_n$ is asymptotically less efficient than the pooled estimator $\hat{\beta}_{n,T}$. This is to be expected because the pooled estimator $\hat{\beta}_{n,T}$ uses all the time series information while the cross section estimator $\tilde{\beta}_{n,[Tr]}$ uses only single time period information. It is therefore interesting to note that, although time series regression may be spurious, use of all the time series data does reduce the limiting variance in a panel regression. Heuristically, this is because when we pool the data we average the limiting information and quantities like $\int_0^1 M_{y_i}$ and $\int_0^1 M_{y_i} M'_{x_i}$ have less variation than $M_{y_i}(1)$ and $M_{y_i}(1) M_{x_i}(1)$. For example, $W(1) \equiv N(0,1)$, whereas $\int_0^1 W \equiv N(0, \frac{1}{3})$.

5. PANEL COINTEGRATION

This section considers the case where the variables in $Z_{i,t}$ are cointegrated. As discussed in Phillips (1986), there exists a cointegrating relation among the variables in $Z_{i,t}$ if the conditional long-run variance matrix Ω_i of $Z_{i,t}$ has deficient rank. We will discuss three particular types of model: (i) heterogeneous panel cointegration, where there exist different cointegrating relations among the variables in $Z_{i,t}$ across individuals; (ii) homogeneous panel cointegration, where the cointegration relation is the same for all the individuals; and (iii) near-homogeneous panel cointegration, where there exist slightly different cointegrating relations across the individuals.

5.1. *Heterogeneous Panel Cointegration*

We start by strengthening the moment conditions of the random coefficients $C_{i,t}$ in (2.3) and the summability conditions as follows. These conditions help to ensure the existence of a valid BN decomposition for the equation errors in the panel cointegration model (5.2) given below.

ASSUMPTION 4 (Random Coefficient Conditions′):
(i) *Assumption* 1(*i*) *holds.*
(ii) $EC_{a,i,t}^{16} (= \sigma_{16,a,t}) < \infty$ *for all* $a = 1, \ldots, m^2$.

ASSUMPTION 5 (Summability Condition′): *For all* $a = 1, \ldots, m^2$:
(i) $\sum_{t=0}^{\infty} t^2 \sigma_{2,a,t} < \infty$;
(ii) $\sum_{t=0}^{\infty} t^4 (\sigma_{4,a,t})^{1/4} < \infty$;
(iii) $\sum_{t=0}^{\infty} t^2 (\sigma_{8,a,t})^{1/8} < \infty$;
(iv) $\sum_{t=0}^{\infty} (\sigma_{16,a,t})^{1/16} < \infty$.

The previous section assumes that the conditional long-run covariance matrix Ω_i of the integrated vector $Z_{i,t}$ in (2.1) is positive definite. When Ω_i is singular, important differences arise in the time series case, as is well known, and a different large T time theory applies for each i.

ASSUMPTION 6: *The following conditions hold almost surely.*
(i) Ω_i *has rank* m_x.
(ii) *Each* $(m_x \times m_x)$ *leading submatrix* $\Omega_{x_i x_i}$ *is positive definite.*

In this case the generating mechanism (2.1) has a deficient set of unit roots and the vector $Z_{i,t}$ is cointegrated almost surely. To see this, take an arbitrary element of the probability space for which (i) and (ii) of Assumption 6 hold. Then we have $\Omega_{y_i y_i} = \Omega_{y_i x_i} \Omega_{x_i x_i}^{-1} \Omega_{x_i y_i}$. Let $\alpha_i = (I_{m_y}, -\beta_i)$ and $\beta_i = \Omega_{y_i x_i} \Omega_{x_i x_i}^{-1}$. The $(m_y \times m)$ random matrix α_i is well defined because $\Omega_{x_i x_i}$ is positive definite. Since $\Omega_i = C_i(1) C_i(1)'$, the equality $\Omega_{y_i y_i} - \Omega_{y_i x_i} \Omega_{x_i x_i}^{-1} \Omega_{x_i y_i} = 0$ can be written as

$$\alpha_i C_i(1) C_{y_i}(1)' = 0, \tag{5.1}$$

so that α_i is in the row null space of the matrix $C_i(1)$.

Define $E_{i,t} = \alpha_i Z_{i,t} = Y_{i,t} - \beta_i X_{i,t}$. Note that $\Delta E_{i,t} = \alpha_i \, \Delta Z_{i,t} = \alpha_i U_{i,t} = \alpha_i C_i(1) V_{i,t} - \Delta \alpha_i \tilde{U}_{i,t}$, where the last equality comes from the BN decomposition of $U_{i,t}$. Then, since $\alpha_i C_i(1) = 0$, $\Delta E_{i,t} = -\Delta\alpha_i \tilde{U}_{i,t}$, that is $E_{i,t} = -\alpha_i \tilde{U}_{i,t} = -\alpha_i \sum_{s=0}^{\infty} \tilde{C}_{i,s} V_{i,t-s}$. Lemma 15 in Appendix D shows that $E_{i,t}$ is square integrable and that the random coefficients $\{-\alpha_i \tilde{C}_{i,s}\}_s$ are summable. Hence, Assumption 6 implies the existence of the following panel cointegration model with probability one:

$$(5.2) \qquad Y_{i,t} \overset{\text{a.s.}}{=} \beta_i X_{i,t} + E_{i,t},$$
$$X_{i,t} = X_{i,t-1} + U_{x_i,t},$$

where

$$(5.3) \qquad F_{i,t} = \begin{pmatrix} E_{i,t} \\ U_{x_i,t} \end{pmatrix} = \sum_{s=0}^{\infty} G_{i,s} V_{i,t-s},$$
$$G_{i,s} = \begin{pmatrix} -\alpha_i \tilde{C}_{i,s} \\ \Gamma C_{i,s} \end{pmatrix}, \qquad \text{and} \qquad \Gamma = (0 \vdots I_{m_x})(m_x \times m).$$

The coefficient β_i in model (5.2) is random. This means that β_i differs randomly across i and so the cointegrating relation between $Y_{i,t}$ and $X_{i,t}$ is heterogenous. Also, the random coefficients $G_{i,s}$ in the linear process generating $(E'_{i,t}, U'_{x_i,t})'$ in model (5.2) each involve the cointegrating matrix α_i whose main component is $\beta_i = \Omega_{y_i x_i} \Omega_{x_i x_i}^{-1}$, which depends on the inverse of $\Omega_{x_i x_i}$. From Assumption 6, $\Omega_{x_i x_i}^{-1}$ exists almost surely. But, additionally, we need some moment conditions on $\Omega_{y_i x_i} \Omega_{x_i x_i}^{-1}$ to ensure the existence of moments of the random coefficients $G_{i,s}$, which help in establishing the validity of a panel BN decomposition. Assumptions 1 and 2 alone do not assure the existence of moments of $\Omega_{x_i x_i}^{-1}$. Hence, to avoid heavy tails in the density of $\Omega_{x_i x_i}^{-1}$, we make the following assumption about the distribution of $\Omega_{x_i x_i}$.

ASSUMPTION 7: *The random matrix $\Omega_{x_i x_i}$ has continuous density function f with the following properties.*

(i) $f(\Omega) = O(\operatorname{etr}(-c\Omega))$ *for some* $c > 0$ *when* $\operatorname{tr}(\Omega) \to \infty$, *where* $\operatorname{etr}(-c\Omega)$ *denotes* $\exp\{\operatorname{tr}(-c\Omega)\}$.

(ii) $f(\Omega) = O((\det \Omega)^{\gamma})$ *for some* $\gamma > 7$ *when* $\det(\Omega) \to 0$.

REMARKS: (a) Condition (i) implies that the tail of the density f is exponentially small as $\operatorname{tr}(\Omega) \to \infty$. Condition (ii) restricts the behavior of the density f when $\det \Omega \to 0$. Taken together (i) and (ii) ensure that $(\det \Omega)^s f(\Omega)$ is integrable for $s \geq -8$.

(b) An example of a density f satisfying conditions (i) and (ii) is the Wishart distribution $W_{m_x}(J, I_{m_x})$ whose probability element is

$$f(\Omega)(d\Omega) = \frac{1}{2^{m_x(J/2)\Gamma_{m_x}(J/2)}} \operatorname{etr}\left(-\frac{1}{2}\Omega\right) \det \Omega^{(J-m_x-1)/2}(d\Omega),$$

with degrees of freedom parameter $J > m_x + 15$ and where

$$\Gamma_{m_x}\left(\frac{J}{2}\right) = \int_{\Omega > 0} \operatorname{etr}(-\Omega) \det \Omega^{J - m_x - 1/2} (d\Omega).$$

In this case, Ω^{-1} has an inverse Wishart distribution with $(J + m_x + 1)$ degrees of freedom and $(m_x \times m_x)$ parameter matrix I_{m_x}, $W_{m_x}^{-1}(J + m_x + 1, I_{m_x})$ (e.g., see Muirhead (1982, p. 113).

In view of Lemma 15(a) in Appendix D, $\sum_{s=1}^{\infty} s^2 \|G_{i,s}\|^2 < \infty$. a.s. Then, as in Lemma 2, it follows from Phillips and Solo (1992) that $F_{i,t}$ has a valid panel BN decomposition of the form

$$(5.4) \qquad F_{i,t} \overset{\text{a.s.}}{=} G_i(1)V_{i,t} + \tilde{F}_{i,t-1} - \tilde{F}_{i,t},$$

where $G_i(1)V_{i,t}$ and $\tilde{F}_{i,t}$ are well defined square integrable random vectors in view of Lemma 15. Using (5.4), the partial sum process of $F_{i,t}$ can be written as

$$(5.5) \qquad \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} F_{i,t} \overset{\text{a.s.}}{=} G_i(1) \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} V_{i,t} + \frac{1}{\sqrt{T}} \tilde{F}_{i,0} - \frac{1}{\sqrt{T}} \tilde{F}_{i,[Tr]}.$$

With this BN decomposition in place, we can use the Phillips-Solo approach to deduce a functional law for partial sums of $F_{i,t}$. In particular, we have the following lemma:

LEMMA 7: *If Assumptions* 4–7 *hold, then*

$$(5.6) \qquad \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} F_{i,t} \Rightarrow G_i(1)W_i(r) \qquad \text{as } T \to \infty \text{ for all } i,$$

where $W_i(r)$ is a standard vector Brownian motion independent of F_{c_i}.

Thus, $(1/\sqrt{T})\sum_{t=1}^{[Tr]} F_{i,t}$ converges in distribution to a randomly scaled (or mixed) Brownian motion $G_i(1)W_i(r)$ as $T \to \infty$ for all i. Let $S_{i,t} = \sum_{s=1}^{t} F_{i,s} + S_{i,0}$, where $S_{i,0}$ are iid across i with $E\|S_{i,0}\|^4 < \infty$. The next lemma shows that $(1/T)\sum_{t=1}^{T} S_{i,t}F'_{i,t}$ converges in distribution to a matrix stochastic integral plus an F_{c_i}-measurable random matrix.

LEMMA 8: *Suppose the assumptions in Lemma* 7 *hold. Then,*

$$(5.7) \qquad \frac{1}{T} \sum_{t=1}^{T} F_{i,t} S'_{i,t} \Rightarrow G_i(1) \int dW_i \, W_i' G_i(1)' + \Lambda_i \qquad \text{as } T \to \infty,$$

where $\Lambda_i = \sum_{k=0}^{\infty} E(F_{i,k} F'_{i,0} | F_{c_i}) = \sum_{k=0}^{\infty} \sum_{s=0}^{\infty} G_{i,s+k} G'_{i,s}$.

Partition $G_i(1)$, Λ_i, and $G_i(1)W_i(r)$ conformably as follows:

$$G_i(1) = \begin{pmatrix} G_{e_i}(1) \\ G_{x_i}(1) \end{pmatrix} \begin{matrix} m_y \\ m_x \end{matrix}, \qquad \Lambda_i = \begin{pmatrix} \Lambda_{e_i e_i} & \Lambda_{e_i x_i} \\ \Lambda_{x_i e_i} & \Lambda_{x_i x_i} \end{pmatrix},$$

$$G_i(1)W_i(r) = \begin{pmatrix} G_{e_i}(1)W_i(r) \\ G_{x_i}(1)W_i(r) \end{pmatrix} = \begin{pmatrix} M_{e_i}(r) \\ M_{x_i}(r) \end{pmatrix}.$$

Consider the time series regression of $Y_{i,t}$ on $X_{i,t}$. Using (5.6), (5.7), and the continuous mapping theorem, we find the following large T limit distribution for the OLS estimator of the (random) coefficient β_i:

$$\text{(5.8)} \qquad T\left(\hat{\beta}_i - \beta_i\right) \overset{\text{a.s.}}{=} \left(\frac{1}{T}\sum_{t=1}^{T} E_{i,t}X'_{i,t}\right)\left(\frac{1}{T^2}\sum_{t=1}^{T} X_{i,t}X'_{i,t}\right)^{-1}$$

$$\Rightarrow \left(G_{e_i}(1)\int dW_i\, W_i' G_{x_i}(1)' + \Lambda_{e_i x_i}\right)\left(G_{x_i}(1)\int W_i W_i' G_{x_i}(1)'\right)^{-1}$$

$$= \left(\int dM_{e_i}\, M'_{x_i} + \Lambda_{e_i x_i}\right)\left(\int M_{x_i} M'_{x_i}\right)^{-1} \qquad \text{as } T \to \infty \text{ for all } i.$$

The bias term $\Lambda_{e_i x_i}$ arises in the usual way from the temporal correlation between $E_{i,t}$ and $U_{x_i,t}$ (c.f., Phillips and Durlauf (1986)). Thus, time series regression produces a consistent estimator of the cointegrating matrix β_i, and thereby distinguishes the randomly differing individual long-run relations between $Y_{i,t}$ and $X_{i,t}$.

When both dimensions of the panel data are utilized, a long-run average coefficient β is also identified. This can be accomplished as in the previous section, by means of a pooled panel regression or a limiting cross section regression. The following sections concentrate on pooled panel regression and discuss limiting cross section estimators only briefly.

In the heterogeneous panel cointegration model (5.2) the pooled estimator $\hat{\beta}_{n,T}$ has the same form as that defined in (4.4). The limit theory for this pooled estimator is as follows.

THEOREM 5: *Let the assumptions of Lemma* 15 *hold. Then*:
(a) *as* $(n, T \to \infty)$, $\hat{\beta}_{n,T} \to_p \beta = \Omega_{yx}\Omega_{xx}^{-1}$;
(b) *as* $(n, T \to \infty)$ *with* $n/T \to 0$,

$$\sqrt{n}\left(\hat{\beta}_{n,T} - \beta\right) \Rightarrow N\left(0, 4\left(\Omega_{xx}^{-1} \otimes I_{m_y}\right)\Theta\left(\Omega_{xx}^{-1} \otimes I_{m_y}\right)\right),$$

where

$$\Theta = \tfrac{1}{6}\left(\Omega_{x_i x_i} \otimes (\Omega_{y_i y_i} - \beta\Omega_{x_i x_i} - \Omega_{x_i x_i} - \Omega_{x_i x_i}\beta' + \beta\Omega_{x_i x_i}\beta')\right)$$
$$+ \tfrac{1}{6}E\left((\Omega_{x_i y_i} - \Omega_{x_i x_i}\beta') \otimes (\Omega_{y_i x_i} - \beta\Omega_{x_i x_i})\right)K_{m_x m_y}$$
$$+ \tfrac{1}{4}E\left(\text{vec}(\Omega_{y_i x_i} - \beta\Omega_{x_i x_i})\text{vec}(\Omega_{y_i x_i} - \beta\Omega_{x_i x_i})'\right).$$

REMARKS: (a) Define $\tilde{E}_{i,t} = (\beta_i - \beta) X_{i,t} + E_{i,t}$. Then the heterogeneous panel cointegration model (5.2) becomes

$$(5.9) \qquad Y_{i,t} = \beta X_{i,t} + \tilde{E}_{i,t}.$$

The pooled estimator $\hat{\beta}_{n,T}$ is a least squares estimator of the regression coefficient in (5.9) and consistently estimates the long-run average coefficient β between $Y_{i,t}$ and $X_{i,t}$. Note that the noise, $\tilde{E}_{i,t}$, in this regression involves the integrated random vector $X_{i,t}$. By the same logic as that of the spurious regression case, the long-run coefficient β is consistently estimated by pooling the panel data because cross section pooling attenuates the strength of the noise $\tilde{E}_{i,t}$ relative to the signal in the regression (5.9).

(b) As seen in Theorem 5, the pooled estimator $\hat{\beta}_{n,T}$ is $\sqrt{n}$ consistent for the average long-run regression coefficient β and has a normal limit distribution. Observe that the limit variance matrix for the heterogeneous pooled panel regression estimator in Theorem 5, viz., $4(\Omega_{xx}^{-1} \otimes I_{m_y})(\Omega_{xx}^{-1} \otimes I_{m_y})$ has precisely the same form as the limit variance matrix of the spurious regression pooled panel regression estimator in Theorem 4. This equivalence in form is especially interesting because the individual long-run covariance matrix Ω_i is singular in the heterogeneous cointegration case but nonsingular in the spurious regression case, so that these individual component matrices must be different between the two models. Nevertheless, and in spite of these differences, the average long-run covariance matrix Ω may well be nonsingular in the heterogeneous cointegration model, in which case there is a basis for direct comparison between the two results. Obviously, the effect of the heterogeneity in the cointegration parameter is to slow down the rate of convergence of the pooled estimator. In particular, the convergence rate is $\sqrt{n}$ and, interestingly, this rate is uninfluenced by the time series sample size in spite of the fact that the individual time series regressions are themselves T-consistent (see (5.8)). Thus, there is a correspondence in the limit theory between the heterogeneous cointegration model and the pooled spurious regression model after pooling the data.

(c) In general, of course, $E[\Omega_{y_i x_i} \Omega_{x_i x_i}^{-1}] \neq E[\Omega_{y_i x_i}](E[\Omega_{x_i x_i}])^{-1}$, so there is no reason why the limit of the average of the cointegrating relation $(1/n)\sum_{i=1}^n \beta_i$ should equal β, the average long-run regression coefficient. As we have seen, it is the latter parameter that is the limit of the pooled regression estimator in the heterogeneous cointegration model. One situation where $\lim_{n\to\infty}(1/n)\sum_{i=1}^n \beta_i = \beta$ does hold is when $\Omega_{x_i x_i}$ has a degenerate distribution, namely, $\Omega_{x_i x_i} = \Omega_{xx}$ almost surely. Thus, in the heterogeneous panel cointegration case, the parameter being estimated is *not* the average cointegrating coefficient, but the average long-run regression coefficient, just as in the spurious panel regression case. Again, the two models are much closer than they might appear.

(d) As discussed in (a), the heterogeneous panel cointegration model can be reinterpreted in the form of the panel model (5.9). As such, we may be interested in constructing statistical tests about the long-run coefficients β. For example, to test $\mathbb{H}_0$: $\varphi(\beta) = 0$, where $\varphi(\cdot)$ is a p-vector of smooth functions on a subset of $\mathbb{R}^{m_y \times m_x}$ such that $\partial\varphi/\partial\beta'$ has full rank $p(\leq m_y m_x)$, we may use the Wald statistic

$$W_{\varphi} = n\varphi\left(\hat{\beta}_{n,T}\right)'\hat{V}_{\varphi}^{-1}\varphi\left(\hat{\beta}_{n,T}\right),$$

where $\hat{V}_{\varphi} = (\partial\varphi(\hat{\beta}_{n,T})/\partial\beta')(\partial\varphi(\hat{\beta}_{n,T})'/\partial\beta)$, $\hat{V}_{\beta} = 4(\hat{\Omega}_{xx}^{-1} \otimes I_{m_y})\hat{\Theta}(\hat{\Omega}_{xx}^{-1} \otimes I_{m_y})$,

$$\hat{\Theta} = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{1}{T^4}\sum_{s,t=1}^{T} X_{i,t}X_{i,s}' \otimes \hat{E}_{i,t}\hat{E}_{i,s}'\right\},$$

$$\hat{\Omega}_{xx}^{-1} = \left[\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{2}{T^2}\sum_{t=1}^{T} X_{i,t}X_{i,t}'\right\}\right]^{-1},$$

and $\hat{E}_{it} = Y_{i,t} - \hat{\beta}_{n,T}X_{i,t}$. Some simple manipulations in the case of sequential asymptotics show that this statistic leads to a standard asymptotic χ^2 test as $(T, n \to \infty)_{\text{seq}}$. This limit theory also holds very generally under joint limits as $(T, n \to \infty)$ as the next result reveals.

THEOREM 6: *Under* $\mathbb{H}_0$: $\varphi(\beta) = 0$ *and Assumptions* 4–7, $W_{\varphi} \Rightarrow \chi_p^2$, *as* $(n, T \to \infty)$ *with* $n/T \to 0$.

(e) We may also be interested in testing hypotheses about the coefficients in generalizations of model (5.9) of the following form:

$$Y_{i,t} = \beta_{\mu}X_{i,t} + \tilde{E}_{i,t} \qquad \text{with} \qquad \begin{cases} \beta_{\mu} = \beta_a \text{ for } i \in I_a, \\ \beta_{\mu} = \beta_b \text{ for } i \in I_b, \end{cases} \tag{5.10}$$

where I_a and I_b are index sets corresponding to subgroups of the cross section population for which the long-run average covariance matrices are Ω_a and Ω_b, respectively, leading to long-run average regression coefficients $\beta_a = \Omega_{a,yx}\Omega_{a,xx}^{-1}$ and $\beta_b = \Omega_{b,yx}\Omega_{b,xx}^{-1}$ that may differ between the two populations. Models like (5.10) can be readily extended to multi-category models and will be empirically relevant, for example, in cross country panel regressions where countries are partitioned into classes of similar category like developed (OECD) nations and developing and undeveloped nations. Note that in such cases the model (5.10) allows for intra-class variation (i.e., the regression coefficients β_i for $i \in I_a$ will differ) but our primary interest lies in the inter-class difference $\beta_a - \beta_b$. A natural hypothesis is then: $\mathbb{H}_0$: $\beta_a = \beta_b$. Let $n_a = \#(I_a)$ and $n_b = \#(I_b)$, respectively. Suppose that $n_b/n_a \to \kappa < \infty$ as $n_a, n_b \to \infty$. The null hypothesis can be tested by constructing pooled regression coefficients $\hat{\beta}_a, \hat{\beta}_b$ in each class and computing the Wald statistic

$$W_{a,b} = n_b\left\{\text{vec}\left(\hat{\beta}_a - \hat{\beta}_b\right)'\hat{V}_{a-b}^{-1}\,\text{vec}\left(\hat{\beta}_a - \hat{\beta}_b\right)\right\},$$

where $\hat{V}_{a-b} = (n_a/n_b)\hat{V}_a + \hat{V}_b, \hat{V}_\mu = 4(\hat{\Omega}_{\mu,xx}^{-1} \otimes I_{m_x})\hat{\Theta}_\mu(\hat{\Omega}_{\mu,xx}^{-1} \otimes I_{m_x})$,

$$\hat{\Theta}_\mu = \frac{1}{n_\mu} \sum_{i \in I_\mu} \left\{ \frac{1}{T^4} \sum_{s,t=1}^{T} X_{i,t} X'_{i,s} \otimes \hat{E}_{i,t} \hat{E}'_{i,s} \right\},$$

$$\hat{\Omega}_{\mu,xx}^{-1} = \left[\frac{1}{n_\mu} \sum_{i \in I_\mu} \left\{ \frac{1}{T^2} \sum_{t=1}^{T} X_{i,t} X'_{i,t} \right\} \right]^{-1},$$

and $\hat{E}_{it} = Y_{i,t} - \hat{\beta}_\mu X_{i,t}$ with $\mu \in \{a, b\}$. Again, this leads to an asymptotic χ^2 test. The following result gives the limit theory under joint limits as $(n_a, n_b, T \to \infty)$ and can be obtained in a simple way from Theorems 5 and 6.

THEOREM 7: *Under* $\mathbb{H}_0$: $\beta_a = \beta_b$ *and Assumptions* 4–7, $W_{a,b} \Rightarrow \chi^2_{m_y m_x}$, *as* $(n_a, n_b, T \to \infty)$ *with* $n_a/T, n_b/T \to 0$.

5.2. *Homogeneous Panel Cointegration and Pooled FM Estimation*

This section considers a homogeneous panel cointegration model, where the cointegrating relations are the same across individuals, and develops an asymptotic theory for a pooled FM estimator. We start with the following simplifying assumption.

ASSUMPTION 8: $C_{i,t} =_{\text{a.s.}} C_t$, *where* C_t *is an* $(m \times m)$ *nonrandom matrix for all* t.

Then, under Assumption 6, the panel cointegration model (5.2) becomes

$$Y_{i,t} \overset{\text{a.s.}}{=} \beta X_{i,t} + E_{i,t}, \quad X_{i,t} = X_{i,t-1} + U_{x_i,t}, \tag{5.11}$$

where

$$\beta = \Omega_{yx}\Omega_{xx}^{-1}, \qquad \alpha = \left(I_{m_y}, -\beta\right), \qquad \begin{pmatrix} E_{i,t} \\ U_{x_i,t} \end{pmatrix} = \sum_{s=0}^{\infty} G_s V_{i,t-s},$$

$$G_s = \begin{pmatrix} G_{e,s} \\ G_{x,s} \end{pmatrix} = \begin{pmatrix} -\alpha \tilde{C}_s \\ \gamma C_s \end{pmatrix}, \qquad \text{and} \qquad \tilde{C}_s = \sum_{j=s+1}^{\infty} C_j.$$

In this model the same long-run relation between $Y_{i,t}$ and $X_{i,t}$ applies for all i. Unlike previous models, the error term in model (5.11) is generated by a linear process with nonrandom coefficients $\{G_s\}$, on which we impose the following summability condition.

ASSUMPTION 9: $\sum_{s=0}^{\infty} s^3 \|C_s\| < \infty$.

Define $\tilde{G}_s = \sum_{j=s+1}^{\infty} G_j$. Under Assumption 9, $G(1) = \sum_{s=0}^{\infty} G_s < \infty$ and $\sum_{s=0}^{\infty} s^2 \|\tilde{G}_s\| = \sum_{s=0}^{\infty} s^2 \|\sum_{j=s+1}^{\infty} G_j\|^2 < \infty$. Write $F_{i,t} = (E'_{i,t}, U'_{x_i,t})'$. Then, as $T \to \infty$, we have the functional law $(1/\sqrt{T})\sum_{t=1}^{[Tr]} F_{i,t} \Rightarrow B_i(r) \equiv BM(\Omega_F)$, where $\Omega_F = G(1)G(1)'$ (Theorem 3.4 in Phillips and Solo (1992)). The following assumption is conventional in time series cointegration analysis, although it could be relaxed with some consequential changes in the asymptotics, including changes in convergence rates in directions determined by the singularity.

ASSUMPTION 10: *Ω_F is positive definite.*

Partition $B_i(r) = (B_{e_i}(r)', B_{x_i}(r)')'$ conformably with $F_{i,t}$. Set $S_{i,t} = \sum_{s=1}^{t} F_{i,s} + S_{i,0}$, where $S_{i,0}$ are iid across i with $E\|S_{i,0}\|^4 < \infty$. Then, in the usual way (Phillips (1988)), as $T \to \infty$

$$\frac{1}{T}\sum_{t=1}^{T} F_{i,t} S'_{i,t} \Rightarrow \int dB_i\, B'_i + \Lambda_F,$$

where $\Lambda_F = \sum_{k=0}^{\infty} E(F_{i,k} F'_{i,0}) = \sum_{k=0}^{\infty}\sum_{s=0}^{\infty} G_{s+k} G'_s$. Again, conformably partition Ω_F and Λ_F as

$$\begin{pmatrix} \Omega_{ee} & \Omega_{ex} \\ \Omega_{xe} & \Omega_{xx} \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \Lambda_{ee} & \Lambda_{ex} \\ \Lambda_{xe} & \Lambda_{xx} \end{pmatrix}, \quad \text{respectively.}$$

For each i, model (5.11) is a time series cointegrating regression. The least squares estimator

$$\begin{aligned}\hat{\beta}_i &= \sum_{t=1}^{T} Y_{i,t} X'_{i,t} \left(\sum_{t=1}^{T} X_{i,t} X'_{i,t}\right)^{-1} \\ &\overset{\text{a.s.}}{=} \beta + \sum_{t=1}^{T} E_{i,t} X'_{i,t} \left(\sum_{t=1}^{T} X_{i,t} X'_{i,t}\right)^{-1}\end{aligned}$$

has the following asymptotic distribution (Phillips and Durlauf (1986)):

$$(5.12) \qquad T\left(\hat{\beta}_i - \beta\right) \Rightarrow \left(\int dB_{e_i}\, B'_{x_i} + \Lambda_{ex}\right)\left(\int B_{x_i} B'_{x_i}\right)^{-1} \quad \text{as} \quad T \to \infty.$$

The time series estimator $\hat{\beta}_i$ is therefore consistent for β, the common long-run coefficient for all i, although there may be a second order bias effect entering through the term Λ_{ex} in (5.12) arising from the correlation between $E_{i,t}$ and $X_{i,s}$.

When the panel observations are pooled, as in the estimator $\hat{\beta}_{n,T}$ defined in (4.4), we get

$$\hat{\beta}_{n,T} \overset{\text{a.s.}}{=} \beta + \sum_{i=1}^{n} \sum_{t=1}^{T} E_{i,t} X'_{i,t} \left(\sum_{i=1}^{n} \sum_{t=1}^{T} X_{i,t} X'_{i,t} \right)^{-1}.$$

When $\Lambda_{ex} = 0$, the limit theory of this estimator is as follows.

THEOREM 8: *Suppose that Assumptions* 6, 8–10 *hold. If* $\Lambda_{ex} = 0$, *then as* $(n, T \to \infty)$ *with* $n/T \to 0$,

$$\sqrt{n}\, T\left(\hat{\beta}_{n,T} - \beta \right) \Rightarrow N\left(0, 2(\Omega_{xx}^{-1} \otimes \Omega_{ee})\right).$$

Thus, if $E_{i,t}$ and $U_{x_i,s}$ are uncorrelated, so that the one sided long-run covariance $\Lambda_{ex} = 0$, the pooled estimator $\hat{\beta}_{n,T}$ is $\sqrt{n}\, T$ consistent and has a limiting normal distribution in joint asymptotics as $(n, T \to \infty)$ when $n/T \to 0$.

When $\Lambda_{ex} \neq 0$, we do not attain $\sqrt{n}\, T$ consistency with the pooled least squares estimator $\hat{\beta}_{n,T}$, because of the persistence of bias effects. However, we may 'fully modify' the regressor $Y_{i,t}$ to eliminate the serial correlation Λ_{ex}. Originally, the fully modified (FM) regression method was introduced in Phillips and Hansen (1990) to correct for the presence of endogeneity (the correlation between B_{e_i} and B_{x_i}) and serial correlation in the OLS estimator $\hat{\beta}_i$ of the individual cointegration regression model. Their construction calls for consistent time series estimators $\hat{\Omega}_F$ and $\hat{\Lambda}_F$ of Ω_F and Λ_F. In our case, consistent estimates may be constructed using averages (over $i = 1, \ldots, n$) of the usual consistent (as $T \to \infty$) nonparametric kernel estimates of the corresponding long-run quantities for each i. More specifically, let $\hat{\Gamma}_i(j) = (1/T)\sum_t F_{i,t+j} F'_{i,t}$, where the summation is over $1 \leq t, t+j \leq T$, and define the averaged kernel estimators

$$\hat{\Omega}_F = \frac{1}{n} \sum_{i=1}^{n} \hat{\Omega}_{F,i}, \qquad \hat{\Omega}_{F,i} = \sum_{j=-T+1}^{T-1} w\left(\frac{j}{K} \right) \hat{\Gamma}_i(j), \tag{5.13}$$

$$\hat{\Lambda}_F = \frac{1}{n} \sum_{i=1}^{n} \hat{\Lambda}_{F,i}, \qquad \hat{\Lambda}_{F,i} = \sum_{j=0}^{T-1} w\left(\frac{j}{K} \right) \hat{\Gamma}_i(j),$$

where $w(x)$ is a lag kernel for which $w(0) = 1$, $w(x) = w(-x)$, $\int_{-\infty}^{\infty} w(x)^2\, dx < \infty$, and with Parzen's exponent $q \in (0, \infty)$ such that $k_q = \lim_{x \to \infty} (1 - w(x))/|x|^q < \infty$ (e.g., see Hannan (1970) or Andrews (1991)).[7] As is well know in the nonparametric literature, the choice of the bandwidth K is important in the limit

[7] In determining asymptotic properties of kernel estimates of the long-run variance we usually also impose a smoothness restriction on the spectral density at the origin. This smoothness condition can be formulated as a summability condition on the autocovariance sequence $\Gamma(h) = E(F_{i,t} F'_{i,t+h})$. The summability conditions in Assumption 9 ensure that $\sum_{h=0}^{\infty} h^2 \|\Gamma(h)\| < \infty$, and provide sufficient smoothness for our results here.

behavior of $\hat{\Omega}_F$. Under the summability condition given in Assumption 9, it is known that $\hat{\Omega}_{F,i} \to \Omega_F$ if $K, T \to \infty$ with $K/T \to 0$. However, later in this section (e.g., for Theorem 9) we need the stronger result that $\sqrt{n}(\hat{\Omega}_F - \Omega_F), \sqrt{n}(\hat{\Lambda}_F - \Lambda_F) = o_p(1)$ as $(n, T \to \infty)$ with $n/T \to 0$. The following Assumption about bandwidth choice is made so that these conditions apply.

ASSUMPTION 11: *The lag kernel* $w(\cdot)$ *in* (5.13) *has Parzen exponent* $q > 1/2$, *and the bandwidth parameter K tends to infinity with* $K/T \to 0$ *and* $K^{2q}/T \to \epsilon > 0$.

Define

$$Y^+_{i,t} = Y_{i,t} - \hat{\Omega}_{ex}\hat{\Omega}^{-1}_{xx}\,\Delta X_{i,t} \tag{5.14}$$

and

$$\hat{\Lambda}^+_{ex} = \hat{\Lambda}_{ex} - \hat{\Omega}_{ex}\hat{\Omega}^{-1}_{xx}\hat{\Lambda}_{xx}. \tag{5.15}$$

Equation (5.14) gives the endogeneity correlation and equation (5.15) gives the serial correlation correction.

Using these corrections, a pooled FM (PFM) estimator can be defined as follows:

$$\begin{aligned}\hat{\beta}_{PFM} &= \left(\sum_{i=1}^{n}\sum_{t=1}^{T} Y^+_{i,t}X'_{i,t} - nT\hat{\Lambda}^+_{ex}\right)\left(\sum_{i=1}^{n}\sum_{t=1}^{T} X_{i,t}X'_{i,t}\right)^{-1} \\ &\overset{\text{a.s.}}{=} \beta + \left(\sum_{i=1}^{n}\left(\sum_{t=1}^{T}\hat{E}^+_{i,t}X'_{i,t} - T\hat{\Lambda}^+_{ex}\right)\right)\left(\sum_{i=1}^{n}\sum_{t=1}^{T} X_{i,t}X'_{i,t}\right)^{-1},\end{aligned} \tag{5.16}$$

where $\hat{E}^+_{i,t} = E_{i,t} - \hat{\Omega}_{ex}\hat{\Omega}^{-1}_{xx}\,\Delta X_{i,t}$. Rescaling $\hat{\beta}_{FPM} - \beta$ by $\sqrt{n}\,T$ and letting $T \to \infty$ for fixed n, we have

$$\sqrt{n}\,T\left(\hat{\beta}_{PFM} - \beta\right) \Rightarrow \left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int dB_{e_i,x_i}B'_{x_i}\right)\left(\frac{1}{n}\sum_{i=1}^{n}\int B_{x_i}B'_{x_i}\right)^{-1},$$

where $B_{e_i,x_i}(r) \equiv BM(\Omega_{e.x})$ and $\Omega_{e.x} = \Omega_{ee} - \Omega_{ex}\Omega^{-1}_{xx}\Omega_{xe}$.

Note that $B_{e_i.x_i}(r)$ and $B_{x_i}(r)$ are independent, so $E\int dB_{e_i.x_i}B'_{x_i} = 0$ and

$$E\left(\operatorname{vec}\int dB_{e_i.x_i}B'_{x_i}\right)\left(\operatorname{vec}\int dB_{e_i.x_i}B'_{x_i}\right)' = \tfrac{1}{2}(\Omega_{xx} \otimes \Omega_{e.x}).$$

Thus, applying the multivariate Lindeberg-Levy theorem to

$$(1/\sqrt{n})\textstyle\sum_{i=1}^{n}\int dB_{e_i.x_i}B'_{x_i}$$

and combining this with the limit of $(1/n)\sum_{i=1}^{n}\int B_{x_i}B'_{x_i}$, we find that as $n \to \infty$

$$\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int dB_{e_i.x_i}B'_{x_i}\right)\left(\frac{1}{n}\sum_{i=1}^{n}\int B_{x_i}B'_{x_i}\right)^{-1} \Rightarrow N\left(0, 2(\Omega^{-1}_{xx} \otimes \Omega_{e.x})\right).$$

Thus, $\sqrt{n}\,T(\hat{\beta}_{FPM}-\beta)\Rightarrow N(0,2(\Omega_{xx}^{-1}\otimes\Omega_{e.x}))$ in sequential limit as $(T,n\to\infty)_{\text{seq}}$. The following theorem shows that these asymptotics also hold for joint limits.

THEOREM 9: *Under Assumptions* 6, 8–11 *we have*

$$\sqrt{n}\,T\left(\hat{\beta}_{PFM}-\beta\right)\Rightarrow N\left(0,2(\Omega_{xx}^{-1}\otimes\Omega_{e.x})\right)$$

as $(n,T\to\infty)$ *with* $n/T\to 0$.

REMARKS: (a) The pooled FM estimator $\hat{\beta}_{PFM}$ is $\sqrt{n}\,T$ consistent and has a normal limit distribution.

(b) When $\Lambda_{ex}=0$, observe that $\hat{\beta}_{PFM}$ is more efficient than $\hat{\beta}_{n,T}$ because $\Omega_{e.x}<\Omega_{ee}$. The efficiency gain in $\hat{\beta}_{PFM}$ is obtained from the endogeneity correction that adjusts $Y_{i,t}$ in the fully modified estimator. This effectively reduces the long-run variance of the noise in the panel cointegrating equation.

(c) Asymptotic χ^2 tests follow from Theorem 9 in the usual way. A consistent estimate of the covariance matrix, $2(\hat{\Omega}_{xx}^{-1}\otimes\hat{\Omega}_{e.x})$, can be constructed from $\hat{\Omega}$ in (5.13) by defining $\hat{\Omega}_{e.x}=\hat{\Omega}_{ee}-\hat{\Omega}_{ex}\hat{\Omega}_{xx}^{-1}\hat{\Omega}_{xe}$. A Wald test of $\mathbb{H}_0$: $\varphi(\beta)=0$, where $\varphi(\cdot)$ is a p-vector of smooth functions such that $\partial\varphi/\partial\beta'$ has full rank p, can then be formulated in the usual way as

$$W_{\varphi}=nT^2\varphi\left(\hat{\beta}_{PFM}\right)'\hat{V}_{\varphi}^{-1}\varphi\left(\hat{\beta}_{PFM}\right), \tag{5.17}$$

where

$$\hat{V}_{\varphi}=\left(\partial\varphi\left(\hat{\beta}_{PFM}\right)/\partial\beta'\right)\left[2\,\hat{\Omega}_{xx}^{-1}\otimes\hat{\Omega}_{e.x}\right]\left(\partial\varphi\left(\hat{\beta}_{PFM}\right)'/\partial\beta\right).$$

(d) As in Remark (e) following Theorem 6, it may be of interest to generalize model 5.11 to allow for subgroups of the population in which the regression coefficient is the same. In effect, we may replace model (5.11) with

$$Y_{i,t}\overset{\text{a.s.}}{=}\beta_{\mu}X_{i,t}+E_{i,t} \qquad \text{with} \qquad \begin{cases}\beta_{\mu}=\beta_a & \text{for } i\in I_a,\\ \beta_{\mu}=\beta_b & \text{for } i\in I_b,\end{cases} \tag{5.18}$$

$$X_{i,t}=X_{i,t-1}+U_{x_i,t}.$$

It is then possible to test hypotheses about the vectors β_a and β_a in the generalized model (5.18). For example, to test $\mathbb{H}_0$: $\beta_a=\beta_b$, letting $n_a=\#(I_a)$ and $n_b=\#(I_b)$, respectively, and assuming that $n_b/n_a\to\kappa\to\infty$ as $n_a,n_b\to\infty$, we may construct pooled FM regression coefficients $\hat{\beta}_{a,PFM},\hat{\beta}_{b,PFM}$ in each class and then the Wald statistic

$$W_{a-b,PFM}=n_bT^2\Big\{\text{vec}\left(\hat{\beta}_{a,PFM}-\hat{\beta}_{b,PFM}\right)' \times\hat{V}_{a-b,PFM}^{-1}\,\text{vec}\left(\hat{\beta}_{a,PFM}-\hat{\beta}_{b,PFM}\right)\Big\}.$$

Here, $\hat{V}_{a-b,\,PFM} = \kappa \hat{V}_{a,\,PFM} + \hat{V}_{b,\,PFM}$, $\hat{V}_{\mu,\,PFM} = 2\,\hat{\Omega}_{\mu,\,xx}^{-1} \otimes \hat{\Omega}_{\mu,\,e.x}$, and $\hat{\Omega}_{\mu,\,xx}$, $\hat{\Omega}_{\mu,\,e.x}$ are the respective estimates of the long-run conditional covariance matrices of the regressors and the fully modified error processes in the classes I_μ with $n_\mu = \#(I_\mu)$ where $\mu \in \{a, b\}$. As in the earlier case of heterogeneous cointegration, this leads to an asymptotic χ^2 test based on the null distribution $W_{a-b,\,PFM} \Rightarrow \chi^2_{m_y m_x}$, which follows in a manner analogous to that of Theorem 7.

5.3. *Near-Homogeneous Panels*

The homogeneous panel model (5.11) discussed above is somewhat unrealistic because it assumed that each individual has exactly the same cointegrating relation. Here we study briefly a panel cointegration model with nearly homogeneous cointegrating vectors of the form

$$\beta_i = \beta + \frac{\theta_i}{\sqrt{n}\,T}, \tag{5.19}$$

where the sequence of $(m_y \times m_x)$ random matrices θ_i is iid across i with mean θ and finite variance.

ASSUMPTION 12: *θ_i is independent of $(E_{i,t}, U_{x_i,t})$ for all i and t.*

We again consider the pooled FM estimator $\hat{\beta}_{PFM}$ given in (5.16) and the limit theory follows in Theorem 9 above.

THEOREM 10: *Suppose there exists near-homogeneous panel cointegration of the form* (5.19). *Let Assumptions* 9–12 *hold. Then, as* $(n, T \to \infty)$ *with* $n/T \to 0$

$$\sqrt{n}\,T\big(\hat{\beta}_{PFM} - \beta\big) \Rightarrow N\big(\theta, 2(\Omega_{xx}^{-1} \otimes \Omega_{e.x})\big). \tag{5.20}$$

Theorem 5.20 is useful in calculating the asymptotic local power of the test statistic for the null hypothesis

$$H_0\colon \beta_i = \beta_0 \qquad \forall i. \tag{5.21}$$

According to remark (c) of the previous subsection, the Wald statistic in (5.17) for the null hypothesis in (5.21) is W_φ with $\varphi(\beta) = \mathrm{vec}(\beta - \beta_0)$ and its limit distribution is $\chi^2_{m_y m_x}$. A sequence of local alternatives to the null (5.21) can be formulated as

$$H_{LA}\colon \beta_i = \beta_0 + \frac{\theta_i}{\sqrt{n}\,T}, \tag{5.22}$$

where the θ_i are iid across i with mean $\theta \neq 0$, have finite variance and satisfy Assumption 12. In this case, under the local alternative hypothesis (5.22) and the assumptions of Theorem 5.20, the Wald statistic W_φ has an asymptotic noncentral chi-square distribution as $(n, T \to \infty)$ with $n/T \to 0$, i.e.,

$$W_\varphi \Rightarrow \chi^2_{m_y m_x}(\lambda),$$

where the noncentrality parameter is $\lambda = \mathrm{vec}(\theta)' V_\varphi^{-1} \mathrm{vec}(\theta)/2$.

6. MODELS WITH INDIVIDUAL EFFECTS

Much of the preceding asymptotic theory can be extended in a straightforward way to panel models with individual specific effects and time trends. We illustrate what is involved in these extensions by taking the case of primary importance where the panel regression equation involves individual special effects. To motivate the analysis, consider the following model of heterogeneous panel cointegration in place of (5.2):

$$(6.1) \qquad Y_{i,t} \overset{\text{a.s.}}{=} \gamma_i + \beta_i X_{i,t} + E_{i,t}, \qquad X_{i,t} = X_{i,t-1} + U_{x_i,t}.$$

Here, the γ_i are individual effects in the cointegrating equation. They could be fixed or random effects. We can also allow for individual effects in the equation for $X_{i,t}$ in (6.1). In that case, the $X_{i,t}$ have individual deterministic trends as well as stochastic trends and in what follows we would proceed using detrended rather than demeaned data in the pooled panel regression, with some associated change in the final formulae.

The individual effect in (6.1) can be eliminated in the usual way by removing individual specific means, i.e., $\overline{Y}_{i,\cdot} = (1/T)\sum_{t=1}^T Y_{i,t}$ and $\overline{X}_{i,\cdot} = (1/T)\sum_{t=1}^T X_{i,t}$. Then pooled panel regression leads to the estimator

$$\tilde{\beta}_{n,T} = \left(\sum_{i=1}^n \sum_{t=1}^T \tilde{Y}_{i,t} \tilde{X}'_{i,t} \right) \left(\sum_{i=1}^n \sum_{t=1}^T \tilde{X}_{i,t} \tilde{X}'_{i,t} \right)^{-1},$$

where $\tilde{Y}_{i,t} = Y_{i,t} - \overline{Y}_{i,\cdot}$, and $\tilde{X}_{i,t} = X_{i,t} - \overline{X}_{i,\cdot}$.

As in our earlier theory, some quick asymptotic results for $\tilde{\beta}_{n,T}$ can be obtained using sequential limits. First consider the case where there is no cointegration and the true data generating mechanism is (2.1), even though it is model (6.1) that is estimated. Define the demeaned limiting process $\tilde{M}_i(r) = (\tilde{M}'_{y_i}(r), \tilde{M}'_{x_i}(r))' = M_i(r) - \int M_i(s)\, ds$. According to (2.6) and the continuous mapping theorem, under Assumptions 1–3, the pooled estimator $\tilde{\beta}_{n,T}$ has the following limit distribution as $T \to \infty$ for any fixed n:

$$\tilde{\beta}_{n,T} \Rightarrow \left(\frac{1}{n} \sum_{i=1}^n \int \tilde{M}_{y_i} \tilde{M}'_{x_i} \right) \left(\frac{1}{n} \sum_{i=1}^n \int \tilde{M}_{x_i} \tilde{M}'_{x_i} \right)^{-1}.$$

A simple calculation shows that $E(\int \tilde{M}_i \tilde{M}'_i) = E(\int M_i M'_i) - E(\int M_i \int M'_i) = \frac{1}{6}\Omega$. Thus, applying the strong law of large numbers to $(1/n)\sum_{i=1}^n \int \tilde{M}_{y_i} \tilde{M}'_{x_i}$ and

$(1/n)\Sigma_{i=1}^{n}\tilde{M}_{x_i}\tilde{M}_{x_i}'$, we get $(1/n)\Sigma_{i=1}^{n}\int\tilde{M}_{y_i}\tilde{M}_{x_i}' \to_{\text{a.s.}} \frac{1}{6}\Omega_{yx}$, and $(1/n)\Sigma_{i=1}^{n}\int\tilde{M}_{x_i}\tilde{M}_{x_i}' \to_{\text{a.s.}} \frac{1}{6}\Omega_{xx}$. It follows that $\tilde{\beta}_{n,T} \to_p \beta = \Omega_{yx}\Omega_{xx}^{-1}$ as $(T, n\to\infty)_{\text{seq}}$.

The asymptotic normality of $\tilde{\beta}_{n,T}$ follows by arguments analogous to those of Section 4. Rescaling the centered estimator $(\tilde{\beta}_{n,T}-\beta)$ by $\sqrt{n}$ and letting $T\to\infty$ for fixed n, we have

$$\sqrt{n}\left(\tilde{\beta}_{n,T}-\beta\right) \Rightarrow \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\int\tilde{M}_{y_i}\tilde{M}_{x_i}'-\beta\int\tilde{M}_{x_i}\tilde{M}_{x_i}'\right)\left(\frac{1}{n}\sum_{i=1}^{n}\int\tilde{M}_{x_i}\tilde{M}_{x_i}'\right)^{-1}.$$

Note that $E(\int\tilde{M}_{y_i}\tilde{M}_{x_i}'-\beta\int\tilde{M}_{x_i}\tilde{M}_{x_i}')=0$, so demeaning the data does not affect the asymptotic centering. After some lengthy calculations, we find the variance matrix

$$\begin{aligned} &E\left(\operatorname{vec}\left(\int\tilde{M}_{y_i}\tilde{M}_{x_i}'-\beta\int\tilde{M}_{x_i}\tilde{M}_{x_i}'\right)\operatorname{vec}\left(\int\tilde{M}_{y_i}\tilde{M}_{x_i}'-\beta\int\tilde{M}_{x_i}\tilde{M}_{x_i}'\right)'\right)\\ &\quad= \tfrac{1}{90}E\left(\Omega_{x_ix_i}\otimes(\Omega_{y_iy_i}-\beta\Omega_{x_iy_i}-\Omega_{y_ix_i}\beta'+\beta\Omega_{x_ix_i}\beta')\right)\\ &\qquad+\tfrac{1}{90}E\left((\Omega_{x_iy_i}-\Omega_{x_ix_i}\beta')\otimes(\Omega_{y_ix_i}-\beta\Omega_{x_ix_i})K_{m_ym_x}\right)\\ &\qquad+\tfrac{1}{36}E\left(\operatorname{vec}(\Omega_{y_ix_i}-\beta\Omega_{x_ix_i})\left(\operatorname{vec}(\Omega_{y_ix_i}-\beta\Omega_{x_ix_i})\right)'\right)\\ &\quad=\Theta_f, \qquad \text{say.}\end{aligned}$$

Note that this covariance matrix differs in the coefficients of its components from the earlier matrix Θ given in (4.6) for the case where there is no demeaning to remove possible individual effects. As is clear from the formulae for these two cases (see (6.2) below), $\Theta_f<\Theta$, so one effect of demeaning is to reduce time series variability.

Applying the multivariate Lindeberg-Levy Theorem to $(1/\sqrt{n})\Sigma_{i=1}^{n}(\int\tilde{M}_{y_i}\tilde{M}_{x_i}'-\beta\int\tilde{M}_{x_i}\tilde{M}_{x_i}')$ and combining this with the limit of $((1/n)\Sigma_{i\text{-}1}^{n}\int\tilde{M}_{x_i}\tilde{M}_{x_i}')^{-1}$ as $n\to\infty$, we have

$$\begin{aligned}&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\int\tilde{M}_{y_i}\tilde{M}_{x_i}'-\beta\int\tilde{M}_{x_i}\tilde{M}_{x_i}'\right)\left(\frac{1}{n}\sum_{i=1}^{n}\int\tilde{M}_{x_i}\tilde{M}_{x_i}'\right)^{-1}\\ &\qquad\Rightarrow N\left(0, 36\left(\Omega_{xx}^{-1}\otimes I_{m_y}\right)\Theta_f\left(\Omega_{xx}^{-1}\otimes I_{m_y}\right)\right).\end{aligned}$$

Hence, as $(T, n\to\infty)_{\text{seq}}$ we have

$$\sqrt{n}\left(\tilde{\beta}_{n,T}-\beta\right)\Rightarrow N\left(0, 36\left(\Omega_{xx}^{-1}\otimes I_{m_y}\right)\Theta_f\left(\Omega_{xx}^{-1}\otimes I_{m_y}\right)\right).$$

These sequential limit results can be extended to joint limit results, just as in the proof of Theorem 4, and we merely state the final results here.

THEOREM 11: *Suppose Assumptions* 1, 2, *and* 3 *hold and the data generating mechanism is* (2.1). *Then*:

(a) *as* $(n, T \to \infty)$, *we have* $\tilde{\beta}_{n,T} \to_p \beta$;

(b) *if* $(n, T \to \infty)$ *and* $n/T \to 0$,

$$\sqrt{n}\left(\tilde{\beta}_{n,T} - \beta\right) \Rightarrow N\left(0, 36\left(\Omega_{xx}^{-1} \otimes I_{m_y}\right)\Theta_f\left(\Omega_{xx}^{-1} \otimes I_{m_y}\right)\right).$$

REMARKS: (a) Comparing the limit variance of $\tilde{\beta}_{n,T}$ in Theorem 11 to that of $\tilde{\beta}_{n,T}$ in Theorem 4, we find that $\tilde{\beta}_{n,T}$ has the smaller asymptotic covariance. In fact, it is apparent from the formulae that

$$(6.2)\qquad 4\Theta - 36\Theta_f = \tfrac{4}{15}\left(\Omega_{x_ix_i} \otimes (\Omega_{y_iy_i} - \beta\Omega_{x_iy_i} - \Omega_{y_ix_i}\beta' + \beta\Omega_{x_ix_i}\beta')\right)$$

$$+ \tfrac{4}{15}E\left((\Omega_{x_iy_i} - \Omega_{x_ix_i}\beta') \otimes (\Omega_{y_ix_i} - \beta\Omega_{x_ix_i})K_{m_ym_x}\right)$$

$$= \tfrac{4}{15}E\Big\{\left(C_{x_i}(1) \otimes \left(C_{y_i}(1) - \beta C_{x_i}(1)\right)\right)(I_{m^2} + K_m)$$

$$\times\left(C_{x_i}(1) \otimes \left(C_{y_i}(1) - \beta C_{x_i}(1)\right)\right)'\Big\}$$

$$(6.3)\qquad > 0.$$

As remarked above, this reduction in variance occurs because demeaning the data by removing individual effects reduces time series variability. Similar effects occur when higher order time trends are removed from the data in the construction of pooled panel estimators.

(b) In the heterogenous panel cointegration case, the data are generated by (6.1). The individual effects γ_i can now be consistently estimated by time series regression on (6.1) leading to $\hat{\gamma}_i = \overline{Y}_{i,.} - \tilde{\beta}_i \overline{X}_{i,.}$ and $\tilde{\beta}_i = (\sum_{t=1}^T \tilde{Y}_{i,t}\tilde{X}'_{i,t})(\sum_{t=1}^T \tilde{X}_{i,t}\tilde{X}'_{i,t})^{-1}$. These least squares estimates and their fully modified variants have asymptotic properties that are well known (Phillips and Hansen (1990)). Following the same line of argument as in Section 5.1, the pooled panel estimator $\tilde{\beta}_{n,T}$ can be shown to have the same limit distribution as that given in Theorem 11 for the spurious regression case, although the long run covariance matrices Ω_i are now singular, just as in Section 5.1. Under the assumptions of Theorem 5, the asymptotic theory holds for joint limits as $(n, T \to \infty)$ and $n/T \to 0$, as well as sequential limits. Again $\tilde{\beta}_{n,T}$ estimates the long run average coefficient $\beta = \Omega_{yx}\Omega_{xx}^{-1}$. Wald tests like those discussed in Section 5 can now be constructed with obvious modifications to the estimated covariance matrix formulae that allow for elimination of the individual effects by demeaning.

(c) In the homogeneous panel cointegration case, the data are generated by (6.1) with $\beta_i = \beta = \Omega_{yx}\Omega_{xx}^{-1}$ a.s. We can eliminate individual effects by removing individual specific means as above, and may proceed with FM estimation as in Section 5.2. The data are now corrected according to the formula $\tilde{Y}^+_{i,t} = \tilde{Y}_{i,t} -$

$\hat{\Omega}_{ex}\hat{\Omega}_{xx}^{-1}\,\Delta\tilde{X}_{i,t}$ rather than as in (5.14). The pooled FM estimator in this case is given by

$$\tilde{\beta}_{PFM} = \left(\sum_{i=1}^{n} \sum_{t=1}^{T} \tilde{Y}_{i,t}^{+} \tilde{X}_{i,t}' - nT\tilde{\Lambda}_{ex}^{+} \right) \left(\sum_{i=1}^{n} \sum_{t=1}^{T} \tilde{X}_{i,t} \tilde{X}_{i,t}' \right)^{-1}.$$

Under the same assumptions as Theorem 9, we find that $\sqrt{n}\,T(\tilde{\beta}_{PFM} - \beta) \Rightarrow N(0, 6(\Omega_{xx}^{-1} \otimes \Omega_{e.x}))$ as $(n, T \to \infty)$ with $n/T \to 0$. Note that in this case, the effect of eliminating individual specific means is to increase the limit variance matrix in comparison with Theorem 9. Wald tests can be constructed as described in Section 5.2 with obvious modifications for the use of demeaned data, and a noncentral limit theory follows as in Section 5.3.

7. CONCLUSION

This paper has developed a linear regression limit theory for nonstationary panel data with large numbers of cross section and time series observations. A central result is the existence of interesting long-run relations between two integrated panel vectors where there is no individual time series cointegration or where there are heterogeneous cointegrating relations. The new relations are characterized as long-run average relationships over the cross section and are parameterized in terms of the matrix regression coefficient, $\Omega_{yx}\Omega_{xx}^{-1}$, of the cross section long-run average covariance matrix, Ω. They are analogous to population regression coefficients in conventional cross section regression of iid variates. The limit theory can be used to construct tests of hypotheses about the long-run average regression coefficients and to compare these coefficients in subgroups of the cross section population. These tests are given explicitly for the two cases of heterogeneous panel cointegration and homogeneous panel cointegration, which seem to be the important cases for empirical applications. The local asymptotic power function for these tests is also derived.

The limit theory developed in this paper is designed for two dimensional arrays where both time series and cross section sample sizes pass to infinity. It allows for both sequential limits as $T \to \infty$ and $n \to \infty$ in that sequence, and joint limits where $T, n \to \infty$ jointly. As the proofs in the Appendices demonstrate, convergence for joint limit is more difficult to obtain. However, apart from some stricter moment and summability conditions, the only additional requirement we use in the development of this theory is the rate condition that $n/T \to 0$. This condition indicates that the limit theory given herein is likely to be most useful in cases where T is large and n is moderately large. The usefulness of this asymptotic theory in describing finite sample behavior in panel regressions now needs to be systematically explored in simulation experiments.

An important assumption that is common in panel data work and is used here in deriving asymptotics is cross section independence. For many nonstationary panel data applications, this independence condition is restrictive and it is an important limitation of our theory. For instance, multi-country GDP series, exchange rates, and financial assets prices all involve cross section dependence arising from global shocks and complicated interdependencies among the variables. As is apparent from our approach, certain strong laws and central limit results will continue to apply when the cross sectional dependence is of the weak memory variety, but in this case the limit variance matrices will change according to the dependence. More significantly, when there are strong correlations in a cross section (as there will be in the face of global shocks) we can expect failures in the strong laws and central limit theory arising from the nonergodicity. However, even in this event, theorems like the ergodic theorem will still apply but the limits will be random and measurable with respect to the invariant algebra generated by the global shocks.

In the present case and, indeed, quite commonly in panel data theory, cross section independence is assumed in part because of the difficulties of characterizing and modeling cross section dependence. In general, finding a natural ordering for cross section indices in economic data is not easy, and this has been a serious obstacle in the development of a satisfactory approach. While some recent research has attempted to resolve the difficulty by employing a framework for spatial data based on the economic distance between individuals (e.g. Conley (1997)), the successful simultaneous modeling of both cross section dependence and time series dependence remains a challenging problem and is a major area for future research in multi-index asymptotics of the type considered here.

Cowles Foundation for Research in Economics, Yale University, Box 208281, New Haven, CT 06520-8281, U.S.A., and University of Auckland; peter.phillips@yale.edu

and

Dept. of Economics, University of California, Santa Barbara, CA 93106, U.S.A.; rmoon@econ.ucsb.edu

Manuscript received April, 1997; final revision received September, 1998.

APPENDICES

APPENDIX A: PRELIMINARY LEMMAS AND PROOFS

We start with some lemmas that are useful in following arguments. The results are straightforward and proofs are omitted here, but are available in PM[b].

LEMMA 9: (a) *For any $p \geq 1$ and any $m \times n$ matrix A, there exists a constant $M > 0$ such that*

$$\|A\|^p \leq M \sum_{i=1}^{m} \sum_{j=1}^{n} |a_{i,j}|^p, \tag{8.1}$$

where $a_{i,j}$ is the (i, j)th element of A.

(b) *For any $m \times m$ matrix A*

$$(\mathrm{tr}(A))^2 \leq m\|A\|^2. \tag{8.2}$$

LEMMA 10: *Suppose that* $A\ (=\{a_{i,j}\}_{i,j})$ *and* $B\ (=\{b_{i,j}\}_{i,j})$ *are* $(m \times m)$ *matrices and* K_m *is the commutation matrix. Then,*

$$tr[(A \otimes B)K_m] \leq \|A\|\,\|B\|.$$

If A is symmetric, then

$$tr[(A \otimes A)K_m] = \|A\|^2.$$

LEMMA 11: (a) *Under Assumptions* 1 *and* 2

$$E\left\|\sum_{t=0}^{\infty} C_{i,t}\right\|^4 < \infty. \tag{8.3}$$

(b) *Under Assumptions* 4 *and* 5

$$E\left\|\sum_{t=0}^{\infty} tC_{i,t}\right\|^8 < \infty. \tag{8.4}$$

(c) *Under Assumptions* 4 *and* 5

$$E\left\|\sum_{t=0}^{\infty} C_{i,t}\right\|^{16} < \infty. \tag{8.5}$$

1. PROOF OF LEMMA 2: The panel BN decomposition

$$U_{i,t} = C_i(1)V_{i,t} + \tilde{U}_{i,t-1} - \tilde{U}_{i,t} \qquad \text{a.s.} \tag{8.6}$$

follows directly from Phillips and Solo (1992) provided $Y_i = \sum_{s=0}^{\infty} s^2 \|C_{i,s}\|^2 < \infty$ a.s. This condition holds if $E(Y_i) < \infty$, which holds by Lemma 1(a). *Q.E.D.*

2. PROOF OF LEMMA 1: See PM[b]. *Q.E.D.*

3. PROOF OF LEMMA 2: It is enough to show that

$$\left|\frac{1}{\sqrt{T}}\tilde{U}_{a,i,0}\right| \xrightarrow{p} 0 \quad \text{and}$$

$$\sup_r \left|\frac{1}{\sqrt{T}}\tilde{U}_{a,i,[Tr]}\right| \xrightarrow{p} 0 \quad \text{as } T \to \infty \text{ for all } a, i.$$

But, these follow because $\tilde{U}_{a,i,t}$ is strictly stationary in t and square integrable by Lemma 1, so that the results hold by the same argument as that given in Phillips and Solo (1992, p. 978). The functional law follows directly. *Q.E.D.*

4. PROOF OF LEMMA 4: Substituting $M_i = C_i(1)W_i$, we have

$$E\left\|\int M_i M_i'\right\|^2 = E\left\|\text{vec}\int M_i M_i'\right\|^2 = E\left\|(C_i(1) \otimes C_i(1))\text{vec}\int W_i W_i'\right\|^2$$

$$\leq E\|C_i(1) \otimes C_i(1)\|^2 E\left\|\text{vec}\int W_i W_i'\right\|^2,$$

where the last inequality holds because $\|AB\| \le \|A\|\,\|B\|$ and because $C_i(1)$ is independent of W_i. We know $E\|\mathrm{vec}\int W_i W_i'\|^2 < \infty$ and $E\|C_i(1) \otimes C_i(1)\|^2 = E\|C_i(1)\|^4 < \infty$ by Lemma 1. Therefore, $E\|\int M_i M_i'\|^2 < \infty$, as required. *Q.E.D.*

Appendix B: Proofs for Section 3—Multidimensional Limit Theory

1. Construction of Random Vectors Y_i in (3.3) to Exist on the Same Probability Space: According to Skorohod's Theorem in $\mathbb{R}^m$, Theorem 29.6 in Billingsley (1986),[8] we can construct a probability space $(\Omega_i^*, \boldsymbol{F}_i^*, P_i^*)$ where there exist random vectors $Y_{i,T}^*$ and Y_i^* such that $Y_{i,T} \equiv Y_{i,T}^*$, $Y_i \equiv Y_i^*$, and $Y_{i,T}^* \to_{\text{a.s.}} Y_i^*$ as $T \to \infty$ for all i. Also, we can choose independent Y_i^* because the $Y_{i,T}$ are independent across i for all T. Now we define $\Omega^* = \prod_{i=1}^{\infty} \Omega_i^*$, the Cartesian product of Ω_i^*, and let π_i be the natural projection of Ω^* onto Ω_i^* for each i. Let $\boldsymbol{F}^*$ be the smallest σ-field containing all the sets $\pi_i^{-1}(F)$ for all i and $F \in \boldsymbol{F}_i^*$. Define $\boldsymbol{R}$ to be the collection of all finite dimensional rectangles, $\prod_{i=1}^{\infty} F_i$ where $F_i \in \boldsymbol{F}_i^*$ for all i and $F_i = \Omega_i^*$, except for at most finite many values of i. Now define $P^*(\prod_{i=1}^{\infty} F_i) = \prod_{i=1}^{\infty} P_i^*(F_i)$. Then, by Theorem 8.2.2 (p. 201) in Dudley (1989), P^* on $\boldsymbol{R}$ extends uniquely to a probability measure on $\boldsymbol{F}^*$. Let $\tilde{Y}_i(\omega) = Y_i^*(\pi_i^{-1}(\omega))$ for all $\omega \in \Omega^*$. By the way of their construction, the $\tilde{Y}_i(\omega)$ are random vectors on the probability space $(\Omega^*, \boldsymbol{F}^*, P^*)$ and $\tilde{Y}_i \equiv Y_i^* \equiv Y_i$. Choose Y_i in (3.3) to be $\tilde{Y}_i$ and we have the desired result. Q.*E.D.*

2. Proof of Lemma 5: We prove part (b). Then part (a) holds by the same principle. Suppose that $f \in \boldsymbol{C}$ is given. From $X_{n,T} \Rightarrow X$ as $n, T \to \infty$, for any given $\varepsilon > 0$, we can chosen n_0 and T_1 such that whenever $n \ge n_0$ and $T \ge T_1$, the following inequality holds:

$$|Ef(X_{n,T}) - Ef(X)| < \frac{\varepsilon}{2}. \tag{8.7}$$

From $X_{n,T} \Rightarrow X_n$ as $T \to \infty$ $\forall_n$, we can choose T_2 depending on n and ε such that

$$|Ef(X_{n,T}) - Ef(X_n)| < \frac{\varepsilon}{2} \qquad \text{if } T \ge T_2. \tag{8.8}$$

For each $n \ge n_0$ choose $T_2(n, \varepsilon)$, and choose a fixed T_0 greater than both T_1 and T_2. Then both (8.7) and (8.8) hold and therefore

$$|Ef(X_n) - Ef(X)| < \varepsilon \qquad \text{if } n \ge n_0$$

and $X_{n,T} \Rightarrow X$ sequentially as $(T, n \to \infty)_{\text{seq}}$. *Q.E.D.*

3. Proof of Lemma 6: We show part (b). Part (a) can be established by similar arguments.

Suppose that $f \in \boldsymbol{C}$ is given. Assume that (3.9) holds. From (3.9) and $X_n \Rightarrow X$ as $n \to \infty$, for any given $\varepsilon > 0$, we can choose n_0 and T_0 such that whenever $n \ge n_0$ and $T \ge T_0$, we have

$$\sup_{n \ge n_0, T \ge T_0} |Ef(X_{n,T}) - Ef(X_n)| < \frac{\varepsilon}{2},$$

and

$$|Ef(X_n) - Ef(X)| < \frac{\varepsilon}{2}.$$

[8] For Skorohod's theorem on function spaces refer to the representation theorem in Pollard (1984) or Theorem 4 on p. 47 in Shorack and Wellner (1986).

Thus, if $n \geq n_0$ and $T \geq T_0$,

$$|Ef(X_{n,T}) - Ef(X)| \leq \sup_{n \geq n_0, T \geq T_0} |Ef(X_{n,T}) - Ef(X_n)| + |Ef(X_n) - Ef(X)| < \varepsilon.$$

Hence, $X_{n,T} \Rightarrow X$ as $(n, T \to \infty)$.

Now assume that $X_{n,T} \Rightarrow X$ and $X_n \Rightarrow X$ as $(n, T \to \infty)$. The necessity of the condition follows because

$$\limsup_{n,T} |Ef(X_{n,T}) - Ef(X_n)|$$
$$\leq \limsup_{n,T} |Ef(X_{n,T}) - Ef(X)| + \limsup_{n} |Ef(X_n) - Ef(X)| = 0. \qquad Q.E.D.$$

Before starting the proof of Theorem 1 we give the following lemma.

LEMMA 12: *Suppose $Y_{i,T}$ are independent across i. Assume that $Y_{i,T} \Rightarrow Y_i$ as $T \to \infty$ for all i. Then,* $\limsup_{n\to\infty}(1/n)\sum_{i=1}^n E|Y_{i,T}| < \infty$ *implies* $\limsup_{n\to\infty}(1/n)\sum_{i=1}^n E|Y_i| < \infty$.

PROOF: Note that, since $|Y_{i,T}| \Rightarrow |Y_i|$ as $T \to \infty$ by the continuous mapping theorem, it follows that $E|Y_i| \leq \liminf_T E|Y_{i,T}|$ (see Theorem 5.3 in Billingsley (1968)). Thus,

$$\limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^n E|Y_i| \leq \limsup_{n\to\infty} \liminf_{T\to\infty} \frac{1}{n}\sum_{i=1}^n E|Y_{i,T}|$$
$$\leq \limsup_{n,T\to\infty} \frac{1}{n}\sum_{i=1}^n E|Y_{i,T}| < \infty. \qquad Q.E.D.$$

4. PROOF OF THEOREM 1: Part (b) follows easily from Lemma 6 and part (a). In particular, from the assumptions of the theorem we know that $X_{n,T} = (1/n)\sum_{i=1}^n Y_{i,T} \Rightarrow X_n = (1/n)\sum_{i=1}^n Y_i$ as $T \to \infty$ for all n and $X_n = (1/n)\sum_{i=1}^n Y_i \to_p \tilde{\mu}_X = \lim_n (1/n)\sum_{i=1}^n EY_i$. Then, since condition (3.9) holds from part (a), the desired result $X_{n,T} \to_p \tilde{\mu}_X$ as $(n, T \to \infty)$ follows by Lemma 6.

Now, we prove part (a). First, we establish condition (3.9) in the scalar class. It is sufficient for condition (3.9) to restrict C to C^∞, the class of all the bounded, continuous real functions with bounded, continuous derivatives of all orders (see Theorem 7.1 in Billingsley (1968) or Theorem 12 in Pollard (1984)). Without loss of generality, let the functions be such that $|f^{(k)}(x)| \leq 1$, $\forall k$, where $f^{(k)}(x)$ denotes the kth derivative function of $f(x)$.

Before proceeding, we need to ensure that the probability space on which the variates are defined is large enough to permit the arguments that follow. Limits such as $X_{n,T} = (1/n)\sum_{i=1}^n Y_{i,T} \Rightarrow X_n = (1/n)\sum_{i=1}^n Y_i$ as $T \to \infty$ involve the joint distributions of the random vectors $(Y_{1,T}, \ldots, Y_{n,T})'$ and $(Y_1, \ldots, Y_n)'$, not any properties of the probability space on which they are defined. However, we need to ensure that we can relate these variates on the same space. This can be accomplished by passing to a new probability space, using Skorohod's Theorem in $\mathbb{R}^m$ (e.g., Theorem 29.6 in Billingsley (1986)), in which they are defined new random variables $(\tilde{Y}_{1,T}, \ldots, \tilde{Y}_{n,}, \tilde{Y}_1, \ldots, \tilde{Y}_n)'$ such that $\tilde{Y}_{i,T} \equiv Y_{i,T}$ and $\tilde{Y}_i \equiv Y_i$ for all i and the $2n$ random variables $\tilde{Y}_{1,T}, \ldots, \tilde{Y}_{n,T}, \tilde{Y}_1, \ldots, \tilde{Y}_n$ are independent. Without loss of generality, we can assume that $\tilde{Y}_{i,T} = Y_{i,T}$ and $\tilde{Y}_i = Y_i$ for all i and T.[9]

[9]As in Appendix B(1) above, we can construct an infinite dimensional probability space where the two independent random vectors $(Y_{1,T}, \ldots, Y_{n,T})'$ and $(Y_1, \ldots, Y_n)'$ coexist. However, the argument given is enough for the proof that follows.

Now we can define the quantities $\zeta_{k,n,T} = \sum_{k>i\geq 1} Y_{i,T} + \sum_{k<i\leq n} Y_i$, for $1 \leq k \leq n$, all on the same probability space. By virtue of the definitions of $X_{n,T}$, X_n, and $\zeta_{k,n,T}$, we have

$$f\left(\frac{1}{n}\sum_{i=1}^{n} Y_{i,T}\right) - f\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right) = f\left(\frac{1}{n}(\zeta_{n,n,T} + Y_{n,T})\right) - f\left(\frac{1}{n}(\zeta_{1,n,T} + Y_1)\right)$$

$$= \sum_{k=1}^{n}\left\{ f\left(\frac{1}{n}(\zeta_{k,n,T} + Y_{k,T})\right) - f\left(\frac{1}{n}(\zeta_{k,n,T} + Y_k)\right)\right\}.$$

It follows that

$$(8.9) \qquad \limsup_{n,T\to\infty} |Ef(X_{n,T}) - Ef(X_n)|$$

$$= \limsup_{n,T\to\infty}\left| Ef\left(\frac{1}{n}\sum_{i=1}^{n} Y_{i,T}\right) - Ef\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right)\right|$$

$$= \limsup_{n,T\to\infty}\left|\sum_{k=1}^{n} E\left\{\begin{array}{l} f\left(\frac{1}{n}(\zeta_{k,n,T} + Y_{k,T})\right) - f\left(\frac{1}{n}\zeta_{k,n,T}\right) \\ +f\left(\frac{1}{n}\zeta_{k,n,T}\right) - f\left(\frac{1}{n}(\zeta_{k,n,T} + Y_k)\right)\end{array}\right\}\right|.$$

Let $g(h) = \sup_x |f(x+h) - f(x) - f'(x)h|$. Take $x = \zeta_{k,n,T}/n$ and $h = Y_{k,T}/n$ in the case of $f((1/n)(\zeta_{k,n,T} + Y_{k,T})) - f((1/n)\zeta_{k,n,T})$, and take $x = \zeta_{k,n,T}/n$ and $h = Y_k/n$ in the case of $f((1/n)(\zeta_{k,n,T} + Y_i)) - f((1/n)\zeta_{k,n,T})$. By the triangle inequality, it follows that (8.9) is bounded above by

$$(8.10) \qquad \limsup_{n,T\to\infty}\left|\sum_{i=1}^{n} E\left\{ f'\left(\frac{\zeta_{i,n,T}}{n}\right)\left(\frac{Y_{i,T}}{n} - \frac{Y_i}{n}\right)\right\}\right|$$

$$+ \limsup_{n,T\to\infty}\sum_{i=1}^{n} Eg\left(\frac{Y_{i,T}}{n}\right) + \limsup_{n,T\to\infty}\sum_{i=1}^{n} Eg\left(\frac{Y_i}{n}\right).$$

By the triangle inequality, the first term in (8.10) is less than

$$\limsup_{n,T\to\infty}\sum_{i=1}^{n}\left| E\left\{ f'\left(\frac{\zeta_{i,n,T}}{n}\right)\left(\frac{Y_{i,T}}{n} - \frac{Y_i}{n}\right)\right\}\right|$$

$$= \limsup_{n,T\to\infty}\sum_{i=1}^{n}\left| Ef'\left(\frac{\zeta_{i,n,T}}{n}\right) E\left(\frac{Y_{i,T}}{n} - \frac{Y_i}{n}\right)\right|$$

$$\leq \limsup_{n,T\to\infty}\sum_{i=1}^{n}\left| E\left(\frac{Y_{i,T}}{n} - \frac{Y_i}{n}\right)\right| = 0.$$

The first line above uses the fact that $\zeta_{i,n,T}/n$, $Y_{i,T}/n$ and Y_i/n are independent, the inequality in the second line holds because $|f'| \leq 1$, and the third line follows directly from condition (ii).

For the second term in (8.10), note by the mean value theorem that $g(h) \le M_1 \min\{|h|, h^2\}$ for some constant M_1 which depends on f alone. Then, for any $\varepsilon > 0$

$$\begin{aligned}
&\limsup_{n,T\to\infty} \sum_{i=1}^{n} Eg\left(\frac{Y_{i,T}}{n}\right) \\
&\quad \le \limsup_{n,T\to\infty} \sum_{i=1}^{n} E\left[g\left(\frac{Y_{i,T}}{n}\right) 1\left\{\left|\frac{Y_{i,T}}{n}\right| \le \varepsilon\right\}\right] \\
&\qquad + \limsup_{n,T\to\infty} \sum_{i=1}^{n} E\left[g\left(\frac{Y_{i,T}}{n}\right) 1\left\{\left|\frac{Y_{i,T}}{n}\right| > \varepsilon\right\}\right] \\
&\quad \le \varepsilon M_1^2 \limsup_{n,T\to\infty} \sum_{i=1}^{n} E\left|\frac{Y_{i,T}}{n}\right| + M_1 \limsup_{n,T\to\infty} \sum_{i=1}^{n} E\left[\left|\frac{Y_{i,T}}{n}\right| 1\left\{\left|\frac{Y_{i,T}}{n}\right| > \varepsilon\right\}\right] = \varepsilon M_2,
\end{aligned}$$

where the first inequality holds by applying $g(h) \le M_1 h^2$ on $1\{|Y_{i,T}/n| \le \varepsilon\}$ and $g(h) \le M_1 |h|$ on $1\{|Y_{i,T}/n| > \varepsilon\}$ and the last inequality holds by conditions (i) and (iii) with $M_2 = M_1^2 \limsup_{n,T} \sum_{i=1}^{n} E|Y_{i,T}/n|$.

By Lemma 12, condition (i) implies

$$\limsup_{n,T} \frac{1}{n} \sum_{i=1}^{n} E|Y_i| < \infty,$$

and by condition (iv) we have

$$\limsup_{n,T} \frac{1}{n} \sum_{i=1}^{n} E|Y_i| 1\{|Y_i| > \varepsilon n\} < \infty.$$

Thus, applying the same arguments as those used for $\limsup_{n,T} \sum_{i=1}^{n} Eg(Y_{i,T}/n)$ to $\limsup_{n,T} \sum_{i=1}^{n} Eg(Y_i/n)$, we have $\limsup_{n,T} \sum_{i=1}^{n} Eg(Y_i/n) = 0$. It follows from (8.9) and (8.10) that condition (3.9) holds.

When the $Y_{i,T}$ are m-vectors, the Cramér-Wold device can be used. That is, using the above argument, we obtain $s' X_{n,T} \to_p s'\tilde{\mu}_X$ as $(n, T \to \infty)$ $\forall s \in \mathbb{R}^m$, and it follows that $X_{n,T} \to_p \tilde{\mu}_X$ as $n, T \to \infty$. *Q.E.D.*

5. PROOF OF COROLLARY 1: Define $X_{n,T} = (1/n)\sum_{i=1}^{n} Y_{i,T} = (1/n)\sum_{i=1}^{n} C_i Q_{i,T}$ and $X_n = (1/n)\sum_{i=1}^{n} Y_i = (1/n)\sum_{i=1}^{n} C_i Q_i$. Assume $\sup_i \|C_i\| > 0$, for if this does not hold, the result is trivial. We know that $X_{n,T} \Rightarrow X$ as $T \to \infty$ for all n by the conditions in the corollary. By assumption the $Q_{i,T}$ are uniformly integrable and $Q_{i,T} \Rightarrow Q_i$, so $E\|Q_i\| < \infty$. Also, $C = \lim_n (1/n)\sum_{i=1}^{n} C_i$ exists, so we have $X_n \to_p CE(Q_i)$ as $n \to \infty$. Hence, if we establish conditions (i)–(iv) of Theorem 1, then $X_{n,T} \to_p CE(Q_i)$ as $(n, T \to \infty)$.

By the uniform integrability of $\|Q_{i,T}\|$ and $\sup_i \|C_i\| < \infty$, we have

$$\limsup_{n,T} \frac{1}{n} \sum_{i=1}^{n} E\|Y_{i,T}\| \le \left(\sup_i \|C_i\|\right) \sup_T E\|Q_{i,T}\| < \infty,$$

verifying condition (i), and

$$\limsup_{n,T} \frac{1}{n} \sum_{i=1}^{n} \|EY_{i,T} - EY_i\| \le \left(\sup_i \|C_i\|\right) \limsup_T \|EQ_{i,T} - EQ_i\| = 0.$$

Condition (iii) is satisfied since

$$(8.11)\qquad \frac{1}{n}\sum_{i=1}^{n} E[\|Y_{i,T}\|1\{\|Y_{i,T}\|>n\varepsilon\}] \le \left(\sup_i\|C_i\|\right)\sup_T E\left[\|Q_{i,T}\|1\left\{\|Q_{i,T}\|>\frac{n\varepsilon}{\sup_i\|C_i\|}\right\}\right]$$

which converges to zero as $n\to\infty$, again by virtue of uniform integrability and $\sup_i\|C_i\|<\infty$.

Condition (iv) of Theorem 1 holds because

$$\frac{1}{n}\sum_{i=1}^{n} E\|Y_i\|1\{\|Y_i\|>n\varepsilon\} \le \left(\sup_i\|C_i\|\right)E\left[\|Q_i\|1\left\{\|Q_i\|>\frac{n\varepsilon}{\sup_i\|C_i\|}\right\}\right]\to 0$$

by $\sup_i\|C_i\|<\infty$ and dominated convergence since $E\|Q_i\|<\infty$. *Q.E.D.*

6. PROOF OF THEOREM 2: The proof follows that of Lindeberg's theorem given in Billingsley (1968, Theorem 7.2). The only change is that the additional index T appears in the component variates $\xi_{i,n,T}$ and limits are taken as $(n,T\to\infty)$. The fact that T passes to infinity with n is incidental to the main argument. For example, we still have

$$\frac{\Omega_{i,T}}{s_{n,T}^2} \le \varepsilon^2 + E[\xi_{i,n,T}^2 1\{|\xi_{i,n,T}|>\varepsilon\}]$$

and, as a consequence of the Lindeberg condition (3.20),

$$\max_{i\le n}\frac{\Omega_{i,T}}{s_{n,T}^2}\to 0$$

as $(n,T\to\infty)$. *Q.E.D.*

7. PROOF OF THEOREM 3: Define

$$\xi_{i,n,T} = \Omega_{n,T}^{-1/2}C_iQ_{i,T},$$

where $\Omega_{n,T}=\sum_{i=1}^n C_i\Sigma_T C_i'$. By the Cramér-Wold device, $\sum_{i=1}^n \xi_{i,n,T}\Rightarrow N(0,I_m)$ as $(n,T\to\infty)$, if $\forall t\in\mathbb{R}^m$ with $\|t\|=1$

$$(8.12)\qquad t'\sum_{i=1}^{n}\xi_{i,n,T}\Rightarrow N(0,1)\qquad \text{as}\qquad n,T\to\infty.$$

Then, by condition (iv) $(1/n)\sum_{i=1}^n Y_{i,T}\Rightarrow N(0,\Omega)$ as $(n,T\to\infty)$.

To establish (8.12), it is sufficient to verify condition (3.20). For given $\varepsilon>0$ and $t\in\mathbb{R}^m$ with $\|t\|=1$, we have

$$(8.13)\qquad t'\sum_{i=1}^{n}E[\xi_{i,n,T}\xi_{i,n,T}'1\{|t'\xi_{i,n,T}\xi_{i,n,T}'t|>\varepsilon\}]t$$

$$= t'\Omega_{n,T}^{-1/2}\sum_{i=1}^{n}E[C_iQ_{i,T}Q_{i,T}'C_i'1\{|t'\Omega_{n,T}^{-1/2}C_iQ_{i,T}Q_{i,T}'C_i'\Omega_{n,T}^{-1/2}t|>\varepsilon\}]\Omega_{n,T}^{-1/2}t.$$

Take the indicator function first. Note that

$$
\begin{aligned}
&1\{|t'\Omega_{n,T}^{-1/2} C_i Q_{i,T} Q_{i,T}' C_i' \Omega_{n,T}^{-1/2} t| > \varepsilon\} \\
&\le 1\left\{ \max_{\|t\|=1} |t'\Omega_{n,T}^{-1/2} C_i Q_{i,T} Q_{i,T}' C_i' \Omega_{n,T}^{-1/2} t| > \varepsilon \right\} \\
&= 1\{\lambda_{\max}(\Omega_{n,T}^{-1/2} C_i Q_{i,T} Q_{i,T}' C_i' \Omega_{n,T}^{-1/2}) > \varepsilon\} \\
&\le 1\left\{ \lambda_{\max}(\Omega_{n,T}^{-1}) \left(\max_{j \le n} \|C_j\|^2 \right) \|Q_{i,T}\|^2 > \varepsilon \right\} \\
&= 1\left\{ \|Q_{i,T}\|^2 > \varepsilon \frac{\lambda_{\min}(\Omega_{n,T})}{\max_{j \le n} \|C_j\|^2} \right\} \le 1\left\{ \|Q_{i,T}\|^2 > \varepsilon \frac{\sigma_T^2 \lambda_{\min}(\sum_{j=1}^n C_j C_j')}{\max_{j \le n}(\|C_j\|^2)} \right\}.
\end{aligned}
$$

Next, expression (8.13) is bounded above by

$$
(8.14) \qquad \max_{\|t\|=1} \left[t'\Omega_{n,T}^{-1/2} \sum_{i=1}^n E\left[C_i Q_{i,T} Q_{i,T}' C_i' 1\left\{ \|Q_{i,T}\|^2 > \varepsilon \frac{\sigma_T^2 \lambda_{\min}(\sum_{j=1}^n C_j C_j')}{\max_{j \le n} \|C_j\|^2} \right\} \right] \Omega_{n,T}^{-1/2} t \right]
$$

$$
\begin{aligned}
&\le \lambda_{\min}^{-1}(\Omega_{n,T}) \sum_{i=1}^n E\left[\|C_i Q_{i,T}\|^2 1\left\{ \|Q_{i,T}\|^2 1 > \varepsilon \frac{\sigma_T^2 \lambda_{\min}(\sum_{j=1}^n C_j C_j')}{\max_{j \le n}(\|C_j\|^2)} \right\} \right] \\
&\le \frac{\sum_{i=1}^n \|C_i\|^2}{\lambda_{\min}(\Omega_{n,T})} E\left[\|Q_{1,T}\|^2 1\left\{ \|Q_{1,T}\|^2 > \varepsilon \frac{\sigma_T^2 \lambda_{\min}(\sum_{j=1}^n C_j C_j')}{\max_{j \le n}(\|C_j\|^2)} \right\} \right] \\
&\le \frac{n \max_i \|C_i\|^2}{\sigma_T^2 \lambda_{\min}(\sum_{i=1}^n C_i C_i')} E\left[\|Q_{1,T}\|^2 1\left\{ \|Q_{1,T}\|^2 > \varepsilon \frac{\sigma_T^2 \lambda_{\min}(\sum_{i=1}^n C_i C_i')}{\max_{i \le n}(\|C_i\|^2)} \right\} \right].
\end{aligned}
$$

By conditions (i) and (ii),

$$
\frac{n \max_i \|C_i\|^2}{\sigma_T^2 \lambda_{\min}(\sum_{i=1}^n C_i C_i')} = O(1) \qquad \text{and} \qquad \frac{\sigma_T^2 \lambda_{\min}(\sum_{i=1}^n C_i C_i')}{\max_{i \le n}(\|C_i\|^2)} \to \infty,
$$

as $(n, T \to \infty)$. Then, since $\|Q_{i,T}\|^2$ is uniformly integrable in T by condition (iii), it follows that (8.14) $\to 0$ as $(n, T \to \infty)$. *Q.E.D.*

APPENDIX C: PROOFS FOR SECTION 4—SPURIOUS PANEL REGRESSION LIMIT THEORY

The next lemma gives the joint limit theory needed for Theorem 4.

LEMMA 13: *Suppose that Assumptions* 1–3 *hold.*
(a) *As* $(n, T \to \infty)$,

$$
\frac{1}{n} \sum_{i=1}^n \frac{1}{T^2} \sum_{t=1}^T Z_{i,t} Z_{i,t}' \xrightarrow{p} \frac{1}{2} E\Omega_i = \frac{1}{2}\Omega.
$$

(b) *If* $(n, T \to \infty)$ *and* $n/T \to 0$, *then*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{1}{T} \sum_{t=1}^{T} (Y_{i,t} X'_{i,t} - \beta X_{i,t} X'_{i,t}) \Rightarrow N(0, \Theta).$$

Proof of Lemma 13: (a) From the BN decomposition of $U_{i,t}$ in (2.4) we have

$$Z_{i,t} \overset{\text{a.s.}}{=} C_i(1) P_{i,t} + \tilde{U}_{i,0} - \tilde{U}_{i,t} + Z_{i,0},$$

where $P_{i,t} = \sum_{s=1}^{t} V_{i,s}$, which leads to

$$\frac{1}{n} \sum_{i=1}^{n} \frac{1}{T^2} \sum_{t=1}^{T} Z_{i,t} Z'_{i,t} \overset{\text{a.s.}}{=} \frac{1}{n} \sum_{i=1}^{n} (Q_{i,T} + R_{i,T}),$$

where

$$Q_{i,T} = \frac{1}{T^2} \sum_{t=1}^{T} C_i(1) P_{i,t} P'_{i,t} C_i(1),$$

$$R_{i,T} = R_{1,i,T} + R'_{1,i,T} + R_{2,i,T},$$

$$R_{1,i,T} + \frac{1}{T^2} \sum_{t=1}^{T} C_i(1) P_{i,t} (\tilde{U}_{i,0} - \tilde{U}_{i,t} + Z_{i,0})', \qquad \text{and}$$

$$R_{2,i,T} = \frac{1}{T^2} \sum_{t=1}^{T} (\tilde{U}_{i,0} - \tilde{U}_{i,t} + Z_{i,0})(\tilde{U}_{i,0} - \tilde{U}_{i,t} + Z_{i,0})'.$$

We show that as $(n, T \to \infty)$, $(1/n)\sum_{i=1}^{n} Q_{i,T} \to_p \frac{1}{2}\Omega$ and $(1/n)\sum_{i=1}^{n} R_{k,i,T} \to_p 0$, for $k = 1, 2$.

The $Q_{i,T}$ are iid across i for all T. Also, as $T \to \infty$, $Q_{i,T} \Rightarrow Q_i = C_i(1) \int W_i W_i' C_i(1)'$ and $(1/n)\sum_{i=1}^{n} Q_i \to_{\text{a.s.}} \frac{1}{2}\Omega$. That is, in sequential asymptotics such as $(T, n \to \infty)_{\text{seq}} (1/n)\sum_{i=1}^{n} Q_{i,T} \to_p \frac{1}{2}\Omega$. According to Corollary 1 (set $C_i = I_m$ so that the second condition is automatically satisfied), if we show that $\|Q_{i,T}\|$ is uniformly integrable in T, then it follows that

$$\text{(8.15)} \qquad \frac{1}{n} \sum_{i=1}^{n} Q_{i,T} \overset{p}{\to} \frac{1}{2}\Omega$$

as $(n, T \to \infty)$.

By $\|AB\| \le \|A\| \|B\|$ and the triangle inequality

$$\text{(8.16)} \qquad \|Q_{i,T}\| \le \|C_i(1)\|^2 \frac{1}{T^2} \sum_{t=1}^{T} \|P_{i,t}\|^2.$$

Also, as $T \to \infty$

$$\frac{1}{T^2} \sum_{t=1}^{T} \|P_{i,t}\|^2 \Rightarrow \int \|W_i\|^2,$$

and we have

$$E\left(\frac{1}{T^2} \sum_{t=1}^{T} \|P_{i,t}\|^2 \right) = \text{tr}\left(\frac{1}{T^2} \sum_{t=1}^{T} E(P_{i,t}, P'_{i,t}) \right) \to E\left(\int \|W_i\|^2 \right) = \frac{1}{2} \text{tr}(I_m).$$

It follows (e.g., Billingsley (1968, Theorem 5.4)) that $(1/T^2)\sum_{t=1}^{T} \|P_{i,t}\|^2$ is uniformly integrable in T. Since $E\|C_i(1)\|^2 < \infty$ by Lemma 1, we deduce that $\|C_i(1)\|^2 (1/T^2)\sum_{t=1}^{T} \|P_{i,t}\|^2$ is uniformly integrable in T. Thus, $\|Q_{i,T}\|$ is uniformly integrable in T, and (8.15) follows.

Next, $(1/n)\Sigma_{i=1}^{n} R_{1,i,T}$ and $(1/n)\Sigma_{i=1}^{n} R_{2,i,T}$ converge in probability to zero if $E\|R_{1,i,T}\|$, $E\|R_{2,i,T}\| \to 0$ as $(n, T \to \infty)$. Note that

$$E\|R_{1,i,T}\| \leq \frac{1}{T^2}\sum_{t=1}^{T} E[\|C_i(1)P_{i,t}\|\,\|\tilde{U}_{i,0} - \tilde{U}_{i,t} + Z_{i,0}\|] \tag{8.17}$$

$$\leq \frac{1}{\sqrt{T}}\frac{1}{T}\sum_{t=1}^{T}\sqrt{E\left\|C_i(1)\frac{P_{i,t}}{\sqrt{T}}\right\|^2 E\|\tilde{U}_{i,0} - \tilde{U}_{i,t} + Z_{i,0}\|^2}$$

$$= \frac{1}{\sqrt{T}}O(1),$$

where the first inequality holds by the triangle inequality and $\|AB\| \leq \|A\|\,\|B\|$, the second inequality holds by the Cauchy-Schwarz inequality and the last line holds because $E\|C_i(1)(P_{i,t}/\sqrt{T})\|^2 = O(1)$, $E\|Z_{i,0}\|^2 < M_1$, and $E\|\tilde{U}_{i,t}\|^2 < M_2$ $\forall t$ and for some $M_1, M_2 < \infty$ by Lemma 1(c). Thus, $E\|R_{1,i,T}\| \to 0$, as $(n, T \to \infty)$.

Similar arguments show that $E\|R_{2,i,T}\| = (1/T)O(1)$. So, all the desired results hold and part (a) is proved.

(b) Write $C_i(1) = (C_{y_i}(1)', C_{x_i}(1)')'$, and $\tilde{U}_{i,t} = (\tilde{U}'_{y_i,t}, \tilde{U}'_{x_i,t})'$, conformably with the partition of $Z_{i,t}$ into $Y_{i,t}$ and $X_{i,t}$. Using the BN decomposition of $U_{i,t}$ in (2.4), we have

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{1}{T^2}\sum_{t=1}^{T}(Y_{i,t}X'_{i,t} - \beta X_{i,t}X'_{i,t}),$$

$$= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(Q_{i,T} + R_{1,i,T} + R_{2,i,T} + R_{3,i,T} + R_{4,i,T} + R_{5,i,T}),$$

where

$$Q_{i,T} = \frac{1}{T^2}\sum_{t=1}^{T}\{C_{y_i}(1)P_{i,t}P'_{i,t}C_{x_i}(1)' - \beta C_{x_i}(1)P_{i,t}P'_{i,t}C_{x_i}(1)'\},$$

$$R_{1,i,T} = \frac{1}{T^2}\sum_{t=1}^{T}\left\{C_{y_i}(1)P_{i,t}\left(\tilde{U}_{x_i,0} - \tilde{U}_{x_i,t} + X_{i,0}\right)'\right\},$$

$$R_{2,i,T} = \frac{1}{T^2}\sum_{t=1}^{T}\left\{\left(\tilde{U}_{y_i,0} - \tilde{U}_{y_i,t} + Y_{i,0}\right)P'_{i,t}C_{x_i}(1)'\right\},$$

$$R_{3,i,T} = \frac{1}{T^2}\sum_{t=1}^{T}\left\{\left(\tilde{U}_{y_i,0} - \tilde{U}_{y_i,t} + Y_{i,0}\right)\left(\tilde{U}_{x_i,0} - \tilde{U}_{x_i,t} + X_{i,0}\right)'\right\},$$

$$R_{4,i,T} = \frac{1}{T^2}\sum_{t=1}^{T}\left\{\beta C_{x_i}(1)P_{i,t}\left(\tilde{U}_{x_i,0} - \tilde{U}_{x_i,t} + X_{i,0}\right)' + \beta\left(\tilde{U}_{x_i,0} - \tilde{U}_{x_i,t} + X_{i,0}\right)P'_{i,t}C_{x_i}(1)'\right\},$$

$$R_{5,i,T} = \frac{1}{T^2}\sum_{t=1}^{T}\left\{\beta\left(\tilde{U}_{x_i,0} - \tilde{U}_{x_i,t} + X_{i,0}\right)\left(\tilde{U}_{x_i,0} - \tilde{U}_{x_i,t} + X_{i,0}\right)'\right\}.$$

We show that as $(n, T \to \infty)$

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Q_{i,T} \Rightarrow N(0, \Theta) \tag{8.18}$$

and as $(n, T \to \infty)$ with $n/T \to 0$,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} R_{k,i,T} \xrightarrow{p} 0 \qquad (k = 1, \ldots, 5). \tag{8.19}$$

Note that

$$EQ_{i,T} = \frac{1}{T^2} \sum_{t=1}^{T} t(E[C_{y_i}(1) C_{x_i}(1)'] - \beta E[C_{x_i}(1) C_{x_i}(1)']) = 0.$$

Also,

$$\begin{aligned} \frac{1}{T^4} \sum_{t=1}^{T} \sum_{s=1}^{T} & E[\operatorname{vec}(P_{i,t} P_{i,t}') \operatorname{vec}(P_{i,s} P_{i,s}')'] \\ &= \frac{1}{T^4} \sum_{t=1}^{T} \sum_{s=1}^{T} E[P_{i,t} P_{i,s}' \otimes P_{i,t} P_{i,s}'] \\ &= \frac{1}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} \left\{ \begin{array}{l} \left(\dfrac{t \wedge s}{T} \right)^2 (I_{m^2} + K_m + \operatorname{vec} I_m (\operatorname{vec} I_m)') \\ \qquad + \left(\dfrac{t \wedge s}{T} \right) \left(\left(\dfrac{t \vee s}{T} \right) - \left(\dfrac{t \wedge s}{T} \right) \right) \operatorname{vec} I_m (\operatorname{vec} I_m)' \end{array} \right\} + O\left(\frac{1}{T} \right) \\ &= \Xi_T + O\left(\frac{1}{T} \right), \qquad \text{say,} \end{aligned} \tag{8.20}$$

and

$$\begin{aligned} & E(\operatorname{vec}(Q_{i,T}) \operatorname{vec}(Q_{i,T})') \\ &= E\left[(C_{x_i}(1) \otimes (C_{y_i}(1) - \beta C_{x_i}(1))) \frac{1}{T^4} \sum_{t=1}^{T} \sum_{s=1}^{T} E[\operatorname{vec}(P_{i,t} P_{i,t}') \operatorname{vec}(P_{i,s} P_{i,s}')'] \right. \\ &\qquad \left. \times (C_{x_i}(1)' \otimes (C_{y_i}(1) - \beta C_{x_i}(1))') \right] \\ &= E\left[(C_{x_i}(1) \otimes (C_{y_i}(1) - \beta C_{x_i}(1))) \left(\Xi_T + O\left(\frac{1}{T} \right) \right) \right. \\ &\qquad \left. \times (C_{x_i}(1)' \otimes (C_{y_i}(1) - \beta C_{x_i}(1))') \right] \\ &= \Xi_T, \qquad \text{say.} \end{aligned} \tag{8.21}$$

It is easy to see that $\Theta_T \to \Theta$ as $T \to \infty$. So $\{Q_{i,T}\}_i$ is an iid sequence with mean zero and covariance matrix Θ_T.

Next, apply Theorem 3 with $C_i = I_{m_y m_x}$ to establish that $(1/\sqrt{n}) \sum_{i=1}^{n} Q_{i,T} \Rightarrow N(0, \Theta)$ as $(n, T \to \infty)$. Conditions (i), (ii), and (iv) of the theorem are obviously satisfied in view of the fact that $C_i = I_{m_y m_x}$ and $\Theta_T \to \Theta$ as $T \to \infty$. For the uniform integrability of $\|Q_{i,t}\|^2$, note by the continuous mapping that as $T \to \infty$

$$\|Q_{i,T}\|^2 \Rightarrow \|Q_i\|^2 = \left\| \int \{C_{y_i}(1) W_i W_i' C_{x_i}(1)' - \beta C_{x_i}(1) W_i W_i C_{x_i}(1)'\} \right\|^2.$$

Then, $\|Q_{i,T}\|^2$ is uniformly integrable in T because

$$E\|Q_{i,T}\|^2 = \mathrm{tr}(E(\mathrm{vec}(Q_{i,T})\mathrm{vec}(Q_{i,T})')) = \mathrm{tr}(\Theta_T)$$
$$\to \mathrm{tr}(\Theta) = \mathrm{tr}(E(\mathrm{vec}(Q_i)\mathrm{vec}(Q_i)')) = E\|Q_i\|^2.$$

Next, to prove $(1/\sqrt{n})\sum_{i=1}^n R_{k,i,T} \to_p 0$ as $(n, T\to\infty)$ with $n/T \to 0$, we simply show that $\sqrt{n}\, E\|R_{k,i,T}\| \to 0$ as $n, T \to \infty$ with $n/T \to 0$ for $k = 1, \ldots, 5$. Note that

$$\sqrt{n}\, E\|R_{1,i,T}\| \leq \sqrt{n}\, \frac{1}{T^2} \sum_{t=1}^{T} E\left\{\|C_{y_i}(1) P_{i,t}\| \, \|\tilde{U}_{x_i,0} - \tilde{U}_{x_i,t} + X_{i,0}\|\right\}$$
$$= \sqrt{\frac{n}{T}}\, O(1) \to 0,$$

where the first inequality holds by the triangle inequality and $\|AB\| \leq \|A\|\,\|B\|$ and the last line holds in view of (8.17) above. Similarly we can show that $\sqrt{n}\, E\|R_{2,i,T}\|$, $\sqrt{n}\, E\|R_{4,i,T}\| = \sqrt{n/T}\, O(1)$ and $\sqrt{n}\, E\|R_{3,i,T}\|$, $\sqrt{n}\, E\|R_{5,i,T}\| = (\sqrt{n}/T) O(1)$. So we have the desired limits and part (b) follows. *Q.E.D.*

PROOF OF THEOREM 4: By Lemma 13(a), it is easy to see that as $(n, T \to \infty)$

$$\hat{\beta}_{n,T} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{T^2}\sum_{t=1}^{T} Y_{i,t}X_{i,t}'\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T^2}\sum_{t=1}^{T} X_{i,t}X_{i,t}'\right)^{-1} \xrightarrow{p} \Omega_{yx}\Omega_{xx}^{-1} = \beta.$$

Also, when $(n, T \to \infty)$ with $n/T \to 0$, from Lemma 13(a) and (b) we have

$$\sqrt{n}(\hat{\beta}_{n,T} - \beta) = \left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{1}{T^2}\sum_{t=1}^{T}(Y_{i,t}X_{i,t}' - \beta X_{i,t}X_{i,t}')\right)\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T^2}\sum_{t=1}^{T} X_{i,t}X_{i,t}'\right)^{-1}$$
$$\Rightarrow N\Big(0, 4\big(\Omega_{xx}^{-1} \otimes I_{m_y}\big)\Theta\big(\Omega_{xx}^{-1} \otimes I_{m_y}\big)\Big),$$

giving the required result. *Q.E.D.*

8.4. APPENDIX D: PROOFS FOR SECTION 5.1—HETEROGENEOUS PANEL COINTEGRATION LIMIT THEORY

The following two lemmas give some useful results on the existence of moments of the heterogeneous coefficients β_i in (5.2) and the random coefficients in the linear process representation (5.3). Both results are proved in PM[b].

LEMMA 14: *Under Assumptions* 4, 5, *and* 7, $E\|\beta_i\|^8 < \infty$.

LEMMA 15: *Let* $G_i(1) = \sum_{s=0}^{\infty} G_{i,s}$, $\tilde{G}_{i,s} = \sum_{t=s+1}^{\infty} G_{i,t}$, *and* $\tilde{F}_{i,t} = \sum_{s=0}^{\infty} \tilde{G}_{i,s} V_{i,t-s}$. *Suppose that Assumptions* 4–7 *hold. Then*:

(a) $E(\sum_{s=0}^{\infty} s^2 \|G_{i,s}\|^2) < \infty$.
(b) $E\|F_{i,t}\|^2 < M$ *for some constant* $M < \infty$.
(c) $E\|G_i(1)\|^4 < \infty$.
(d) $E\|\tilde{F}_{i,t}\|^4 < M$ *for some constant* $M < \infty$.

2. PROOF OF LEMMA 7: As in the proof of Lemma 3, since $\{\tilde{F}_{i,t}\}_t$ is strictly stationary and $\tilde{F}_{i,t}$ is square integrable from Lemma 15(d), it follows that

$$\sup_{0\leq r\leq 1}\left|\frac{1}{\sqrt{T}}\tilde{F}_{a,i,[Tr]}\right| \xrightarrow{p} 0 \qquad \text{as } T\to\infty \qquad \text{for all } a, i, \tag{8.22}$$

where $\tilde{F}_{a,i,[Tr]}$ is the ath element of $\tilde{F}_{i,[Tr]}$. The functional law follows directly from (5.5) and Donsker's theorem applied to $(1/\sqrt{T})\Sigma_{t=1}^{[Tr]}V_{i,t}$. *Q.E.D.*

3. PROOF OF LEMMA 8: The proof follows the same lines as Phillips (1988) and is omitted here (details are available in PM[b]). *Q.E.D.*

4. PROOF OF THEOREM 5: The proof follows similar lines to that of Lemma 13 and Theorem 4 above but makes use of the bounds established in Lemmas 14 and 15 and the panel BN decomposition (5.4). The details are lengthy and to save space they are not repeated here. They are given in full in PM[b]. *Q.E.D.*

5. PROOF OF THEOREM 6: The proof proceeds by showing that as $(n, T\to\infty)$ with $n, T\to 0$,

$$\hat{\Theta} = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{1}{T^4}\sum_{t=1}^{T}\sum_{s=1}^{T}X_{i,t}X'_{i,s}\otimes\hat{E}_{i,t}\hat{E}'_{i,x}\right\}\xrightarrow{p}\Theta,$$

and

$$\hat{\Omega}_{xx}^{-1} = \left[\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{2}{T^2}\sum_{t=1}^{T}X_{i,t}X'_{i,t}\right\}\right]^{-1}\xrightarrow{p}\Omega_{xx}^{-1}.$$

Then, by Theorem 5 and the delta method, the proof is complete. From Lemma 13, we know that $(1/n)\Sigma_{i=1}^{n}\{(2/T^2)\Sigma_{t=1}^{T}X_{i,t}X'_{i,t}\}\to_p \Omega_{xx}$ as $(n, T\to\infty)$. In consequence, $\hat{\Omega}_{xx}^{-1}\to_p\Omega_{xx}^{-1}$ as $(n, T\to\infty)$ since $\Omega_{xx}>0$. Again, full details are available in PM[b]. *Q.E.D.*

6. PROOF OF THEOREM 7: Under the null hypothesis, using the cross section independence and applying Theorem 5, we have

$$\begin{aligned}\sqrt{n_b}(\hat{\beta}_a-\hat{\beta}_b) &= \sqrt{n_a}(n_b/n_a)^{1/2}(\hat{\beta}_a-\beta_a)-\sqrt{n_b}(\hat{\beta}_b-\beta_b)\\ &\Rightarrow N(0,\kappa V_a+V_b),\end{aligned}$$

when $(T, n_a, n_b\to\infty)$ and $n_a/T, n_b/T\to 0$, and where $V_\mu = 4(\Omega_{\mu,xx}^{-1}\otimes I_{m_x})\Theta_\mu(\Omega_{\mu,xx}^{-1}\otimes I_{m_x})$ for $\mu = a, b$.

As in the proof of Theorem 6, we can show that as $(T, n_a, n_b\to\infty)$ with $n_a/T, n_b/T\to 0$, we have

$$\hat{\Theta}_\mu = \frac{1}{n_\mu}\sum_{i\in I_\mu}\left\{\frac{1}{T^4}\sum_{t=1}^{T}\sum_{s=1}^{T}X_{i,t}X'_{i,s}\otimes\hat{E}_{i,t}\hat{E}'_{i,s}\right\}\xrightarrow{p}\Theta_\mu,$$

and

$$\hat{\Omega}_{\mu,xx}^{1} = \left[\frac{1}{n_\mu}\sum_{i\in I_\mu}\left\{\frac{2}{T^2}\sum_{t=1}^{T}X_{i,t}X'_{i,t}\right\}\right]^{-1}\xrightarrow{p}\Omega_{\mu,xx}^{-1}.$$

Consequently, $\hat{V}_\mu \to_p V_\mu$, and $\hat{V}_{a-b} = (n_b/n_a)\hat{V}_a + \hat{V}_b \to_p \kappa V_a + V_b$. It follows that

$$W_{a,b} = n_b\{\mathrm{vec}(\hat{\beta}_a - \hat{\beta}_b)' \hat{V}_{a-b}^{-1} \mathrm{vec}(\hat{\beta}_a - \hat{\beta}_b)\} \Rightarrow \chi^2_{m_y m_x},$$

giving the required result.

APPENDIX E: PROOFS FOR SECTION 5.2—HOMOGENEOUS PANEL COINTEGRATION LIMIT THEORY

Before we start the proof of Theorem 9, we give the following useful lemma.

THEOREM 16: *Let $F_{i,t} = (E'_{i,t}, U'_{x_i,t})' = \sum_{s=0}^\infty G_s V_{i,t-s}$ be the panel process defined in Model* (5.11). *Also, let $S_{i,t} = \sum_{t=1}^T F_{i,t} + S_{i,0}$, where $S_{i,0}$ are iid with $E\|S_{i,0}\|^4 < \infty$, $\Omega_F = G(1)G(1)'$, $\Lambda_F = \sum_{k=0}^\infty \sum_{s=0}^\infty G_{s+k}G'_s$, and $G(1) = \sum_{s=0}^\infty G_s$. Then, under the summability condition Assumption* 9 *and positive definiteness condition Assumption* 10, *as $(n, T \to \infty)$ with $n/T \to 0$,*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\frac{1}{T} \sum_{t=1}^T F_{i,t} S'_{i,t} - \Lambda_F \right) \Rightarrow N\left(0, \frac{1}{2} \Omega_F \otimes \Omega_F \right).$$

PROOF OF LEMMA 16: Using the BN decomposition as in the proof of Lemma 8, we have

$$\begin{aligned}
&\frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\frac{1}{T} \sum_{t=1}^T F_{i,t} S'_{i,t} - \Lambda_F \right) \\
&\quad = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\frac{1}{T} \sum_{t=1}^T G(1)(V_{i,t}P'_{i,t} - I_m)G(1)' \right) \\
&\qquad + \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\frac{1}{T} \sum_{t=1}^{T-1} \left(\tilde{F}_{i,t}F'_{i,t+1} - \sum_{s=0}^\infty \tilde{G}_s G'_{s+1} \right) \right) + \frac{\sqrt{n}}{T} \sum_{s=0}^\infty \tilde{G}_s G'_{s+1} \\
&\qquad - \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\frac{1}{T} \sum_{t=1}^T (G(1)V_{i,t}\tilde{F}'_{i,t} - G(1)\tilde{G}'_0) \right) \\
&\qquad + \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\frac{1}{T} \sum_{t=1}^T G(1)V_{i,t}(\tilde{F}_{i,0} + S_{i,0})' \right) \\
&\qquad - \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{1}{T} \tilde{F}_{i,T} S'_{i,T} + \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{1}{T} \tilde{F}_{i,0} S'_{i,1} \qquad \text{a.s.} \\
&\quad = \frac{1}{\sqrt{n}} \sum_{i=1}^n (Q_{1,i,T} + R_{1,i,T} + R_{2,i,T} + R_{3,i,T} + R_{4,i,T} + R_{5,i,T}) \\
&\qquad + O\left(\frac{\sqrt{n}}{T} \right) \qquad \text{a.s., say.}
\end{aligned}$$

We show that $(1/\sqrt{n})\sum_{i=1}^n Q_{i,T} \Rightarrow N(0, \frac{1}{2}\Omega_F \otimes \Omega_F)$, and $(1/\sqrt{n})\sum_{i=1}^n R_{k,i,T} \to_p 0$, $k = 1, \ldots, 5$, as $(n, T \to \infty)$ with $n/T \to 0$.

Note that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Q_{i,T} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{t=2}^{T} G(1)V_{i,t}P_{i,t-1}'G(1)'\right)$$
$$+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{i=1}^{T} G(1)(V_{i,t}V_{i,t}'-I_m)G(1)'\right)$$
$$=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(Q_{1,i,T}+Q_{2,i,T}), \qquad \text{say.}$$

Since $E((1/T)\sum_{t=1}^{T} G(1)(V_{i,t}V_{i,t}'-I_m)G(1)')=0$, we have

$$E\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Q_{2,i,T}\right\|^2$$
$$= E\left\|\frac{1}{T}\sum_{t=1}^{T} G(1)(V_{i,t}V_{i,t}'-I_m)G(1)'\right\|^2$$
$$\leq \|G(1)\|^4 \operatorname{tr}\left[\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T} E\{(V_{i,t}V_{i,s}'\otimes V_{i,t}V_{i,s}')-\operatorname{vec}(I_m)(\operatorname{vec}(I_m))'\}\right]$$
$$= O\left(\frac{1}{T}\right).$$

Thus, $(1/\sqrt{n})\sum_{i=1}^{n} Q_{2,i,T}=o_p(1)$. Next, observe that

$$E[\operatorname{vec}(Q_{1,i,T})(\operatorname{vec}(Q_{1,i,T}))']$$
$$=\frac{1}{T^2}\sum_{t=2}^{T}\sum_{s=2}^{T}(G(1)\otimes G(1))E(P_{i,t-1}P_{i,s-1}'\otimes V_{i,t}V_{i,s}')(G(1)'\otimes G(1)')$$
$$=\frac{1}{T}\sum_{t=2}^{T}\left(\frac{t-1}{T}\right)(G(1)G(1)'\otimes G(1)G(1)')\equiv \Xi_T^* \qquad \text{(say)}$$
$$\to \frac{1}{2}(G(1)G(1)'\otimes G(1)G(1)')=\frac{1}{2}(\Omega_F\otimes\Omega_F)\equiv\Xi^*, \qquad \text{say.}$$

Also, note that $\frac{1}{2}(\Omega_F\otimes\Omega_F)>0$. These verify conditions (i), (ii), and (iv) of Theorem 3. Condition (iii) of Theorem 3 holds because

$$\|Q_{i,T}\|^2 \Rightarrow \|Q_i\|^2 = \left\|G(1)\int dW_i\, W_i' G(1)'\right\|^2$$

and

$$Q_{i,T}^2 = \operatorname{tr}(\Xi_T^*)\to \operatorname{tr}(\Xi^*) = E\|Q_i\|^2$$

so that the $\|Q_{i,T}\|^2$ are uniformly integrable in T. By Theorem 3, $(1/\sqrt{n})\sum_{i=1}^{n} Q_{1,i,T}\Rightarrow N(0,\frac{1}{2}\Omega_F\otimes\Omega_F)$.

Next, we show that $(1/\sqrt{n})\sum_{i=1}^{n} R_{1,i,T} \to_p 0$ by proving $E\|(1/\sqrt{n})\sum_{i=1}^{n} R_{1,i,T}\|^2 \to 0$ as $(n, T \to \infty)$. Note that

$$
\begin{aligned}
&E\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} R_{1,i,T}\right\|^2 \\
&\quad = \operatorname{tr}[E(\operatorname{vec}(R_{1,i,T})(\operatorname{vec}(R_{1,i,T}))')] \text{ since } E(R_{1,i,T}) = 0 \\
&\quad = \operatorname{tr}\left[\frac{1}{T^2}\sum_{t=1}^{T-1}\sum_{s=1}^{T-1} E\left\{\begin{array}{l}\operatorname{vec}\left(\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\tilde{G}_j V_{i,t-j}V'_{i,t+1-k}G'_k\right)\\ \qquad \times\left(\operatorname{vec}\left(\sum_{p=0}^{\infty}\sum_{q=0}^{\infty}\tilde{G}_p V_{i,s-p}V'_{i,s+1-q}G'_q\right)\right)' \\ \qquad -\operatorname{vec}\left(\sum_{s=0}^{\infty}\tilde{G}_s G'_{s+1}\right)\left(\operatorname{vec}\left(\sum_{s=0}^{\infty}\tilde{G}_s G'_{s+1}\right)\right)'\end{array}\right\}\right] \\
&\quad = \operatorname{tr}\left[\frac{1}{T}\sum_{h=-T+2}^{T-2}\left(\frac{T-1-|h|}{T}\right)\right. \\
&\qquad \left.\times\left(\begin{array}{l}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\sum_{p=0}^{\infty}\sum_{q=0}^{\infty}(G_k\otimes\tilde{G}_j)\\ \qquad \times E(V_{i,t+1-k}V'_{i,t+h+1-q}\otimes V_{i,t-j}V'_{i,t+h-p})(G_q\otimes\tilde{G}_p)' \\ \qquad -\operatorname{vec}\left(\sum_{s=0}^{\infty}\tilde{G}_s G'_{s+1}\right)\left(\operatorname{vec}\left(\sum_{s=0}^{\infty}\tilde{G}_s G'_{s+1}\right)\right)'\end{array}\right)\right].
\end{aligned}
$$

If we show

$$
\operatorname{tr}\sum_{h=0}^{\infty}\left(\begin{array}{l}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\sum_{p=0}^{\infty}\sum_{q=0}^{\infty}(G_k\otimes\tilde{G}_j)\\ \qquad \times E(V_{i,t+1-k}V'_{i,t+h+1-q}\otimes V_{i,t-j}V'_{i,t+h-p})(G_q\otimes\tilde{G}_p)' \\ \qquad -\operatorname{vec}\left(\sum_{s=0}^{\infty}\tilde{G}_s G'_{s+1}\right)\left(\operatorname{vec}\left(\sum_{s=0}^{\infty}\tilde{G}_s G'_{s+1}\right)\right)'\end{array}\right) < \infty,
$$

then, by Cesaro summability, it follows that $E\|(1/\sqrt{n})\sum_{i=1}^{n} R_{1,i,T}\|^2 = O(1/T)$. Observe that

$$
\begin{aligned}
(8.23)\qquad &= \sum_{h=0}^{\infty}\left(\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\operatorname{tr}\left(G_k G'_{k+h}\otimes\tilde{G}_j\tilde{G}'_{j+h}\right)\right) \\
&\quad + \sum_{h=0}^{\infty}\left(\sum_{j=0}^{\infty}\sum_{k=(0\vee(1-h))}^{\infty}\operatorname{tr}\left(G_k\tilde{G}'_{k+h-1}\otimes\tilde{G}_j G'_{j+h+1}\right)K_m\right) \\
&\quad + (v^4-3)\sum_{h=0}^{\infty}\sum_{j=0}^{\infty}\operatorname{tr}\left((G_{j+1}\otimes\tilde{G}_j)\left(\sum_{l=1}^{m} e_{l,l}\otimes e_{l,l}\right)(G'_{j+h+1}\otimes\tilde{G}'_{j+h})\right) \\
&= I + II + III, \qquad \text{say.}
\end{aligned}
$$

Since $\mathrm{tr}(A \otimes B) = \mathrm{tr}(A)\mathrm{tr}(B)$ and $\mathrm{tr}(A) \le \mathrm{rows}((A))^{1/2}\|A\|$ from (8.2) in Lemma 9, we have

$$\begin{aligned} I &= \sum_{h=0}^{\infty} \mathrm{tr}\left(\sum_{k=0}^{\infty} G_k G'_{k+h} \right) \mathrm{tr}\left(\sum_{j=0}^{\infty} \tilde{G}_j \tilde{G}'_{j+h} \right) \\ &\le \left[\sum_{h=0}^{\infty} \left| \mathrm{tr}\left(\sum_{k=0}^{\infty} G_k G'_{k+h} \right) \right| \right] \left[\sum_{h=0}^{\infty} \left| \mathrm{tr}\left(\sum_{j=0}^{\infty} \tilde{G}_j \tilde{G}'_{j+h} \right) \right| \right] \\ &\le m \left(\sum_{k=0}^{\infty} \|G_k\| \right)^2 \left(\sum_{k=0}^{\infty} \|\tilde{G}_k\| \right)^2 < \infty \qquad \text{by Assumption 9.} \end{aligned}$$

By Lemma 10,

$$\begin{aligned} II &\le \sum_{k=1}^{\infty} \sum_{j=0}^{\infty} \|G_k\| \|\tilde{G}_{k-1}\| \|\tilde{G}_j\| \|G_{j+1}\| + \sum_{h=1}^{\infty} \left(\sum_{k=0}^{\infty} \sum_{j=0}^{\infty} \|G_k\| \|\tilde{G}_{k+h-1}\| \|\tilde{G}_j\| \|G_{j+h+1}\| \right) \\ &\le \left(\sum_{j=0}^{\infty} \|G_j\| \right)^4 + \left(\sum_{h=0}^{\infty} \sum_{k=0}^{\infty} \|G_k\| \|\tilde{G}_{k+h}\| \right) \left(\sum_{h=0}^{\infty} \sum_{j=0}^{\infty} \|\tilde{G}_j\| \|G_{j+h}\| \right) \\ &\le \left(\sum_{j=0}^{\infty} \|G_j\| \right)^4 + \left(\sum_{j=0}^{\infty} j\|G_j\| \right)^2 \left(\sum_{j=0}^{\infty} \|G_j\| \right)^2 < \infty. \end{aligned}$$

Similarly, we can show that for some $M > 0$

$$III \le M \left(\sum_{j=0}^{\infty} \|G_j\| \right)^4 < \infty.$$

Also, we can show by modifying the arguments used above that

$$E\left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} R_{2,i,T} \right\|^2, \qquad E\left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} R_{3,i,T} \right\|^2 = O\left(\frac{1}{T} \right),$$

and

$$E\left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} R_{4,i,T} \right\| = O\left(\sqrt{\frac{n}{T}} \right), \qquad E\left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} R_{5,i,T} \right\| = O\left(\frac{\sqrt{n}}{T} \right),$$

so all the desired results are proved and the lemma follows. *Q.E.D.*

PROOF OF THEOREM 9: To establish joint limit normality of the PFM estimator $\hat{\beta}_{PFM}$, it is enough to show that, as $(n, T \to \infty)$ with $n/T \to 0$,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\frac{1}{T} \sum_{t=1}^{T} (\hat{E}^+_{i,t} X'_{i,t} - \hat{\Lambda}^+_{ex}) \right) \Rightarrow N(0, \tfrac{1}{2}(\Omega_{xx} \otimes \Omega_{e.x})),$$

and

$$\left[\frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{1}{T^2} \sum_{t=1}^{T} X_{i,t} X'_{i,t} \right\} \right]^{-1} \overset{p}{\to} 2\Omega_{xx}^{-1}.$$

The proof of the latter result is similar to the proof of Lemma 13(a), so we concentrate on the former.

Let $\Lambda_{ex}^{+}=\Lambda_{ex}-\Omega_{ex}\Omega_{xx}^{-1}\Lambda_{xx}$ and $E_{i,t}^{+}=E_{i,t}-\Omega_{ex}\Omega_{xx}^{-1}\,\Delta X_{i,t}$. Then

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{t=1}^{T}(\hat{E}_{i,t}^{+}X_{i,t}'-\hat{\Lambda}_{ex}^{+})\right)\\ &=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{t=1}^{T}(E_{i,t}^{+}X_{i,t}'-\Lambda_{ex}^{+})\right)\\ &\quad-(\hat{\Omega}_{ex}\hat{\Omega}_{xx}^{-1}-\Omega_{ex}\Omega_{xx}^{-1})\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{t=1}^{T}(\Delta X_{i,t}\,X_{i,t}'-\Lambda_{xx})\right)\\ &\quad-\sqrt{n}(\hat{\Lambda}_{ex}-\Lambda_{ex})+\hat{\Omega}_{ex}\hat{\Omega}_{xx}^{-1}\sqrt{n}(\hat{\Lambda}_{xx}-\Lambda_{xx}).\end{aligned}$$

First, $(\hat{\Omega}_{ex}\hat{\Omega}_{xx}^{-1}-\Omega_{ex}\Omega_{xx}^{-1})(1/\sqrt{n})\sum_{i=1}^{n}((1/T)\sum_{t=1}^{T}(\Delta X_{i,t}\,X_{i,t}'-\Lambda_{xx}))=o_p(1)$ because $\hat{\Omega}_{ex}\hat{\Omega}_{xx}^{-1}-\Omega_{ex}\Omega_{xx}^{-1}=o_p(1)$ and $(1/\sqrt{n})\sum_{i=1}^{n}((1/T)\sum_{t=1}^{T}(\Delta X_{i,t}\,X_{i,t}'-\Lambda_{xx}))=O_p(1)$ by Lemma 16 and $E\|X_{i,0}\|^4<M$ for some constant M. Next, according to Theorems 9 and 10 in Hannan (1970, pp. 280–283) (or Proposition 1 in Andrews (1991)), we know that $E\|\hat{\Omega}_{F,i}-E\hat{\Omega}_{F,i}\|^2=(K/T)O(1)$, and $\|E\hat{\Omega}_{F,i}-\Omega_F\|^2=(1/K^{2q})O(1)$. Thus,

$$\begin{aligned}E\|\sqrt{n}(\hat{\Omega}_F-\Omega_F)\|^2&=E\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\hat{\Omega}_{F,i}-E\hat{\Omega}_{F,i}+E\hat{\Omega}_{F,i}-\Omega_F)\right\|^2\\ &=E\|\hat{\Omega}_{F,i}-E\hat{\Omega}_{F,i}\|^2+n\|E\hat{\Omega}_{F,i}-\Omega_F\|^2\\ &=\left(\frac{K}{T}+\frac{n}{K^{2q}}\right)O(1).\end{aligned}$$

Since the bandwidth parameter $K\to\infty$ with $K/T\to 0$ and $K^{2q}/T\to\epsilon>0$ for some $q>\frac{1}{2}$ by Assumption 11, it follows that $E\|\sqrt{n}(\hat{\Omega}_F-\Omega_F)\|^2\to 0$ as $(n,T\to\infty)$ with $n/T\to 0$. The same argument can be applied to $\hat{\Lambda}_F$. In consequence, we have

$$\sqrt{n}(\hat{\Lambda}_{ex}-\Lambda_{ex}),\ \hat{\Omega}_{ex}\hat{\Omega}_{xx}^{-1}\sqrt{n}(\hat{\Lambda}_{xx}-\Lambda_{xx})=o_p(1).$$

The remainder of the proof involves showing that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{t=1}^{T}(E_{i,t}^{+}X_{i,t}'-\Lambda_{ex}^{+})\right)\Rightarrow N\left(0,\frac{1}{2}(\Omega_{xx}\otimes\Omega_{e.x})\right),$$

and this is entirely analogous to the proof of Lemma 16. The main contribution of $(1/\sqrt{n})\sum_{i=1}^{n}((1/T)\sum_{t=1}^{T}(E_{i,t}^{+}X_{i,t}'-\Lambda_{ex}^{+}))$ from the BN decomposition is

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{t=1}^{T}(G_e(1)-\Omega_{ex}\Omega_{xx}^{-1}C_x(1))V_{i,t}P_{i,t-1}'C_x(1)'\right)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}Q_{i,T},$$

and it is easy to see that

$$\begin{aligned}E[\operatorname{vec}(Q_{i,T})(\operatorname{vec}(Q_{i,T}))']&=\left(\frac{1}{T}\sum_{t=1}^{T}\frac{t-1}{T}\right)(\Omega_{xx}\otimes\Omega_{e.x})\\ &\to\frac{1}{2}(\Omega_{xx}\otimes\Omega_{e.x})=E[\operatorname{vec}(Q_i)(\operatorname{vec}(Q_i))'],\end{aligned}$$

where $Q_i = (G_e(1) - \Omega_{ex}\Omega_{xx}^{-1}C_x(1))\int dW_i\ W_i' C_x(1)'$. Thus, by Theorem 3, we have the desired result. All the remainder terms in the BN decomposition of $(1/\sqrt{n})\sum_{i=1}^{n}((1/T)\sum_{t=1}^{T}(E_{i,t}^{+}X_{i,t}' - \Lambda_{ex}^{+}))$ converge in probability to zero by Lemma 16 and the moment bound $E\|X_{i,0}\|^4 < M$, for some constant M. *Q.E.D.*

REFERENCES

ANDREWS, D. W. K. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–858.

BALTAGI, B. (1995): *Econometric Analysis of Panel Data*. New York: Wiley.

BILLINGSLEY, P. (1968): *Convergence of Probability Measures*. New York: Wiley.

——— (1986): *Probability and Measure*. New York: Wiley.

CHAMBERLAIN, G. (1984): "Panel Data," in *Handbook of Econometrics*, Vol. 2, ed. by Z. Griliches and M. Intriligator. Amsterdam: North-Holland.

CONLEY, T. (1997): "Econometric Modelling of Cross Sectional Dependence," Mimeo.

DUDLEY, R. (1989): *Real Analysis and Probability*. Pacific Grove: Wadsworth & Brooks/Cole Mathematics Series.

EICKER, F. (1963): "Central Limit Theorems for Families of Sequences of Random Variables," *Annals of Mathematical Statistics*, 34, 439–446.

GRANGER, C. W. J., AND P. NEWBOLD (1974): "Spurious Regressions in Econometrics," *Journal of Econometrics*, 2, 111–120.

HSIAO, C. (1986): *Analysis of Panel Data*. Cambridge: Cambridge University Press.

HALL, P., AND C. HEYDE (1980): *Martingale Limit Theory and its Applications*. New York: Academic Press.

HANNAN, E. (1970): *Multiple Time Series*. New York: Wiley.

IM, K., H. PESARAN, AND Y. SHIN (1996): "Testing for Unit Roots in Heterogeneous Panels," Mimeo.

LEVIN, A., AND C. LIN (1993): "Unit Root Tests in Panel Data: New Results," UC San Diego Working Paper.

MAGNUS, J., AND H. NEUDECKER (1988): *Matrix Differential Calculus*. New York: Wiley.

MATYAS, L., AND P. SEVESTRE (EDS.) (1992): *The Econometrics of Panel Data*. Boston, MA: Kluwer Academic Publishers.

MUIRHEAD, R. (1982): *Aspects of Multivariate Statistical Theory*. New York: Wiley.

PEDRONI, P. (1995): "Panel Cointegration; Asymptotic and Finite Sample Properties of Pooled Time Series Tests with an Application to the PPP Hypothesis," Indiana University Working Papers in Economics No. 95-013.

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

PHILLIPS, P. C. B. (1986): "Understanding Spurious Regressions in Econometrics," *Journal of Econometrics*, 33, 311–340.

——— (1988): "Weak Convergence of Sample Covariance Matrices to Stochastic Integrals via Martingale Approximations," *Econometric Theory*, 4, 528–533.

PHILLIPS, P. C. B., AND S. DURLAUF (1986): "Multiple Time Series Regression with Integrated Processes," *Review of Economic Studies*, 53, 473–495.

PHILLIPS, P. C. B., AND B. HANSEN (1990): "Statistical Inference in Instrumental Variables Regression with $I(1)$ Processes," *Review of Economic Studies*, 57, 99–125.

PHILLIPS, P. C. B., AND C. LEE (1996): "Efficiency Gains from Quasi-Differerencing under Nonstationarity," in *Athens Conference on Applied Probability and Time Series: Volume II, Time Series Analysis in Memory of E. J. Hannan*, ed. by P. M. Robinson and M. Rosenblatt. New York: Springer-Verlag.

PHILLIPS, P. C. B., AND H. MOON (1997a): "Linear Regression Limit Theory for Nonstationary Panel Data," University of Auckland Discussion Paper in Economics.

——— (1997b): "Linear Regression Limit Theory for Nonstationary Panel Data," Yale University, Mimeographed.

PHILLIPS, P. C. B., AND V. SOLO (1992): "Asymptotics for Linear Processes," *Annals of Statistics*, 20, 971–1001.

POLLARD, D. (1984): *Convergence of Stochastic Processes*. New York: Springer Verlag.

QUAH, D. (1994): "Exploiting Cross-Section Variations for Unit Root Inference in Dynamic Data," *Economic Letters*, 44, 9–19.

ROBERTSON, D., AND J. SYMONS (1992): "Some Strange Properties of Panel Data Estimators," *Journal of Applied Econometrics*, 7, 175–189.

SHORACK, G., AND J. WELLNER (1986): *Empirical Processes with Applications to Statistics*. New York: Wiley.