On the Estimation and Inference of a Cointegrated Regression in Panel Data

Chihwa Kao and Min-Hsien Chiang*
Syracuse University

First Draft: May 1, 1995
First Revised Draft: January 1997
This Draft: February 27, 1997

Abstract

In this paper, we study the asymptotic distributions for least-squares (OLS), fully modified (FM), and dynamic OLS (DOLS) estimators in cointegrated regression models in panel data. We show that the OLS, FM, and DOLS estimators are all asymptotically normally distributed. However, the asymptotic distribution of the OLS estimator is shown to have a non-zero mean. Monte Carlo results examine the sampling behavior of the proposed estimators and show that (1) the OLS estimator has a non-negligible bias in finite samples, (2) the FM estimator does not improve over the OLS estimator in general, and (3) the DOLS out-performs both the OLS and FM estimators.

Key Words: *Panel Data, OLS Estimator; FM Estimator, DOLS Estimator, Heterogeneous Panels.*

1 Introduction

Evaluating the statistical properties of data along the time dimension has proven to be very different from analysis of the cross-section dimension. As economists have gained access to better data with more observations across time, understanding these properties has grown increasingly important. An area of particular concern in time series econometrics has been the use of non-stationary data. With the desire to study the

*We are thankful to three anonymous referees and Peter Phillips for pointing out several technical errors in an earlier version and providing comments that led to improvements of this paper. We also thank Peter Pedroni for his helpful comments and Bangtian Chen for his research assistance on an earlier draft of this paper. An electronic version of the paper in postscript format can be retrieved from http://web.syr.edu/~cdkao. Address correspondence to: Chihwa Kao, Center for Policy Research, 426 Eggers, Syracuse University, Syracuse, NY 13244; e-mail: cdkao@maxwell.syr.edu.

behavior of a cross-sectional data over time and the increasing use of panel data, one new research area is examining the properties of non-stationary time-series data in panel form. It is an intriguing question to ask: how exactly does this hybrid style of data combine the statistical elements of traditional cross-sectional analysis and time-series analysis? In particular, what is the correct way to analyze non-stationarity, the spurious regression problem, and cointegration in panel data?

Given the immense interest in testing for unit roots and cointegration in time-series data, not much attention has been paid to testing the unit roots in panel data. The only theoretical studies as far as we know in this area are Breitung and Meyer (1994), Quah (1994), Levin and Lin (1993), Im, Pesaran, and Shin (1995), and Maddala and Wu (1996). Breitung and Meyer (1994) derived the asymptotic normality of the Dickey-Fuller test statistic for panel data with a large cross-section dimension and a small time-series dimension. Quah (1994) studied a unit root test for panel data that have simultaneously extensive cross-section and time-series variation. He showed that the asymptotic distribution for the proposed test is a mixture of the standard normal and Dickey-Fuller-Phillips asymptotics. Levin and Lin (1993) derived the asymptotic distributions for unit roots on panel data and showed that the power of these tests increases dramatically as the cross-section dimension increases. Im et al. (1995) critiqued the Levin and Lin panel unit root statistics and proposed alternatives. Maddala and Wu (1996) provided a comparison of the tests of Im et al. (1995) and Levin and Lin (1993). They suggested a new test based on the Fisher test. However, to this date, little is known about cointegration tests and estimation with regression models in panel data. Exceptions are Kao (1996), McCoskey and Kao (1996), and Pedroni (1995, 1996). In the first half of Kao (1996), he studied a spurious regression in panel data. Asymptotic properties of the least-squares (OLS) estimator and other conventional statistics were examined. Kao (1996) showed that the OLS estimator is consistent for its true value, but the t-statistic diverges so that inferences about the regression coefficient, β, are wrong with probability that goes to one as $N \to \infty$ and $T \to \infty$. Furthermore, Kao (1996) examined the Dickey-Fuller (DF) and the augmented Dickey-Fuller (ADF) tests to test the null hypothesis of no cointegration in panel data. McCoskey and Kao (1996) proposed further tests for the null hypothesis of cointegration in panel data. Pedroni (1995) derived asymptotic distributions for residual based tests of cointegration for both homogenous and heterogenous panels. Pedroni (1996) proposed a fully modified estimator for heterogenous panels. On the other hand, Park and Ogaki (1991) derived asymptotic distributions for cointegration coefficient estimators and related t-statistics for panel data using CCR transformations. Although they used an SUR approach rather than N dimension asymptotics, many of the issues they dealt with are similar.

In this paper, we study the limiting distributions for the ordinary least squares (OLS), fully modified (FM), and Dynamic OLS (DOLS) estimators in panel cointegrated regression models.

Section 2 introduces the model and assumptions. Section 3 develops the asymptotic theory for OLS, FM and DOLS estimators. Section 4 gives the limiting distributions of FM and DOLS estimators for heterogeneous panels. Section 5 develops the limiting distributions of the Wald statistics. Section 6 presents some Monte Carlo results to evaluate the finite sample properties of the OLS, FM, and dynamic OLS estimators. Section 7 summarizes the findings. All proofs are in the Appendix.

A word on notation. We write the integral $\int_0^1 W(s)ds$ as $\int W$ when there is no ambiguity over limits. We define $\Omega^{1/2}$ to be any matrix such that $\Omega = \left(\Omega^{1/2}\right)\left(\Omega^{1/2}\right)^{'}$. We use $\|A\|$ to denote $\left\{tr\left(A^{'}A\right)\right\}^{1/2}$, $|A|$ to denote the determinant of A, $\xrightarrow{d}$ to denote convergence in distribution, $\xrightarrow{p}$ to denote convergence in probability, $[x]$ to denote the largest integer $\leq x$, $I(0)$ and $I(1)$ to signify a time series that is integrated of order zero and one, respectively, and $BM\left(\Omega\right)$ to denote Brownian motion with covariance matrix Ω.

2 The Model and Assumptions

Consider the following fixed effect panel regression:

$$y_{it} = \alpha_i + x_{it}^{'}\beta + u_{it},\; i = 1, ..., N,\; t = 1, ...T, \tag{1}$$

where $\{y_{it}\}$ are 1×1, β is a $k \times 1$ vector of the slope parameters, $\{\alpha_i\}$ are the intercepts, and $\{u_{it}\}$ are the stationary disturbance terms. We assume that $\{x_{it}\}$ are $k \times 1$ integrated processes of order one for all i, where

$$x_{it} = x_{it-1} + \varepsilon_{it}.$$

Under these specifications, (1) describes a system of cointegrated regressions, i.e., y_{it} is cointegrated with x_{it}. The initialization of this system is $y_{i0} = x_{i0} = 0$ for all i.

Assumption 1 *$\{y_{it}, x_{it}\}$ are independent across i.*

Assumption 2 *The cross-section dimension is a monotonic function of the time-series dimension, i.e., $N = N(T)$, so that the law of large numbers (Theorem 6.2, Billingsley, 1986, p. 81) and the central limit theorem (Theorem 27.2, Billingsley, 1986, p. 369) for triangular arrays can be applied.*

Next, we characterize the innovation vector $w_{it} = \left(u_{it}, \varepsilon_{it}^{'}\right)^{'}$. We assume that w_{it} is a linear process that satisfies the following assumption.

Assumption 3 *(e.g., Phillips, 1995)*

(a) $w_{it} = \Pi(L)\epsilon_{it} = \sum_{j=0}^{\infty} \Pi_j \epsilon_{it-j}, \sum_{j=0}^{\infty} j^a \left\|\Pi_j\right\| < \infty, |\Pi(1)| \neq 0$ *for some* $a > 1$.

(b) ϵ_{it} is i.i.d. with zero mean, variance matrix Σ_ϵ, and finite fourth order cumulants.

Assumption 3 implies that (e.g., Phillips and Solo, 1992) the partial sum process $\frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]} w_{it}$ satisfies the following multivariate invariance principle:

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]} w_{it} \xrightarrow{d} B_i(r) \equiv BM_i\left(\Omega\right) \text{ as } T\to\infty, \tag{2}$$

where

$$B_i = \begin{bmatrix} B_{ui} \\ B_{\varepsilon i} \end{bmatrix}.$$

The long-run covariance matrix of $\{w_{it}\}$ is given by

$$\begin{aligned} \Omega &= \sum_{j=-\infty}^{\infty} E\left(w_{ij}w_{i0}^{'}\right) \\ &= \Pi(1)\Sigma_\epsilon \Pi(1)^{'} \\ &= \Sigma + \Gamma + \Gamma^{'} \\ &\equiv \begin{bmatrix} \Omega_u & \Omega_{u\varepsilon} \\ \Omega_{\varepsilon u} & \Omega_\varepsilon \end{bmatrix}, \end{aligned}$$

where

$$\Gamma = \sum_{j=1}^{\infty} E\left(w_{ij}w_{i0}^{'}\right) \equiv \begin{bmatrix} \Gamma_u & \Gamma_{u\varepsilon} \\ \Gamma_{\varepsilon u} & \Gamma_\varepsilon \end{bmatrix} \tag{3}$$

and

$$\Sigma = E\left(w_{i0}w_{i0}^{'}\right) \equiv \begin{bmatrix} \Sigma_u & \Sigma_{u\varepsilon} \\ \Sigma_{\varepsilon u} & \Sigma_\varepsilon \end{bmatrix} \tag{4}$$

are partitioned conformably with w_{it}.

Assumption 4 *Ω_ε is non-singular, i.e., $\{x_{it}\}$ are not cointegrated.*

Define

$$\Omega_{u.\varepsilon} = \Omega_u - \Omega_{u\varepsilon}\Omega_\varepsilon^{-1}\Omega_{\varepsilon u}. \tag{5}$$

Then, B_i can be rewritten as

$$B_i = \begin{bmatrix} B_{ui} \\ B_{\varepsilon i} \end{bmatrix} = \begin{bmatrix} \Omega_{u.\varepsilon}^{1/2} & \Omega_{u\varepsilon}\Omega_\varepsilon^{-1/2} \\ 0 & \Omega_\varepsilon^{1/2} \end{bmatrix} \begin{bmatrix} V_i \\ W_i \end{bmatrix},$$

where $\begin{bmatrix} V_i \\ W_i \end{bmatrix} = BM(I)$ is a standardized Brownian motion. Define the one-sided long-run covariance

$$\begin{aligned} \Delta &= \Sigma + \Gamma \\ &= \sum_{j=0}^{\infty} E\left(w_{ij} w_{i0}^{'}\right) \end{aligned}$$

with

$$\Delta = \begin{bmatrix} \Delta_u & \Delta_{u\varepsilon} \\ \Delta_{\varepsilon u} & \Delta_{\varepsilon} \end{bmatrix}.$$

Remark 1 *Here we assume that panels are homogeneous, i.e., the variances are constant across the cross-section units. We will relax this assumption in Section 4 to allow for different variances for different i.*

3 OLS, Fully Modified, and Dynamic OLS Estimators

Let us first study the limiting distribution of the OLS estimator for equation (1). The OLS estimator of β is

$$\widehat{\beta}_{OLS} = \left[\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}-\overline{x}_i)(x_{it}-\overline{x}_i)^{'}\right]^{-1}\left[\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}-\overline{x}_i)(y_{it}-\overline{y}_i)\right]. \tag{6}$$

It follows that

$$\begin{aligned} &\sqrt{N}T\left(\widehat{\beta}_{OLS}-\beta\right) \\ &= \left[\tfrac{1}{N}\textstyle\sum_{i=1}^{N}\tfrac{1}{T^2}\sum_{t=1}^{T}(x_{it}-\overline{x}_i)(x_{it}-\overline{x}_i)^{'}\right]^{-1}\left[\sqrt{N}\tfrac{1}{N}\textstyle\sum_{i=1}^{N}\left(\tfrac{1}{T}\sum_{t=1}^{T}(x_{it}-\overline{x}_i)u_{it}\right)\right] \\ &= \left[\tfrac{1}{N}\textstyle\sum_{i=1}^{N}\zeta_{2iT}\right]^{-1}\left[\sqrt{N}\tfrac{1}{N}\textstyle\sum_{i=1}^{N}\zeta_{1iT}\right] \\ &= \left[\xi_{2NT}\right]^{-1}\sqrt{N}\xi_{1NT}, \end{aligned}$$

where $\overline{x}_i = \frac{1}{T}\sum_{t=1}^{T}x_{it}$, $\overline{y}_i = \frac{1}{T}\sum_{t=1}^{T}y_{it}$, $\zeta_{1iT} = \frac{1}{T}\sum_{t=1}^{T}(x_{it}-\overline{x}_i)u_{it}$, $\zeta_{2iT} = \frac{1}{T^2}\sum_{t=1}^{T}(x_{it}-\overline{x}_i)(x_{it}-\overline{x}_i)^{'}$, $\xi_{1NT} = \frac{1}{N}\sum_{i=1}^{N}\zeta_{1iT}$, and $\xi_{2NT} = \frac{1}{N}\sum_{i=1}^{N}\zeta_{2iT}$. Before going into the next theorem, we need to consider some preliminary results. All limits in $(a)-(d)$ in Lemma 1 are taken as $T \to \infty$. Also, all limits in $(e)-(f)$ in Lemma 1 and Theorems $1-4$ are taken as $N \to \infty$ and $T \to \infty$.

Lemma 1 *If Assumptions* $1-4$ *hold, then*

(a) $\zeta_{1iT} \overset{d}{\to} \Omega_{\varepsilon}^{1/2}\left(\int \widetilde{W}_i dV_i\right)\Omega_{u.\varepsilon}^{1/2} + \Omega_{\varepsilon}^{1/2}\left(\int \widetilde{W}_i dW_i^{'}\right)\Omega_{\varepsilon}^{-1/2}\Omega_{\varepsilon u} + \Delta_{\varepsilon u} \equiv \zeta_{1i}$,

(b) $\zeta_{2iT} \overset{d}{\to} \Omega_{\varepsilon}^{1/2}\left(\int \widetilde{W}_i \widetilde{W}_i^{'}\right)\Omega_{\varepsilon}^{1/2} \equiv \zeta_{2i}$,

(c) $E[\zeta_{1i}] = -\frac{1}{2}\Omega_{\varepsilon u} + \Delta_{\varepsilon u}$,

(d) $E\left[\zeta_{2i}\right] = \frac{1}{6}\Omega_{\varepsilon}$,

(e) $\xi_{1NT} \xrightarrow{p} -\frac{1}{2}\Omega_{\varepsilon u} + \Delta_{\varepsilon u}$,

(f) $\xi_{2NT} \xrightarrow{p} \frac{1}{6}\Omega_{\varepsilon}$,

where $\widetilde{W}_i = W_i - \int W_i$.

Remark 2 $-\frac{1}{2}\Omega_{\varepsilon u}$ *is due to the endogeneity of the regressor* x_{it}, *and* $\Delta_{\varepsilon u}$ *is due to the serial correlation.*

Thus, we have established the following theorem:

Theorem 1 *If Assumptions* $1-4$ *hold, then*

(a) $T\left(\widehat{\beta}_{OLS} - \beta\right) \xrightarrow{p} -3\Omega_{\varepsilon}^{-1}\Omega_{\varepsilon u} + 6\Omega_{\varepsilon}^{-1}\Delta_{\varepsilon u}$,

(b) $\sqrt{N}T\left(\widehat{\beta}_{OLS} - \beta\right) - \sqrt{N}\delta_{NT} \xrightarrow{d} N\left(0, 6\Omega_{\varepsilon}^{-1}\Omega_{u.\varepsilon}\right)$,

where

$$\delta_{NT} = \left[\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T^2}\sum_{t=1}^{T}\left(x_{it} - x_{it}\right)\left(x_{it} - \overline{x}_i\right)'\right]^{-1}\left[\frac{1}{N}\sum_{i=1}^{N}\Omega_{\varepsilon}^{1/2}\left(\int \widetilde{W}_i dW_i'\right)\Omega_{\varepsilon}^{-1/2}\Omega_{\varepsilon u} + \Delta_{\varepsilon u}\right].$$

Remark 3 *We notice that* $\delta_{NT} \xrightarrow{p} -3\Omega_{\varepsilon}^{-1}\Omega_{\varepsilon u} + 6\Omega_{\varepsilon}^{-1}\Delta_{\varepsilon u}$.

Remark 4 *The normality of the OLS estimator comes naturally. When summing across i, the nonstandard asymptotic distribution due to unit root in the time dimension is smoothed out. However, it is important to note that the OLS estimator is asymptotically biased. The asymptotic bias is*

$$\widehat{\beta}_{OLS} - \beta \cong \frac{\delta_{NT}}{T} \cong \frac{-3\Omega_{\varepsilon}^{-1}\Omega_{\varepsilon u} + 6\Omega_{\varepsilon}^{-1}\Delta_{\varepsilon u}}{T}$$

which decreases as T *increases.*

Remark 5 $\Omega_{\varepsilon}^{-1}\Omega_{u.\varepsilon}$ *can be seen as the long-run signal-to-noise ratio.*

Remark 6 *If* $w_{it} = \left(u_{it}, \varepsilon_{it}'\right)'$ *are i.i.d., then*

$$\delta_{NT} \xrightarrow{p} 3\Sigma_{\varepsilon}^{-1}\Sigma_{\varepsilon u}$$

which was examined by Kao and Chen (1995).

Chen, McCoskey, and Kao (1996) investigated the finite sample proprieties of the OLS estimator in (6), the t-statistic, the bias-corrected OLS estimator, and the bias-corrected t-statistic. They found that the bias-corrected OLS estimator does not improve over the OLS estimator in general. The results of Chen, McCoskey, and Kao (1996) suggest that alternatives, such as the FM estimator or DOLS estimator (e.g., Saikkonen, 1991; Stock and Watson, 1993) may be more promising in cointegrated panel regressions. Thus, we begin our study by examining the limiting distribution of FM estimator, $\widehat{\beta}_{FM}$. Following Pedroni (1996), we begin our study by examining the limiting distribution of the FM estimator, $\widehat{\beta}_{FM}$. In contrast to Pedroni (1996), we initially consider the case where Ω is common across members of the panel in order to focus on the role that the signal to noise ratio, $\Omega_{\varepsilon}^{-1}\Omega_{u.\varepsilon}$, can play in the asymptotic distribution of an FM estimator.

The FM estimator is constructed by making corrections for endogeneity and serial correlation to the OLS estimator $\widehat{\beta}_{OLS}$ in (6). Let $\widehat{\Omega}_{\varepsilon u}$ and $\widehat{\Omega}_{\varepsilon}$ are consistent estimates of $\Omega_{u\varepsilon}$ and Ω_{ε}. Define

$$u_{it}^{+} = u_{it} - \Omega_{u\varepsilon}\Omega_{\varepsilon}^{-1}\varepsilon_{it},$$

$$\begin{aligned}\widehat{u}_{it}^{+} &= u_{it} - \widehat{\Omega}_{u\varepsilon}\widehat{\Omega}_{\varepsilon}^{-1}\varepsilon_{it},\\ y_{it}^{+} &= y_{it} - \Omega_{u\varepsilon}\Omega_{\varepsilon}^{-1}\varepsilon_{it},\end{aligned}$$

and

$$\widehat{y}_{it}^{+} = y_{it} - \widehat{\Omega}_{u\varepsilon}\widehat{\Omega}_{\varepsilon}^{-1}\varepsilon_{it}.$$

Note that

$$\begin{bmatrix} u_{it}^{+} \\ \varepsilon_{it} \end{bmatrix} = \begin{bmatrix} 1 & -\Omega_{u\varepsilon}\Omega_{\varepsilon}^{-1} \\ 0 & \mathbf{I}_k \end{bmatrix} \begin{bmatrix} u_{it} \\ \varepsilon_{it} \end{bmatrix},$$

which has the long-run covariance matrix

$$\begin{bmatrix} \Omega_{u.\varepsilon} & 0 \\ 0 & \Omega_{\varepsilon} \end{bmatrix},$$

where $\mathbf{I}_k$ is a $k \times k$ identity matrix. The endogeneity correction is achieved by modifying the variable y_{it} in (1) with the transformation

$$\begin{aligned}\widehat{y}_{it}^{+} &= y_{it} - \widehat{\Omega}_{u\varepsilon}\widehat{\Omega}_{\varepsilon}^{-1}\varepsilon_{it}\\ &= \alpha_i + x_{it}^{'}\beta + u_{it} - \widehat{\Omega}_{u\varepsilon}\widehat{\Omega}_{\varepsilon}^{-1}\varepsilon_{it}.\end{aligned}$$

The serial correlation correction term has the form

$$\begin{aligned}\widehat{\Delta}_{\varepsilon u}^{+} &= \begin{pmatrix} \widehat{\Delta}_{\varepsilon u} & \widehat{\Delta}_{\varepsilon} \end{pmatrix} \begin{pmatrix} 1 \\ -\widehat{\Omega}_{\varepsilon}^{-1}\widehat{\Omega}_{\varepsilon u} \end{pmatrix}\\ &= \widehat{\Delta}_{\varepsilon u} - \widehat{\Delta}_{\varepsilon}\widehat{\Omega}_{\varepsilon}^{-1}\widehat{\Omega}_{\varepsilon u},\end{aligned}$$

where $\widehat{\Delta}_{\varepsilon u}$ and $\widehat{\Delta}_{\varepsilon}$ are kernel estimates of $\Delta_{\varepsilon u}$ and Δ_{ε}. Therefore, the FM estimator is

$$\widehat{\beta}_{FM}=\left[\sum_{i=1}^{N}\sum_{t=1}^{T}\left(x_{it}-\overline{x}_{i}\right)\left(x_{it}-\overline{x}_{i}\right)^{'}\right]^{-1}\left[\sum_{i=1}^{N}\left(\sum_{t=1}^{T}\left(x_{it}-\overline{x}_{i}\right)\widehat{y}_{it}^{+}-T\widehat{\Delta}_{\varepsilon u}^{+}\right)\right]. \tag{7}$$

Now, we state the limiting distribution of $\widehat{\beta}_{FM}$.

Theorem 2 *If Assumptions* $1-4$ *hold, then* $\sqrt{N}T\left(\widehat{\beta}_{FM}-\beta\right)\xrightarrow{d}N\left(0,6\Omega_{\varepsilon}^{-1}\Omega_{u.\varepsilon}\right).$

Remark 7 *Note that Pedroni (1996) allowed the drifts for the integrated regressors in his cointegrated system. This paper only considers the regression in which integrated regressors do not have drifts. Also we propose the FM estimators for multiple regression.*

Next, we propose a DOLS estimator, $\widehat{\beta}_{D}$, which uses the past and future values of $\triangle x_{it}$ as additional regressors. We then show that the limiting distribution of $\widehat{\beta}_{D}$ is the same as the FM estimator, $\widehat{\beta}_{FM}$. But first, we need the following additional assumption:

Assumption 5 *The process* $\{u_{it}\}$ *can be projected on to* $\{\varepsilon_{it}\}$ *to get*

$$u_{it}=\sum_{j=-\infty}^{\infty}c_{ij}\varepsilon_{it+j}+v_{it}, \tag{8}$$

where

$$\sum_{j=-\infty}^{\infty}\|c_{ij}\|<\infty,$$

$\{v_{it}\}$ *is stationary with zero mean, and* $\{v_{it}\}$ *and* $\{\varepsilon_{it}\}$ *are uncorrelated not only contemporaneously but also in all lags and leads.*

Remark 8 *Assumption 5 can be guaranteed by following the conditions in Saikkonen (1991, p. 11).*

Remark 9 *In practice, the leads and lags may be truncated while retaining Assumption 5 approximately, so that*

$$u_{it}=\sum_{j=-q_1}^{q_2}c_{ij}\varepsilon_{it+j}+v_{it}.$$

This is because $\{c_{ij}\}$ *are assumed to be absolutely summable, i.e.,* $\sum_{j=-\infty}^{\infty}\|c_{ij}\|<\infty$.

We also need to require that q_1 and q_2 tend to infinity with T at a suitable rate, i.e.,

Assumption 6 $\frac{q_1^3}{T}\to 0, \frac{q_2^3}{T}\to 0,$ *and*

$$T^{1/2}\sum_{|j|>q_1\ or\ q_2}\|c_{ij}\|\to 0. \tag{9}$$

We then substitute (8) into (1) to have

$$y_{it} \;=\; \alpha_i \;+\; x_{it}^{'}\beta \;+\sum_{j=-q_1}^{q_2} c_{ij}\varepsilon_{it+j} + v_{it}.$$

Therefore, we obtain the DOLS of β, $\widehat{\beta}_D$, by running the following regression:

$$y_{it} \;=\; \alpha_i \;+\; x_{it}^{'}\beta \;+\sum_{j=-q_1}^{q_2} c_{ij}\triangle x_{it+j} + v_{it}. \tag{10}$$

Next, we show that $\widehat{\beta}_D$ has the same limiting distribution $\widehat{\beta}_{FM}$ as in Theorem 2.

Theorem 3 *If Assumptions* 1 − 6 *hold, then* $\sqrt{N}T\left(\widehat{\beta}_D - \beta\right) \xrightarrow{d} N\left(0, 6\Omega_{\varepsilon}^{-1}\Omega_{u.\varepsilon}\right).$

4 Heterogeneous Panels

The paper so far assumes that the panel data are homogeneous. The substantial heterogeneity exhibited by actual data in the cross-sectional dimension severely restricts the practical applicability of such estimators. Also, the estimators in Sections 2 and 3 are not easily extended to cases of broader cross-sectional heterogeneity since the variances and biases are specified in terms of the asymptotic covariance parameters that are assumed to be shared cross sectionally. Recently, Pedroni (1996) proposed an FM estimator for heterogeneous panels. Pedroni (1996) proposed the following panel FM estimator (using his notations):

$$\widehat{\beta}_{NT}^{*} - \beta = \left(\sum_{i=1}^{N}\widehat{L}_{22i}^{-2}\sum_{t=1}^{T}\left(x_{it}-\overline{x}_i\right)^2\right)^{-1}\sum_{i=1}^{N}\widehat{L}_{11i}^{-1}\widehat{L}_{22i}^{-1}\left(\sum_{t=1}^{T}\left(x_{it}-\overline{x}_i\right)u_{it}^{*} - T\widehat{\gamma}_i\right), \tag{11}$$

where $\widehat{L}_i$ is the lower triangular decomposition of a consistent estimator of the asymptotic covariance matrix

$$\Omega_i = \Omega_i^0 + \Gamma_i + \Gamma_i^{'},$$

and where u_{it}^{*} is given by

$$u_{it}^{*} = u_{it} - \frac{\widehat{L}_{21i}}{\widehat{L}_{22i}}\triangle x_{it}$$

and the serial correlation adjustment parameter $\widehat{\gamma}_i$ is given by

$$\widehat{\gamma}_i = \widehat{\Gamma}_{21i} + \widehat{\Omega}_{21i}^{0} - \frac{\widehat{L}_{21i}}{\widehat{L}_{22i}}\left(\widehat{\Gamma}_{21i} + \Omega_i^0\right).$$

Pedroni (1996) then derived the following result (his Proposition 1.2):

$$T\sqrt{N}\left(\widehat{\beta}_{NT}^{*} - \beta\right) \to N\left(0, v\right),$$

where

$$v = \begin{cases} 2 & \text{iff } \overline{x}_i = \overline{y}_i = 0 \\ 6 & \text{else} \end{cases}.$$

In this section, we propose an alternative representation of the panel FM estimator for heterogeneous panels. Again, in contract to Pedroni (1996), this section only considers the regression that integrated regressors do not have drifts. Also we propose an FM estimator for multiple regression. Before we discuss the FM estimator we need the following assumptions:

Assumption 7 *We assume the panels are heterogeneous, i.e., Ω_i, Γ_i and Σ_i are varied for different i. We also assume the invariance principle in (2), (8) in Assumption 5, and (9) in Assumption 6 still hold.*

Let

$$x_{it}^* = \widehat{\Omega}_{i\varepsilon}^{-1/2} x_{it}, \tag{12}$$

$$u_{it}^* = \widehat{\Omega}_{iu.\varepsilon}^{-1/2} \widehat{u}_{it}^+, \tag{13}$$

and

$$y_{it}^* = \widehat{\Omega}_{iu.\varepsilon}^{-1/2} \widehat{y}_{it}^+, \tag{14}$$

where $\widehat{\Omega}_{i\varepsilon}$ and $\widehat{\Omega}_{iu.\varepsilon}$ are consistent estimators of $\Omega_{i\varepsilon}$ and $\Omega_{iu.\varepsilon}$, respectively.

Assumption 8 *$\widehat{\Omega}_{i\varepsilon}$ is not singular for all i.*

Then, we define the FM estimator for heterogeneous panels as

$$\widehat{\beta}_{FM}^* = \left[\sum_{i=1}^{N}\sum_{t=1}^{T}\left(x_{it}^* - \overline{x}_i^*\right)\left(x_{it}^* - \overline{x}_i^*\right)'\right]^{-1}\left[\sum_{i=1}^{N}\left(\sum_{t=1}^{T}\left(x_{it}^* - \overline{x}_i^*\right) y_{it}^* - T\widehat{\Delta}_{i\varepsilon u}^+\right)\right], \tag{15}$$

where

$$\begin{aligned} \widehat{\Delta}_{i\varepsilon u}^+ &= \begin{pmatrix} \widehat{\Delta}_{i\varepsilon u} & \widehat{\Delta}_{i\varepsilon} \end{pmatrix} \begin{pmatrix} 1 \\ -\widehat{\Omega}_{i\varepsilon}^{-1}\widehat{\Omega}_{i\varepsilon u} \end{pmatrix} \\ &= \widehat{\Delta}_{i\varepsilon u} - \widehat{\Delta}_{i\varepsilon}\widehat{\Omega}_{i\varepsilon}^{-1}\widehat{\Omega}_{i\varepsilon u}. \end{aligned}$$

$\widehat{\beta}_{FM}^*$ can be written as

$$\begin{aligned} &\sqrt{N}T\left(\widehat{\beta}_{FM}^* - \beta\right) \\ &= \left[\tfrac{1}{N}\textstyle\sum_{i=1}^{N}\tfrac{1}{T^2}\sum_{t=1}^{T}\left(x_{it}^* - \overline{x}_i^*\right)\left(x_{it}^* - \overline{x}_i^*\right)'\right]^{-1}\left[\sqrt{N}\tfrac{1}{N}\textstyle\sum_{i=1}^{N}\left(\tfrac{1}{T}\sum_{t=1}^{T}\left(x_{it}^* - \overline{x}_i^*\right)u_{it}^* - T\widehat{\Delta}_{i\varepsilon u}^+\right)\right] \\ &= \left[\tfrac{1}{N}\textstyle\sum_{i=1}^{N}\zeta_{4iT}\right]^{-1}\left[\sqrt{N}\tfrac{1}{N}\textstyle\sum_{i=1}^{N}\zeta_{3iT}\right] \\ &= \left[\xi_{4NT}\right]^{-1}\sqrt{N}\xi_{3NT}, \end{aligned}$$

where $\overline{x}_i^* = \frac{1}{T}\sum_{t=1}^T x_{it}^*$, $\zeta_{3iT} = \frac{1}{T}\sum_{t=1}^T \left(x_{it}^* - \overline{x}_i^*\right) u_{it}^* - T\widehat{\Delta}_{i\varepsilon u}^+$, $\zeta_{4iT} = \frac{1}{T^2}\sum_{t=1}^T \left(x_{it}^* - \overline{x}_i^*\right)\left(x_{it}^* - \overline{x}_i^*\right)'$, $\xi_{3NT} = \frac{1}{N}\sum_{i=1}^N \zeta_{3iT}$, and $\xi_{4NT} = \frac{1}{N}\sum_{i=1}^N \zeta_{4iT}$. It is clear that from Lemma 1 that

$$\zeta_{3iT} \xrightarrow{d} \int \widetilde{W}_i dV_i,$$

$$\zeta_{4iT} \xrightarrow{d} \int \widetilde{W}_i \widetilde{W}_i{}',$$

$$\xi_{4NT} \xrightarrow{p} \frac{1}{6}\mathbf{I}_k,$$

and

$$\xi_{3NT} \xrightarrow{p} \frac{1}{6}\mathbf{I}_k.$$

It follows that

$$\sqrt{N}T\left(\widehat{\beta}_{FM}^* - \beta\right) \xrightarrow{d} N\left(0, 6\mathbf{I}_k\right).$$

Hence, we have established the following theorem:

Theorem 4 *If Assumptions* $7-8$ *hold, then* $\sqrt{N}T\left(\widehat{\beta}_{FM}^* - \beta\right) \xrightarrow{d} N\left(0, 6\mathbf{I}_k\right)$.

The DOLS estimator for heterogeneous panels, $\widehat{\beta}_D^*$, can be obtained by running the following regression:

$$y_{it}^* = \alpha_i + x_{it}^{*'}\beta + \sum_{j=-q_1}^{q_2} c_{ij}\triangle x_{it+j}^* + v_{it}^*.$$

It is straightforward to show that $\widehat{\beta}_D^*$ also has the same limiting distribution as $\widehat{\beta}_{FM}^*$.

Theorem 5 *If Assumptions* $7-8$ *hold, then* $\sqrt{N}T\left(\widehat{\beta}_D^* - \beta\right) \xrightarrow{d} N\left(0, 6\mathbf{I}_k\right)$.

Remark 10 *Theorems 4 and 5 show that the limiting distributions of* $\widehat{\beta}_{FM}^*$ *and* $\widehat{\beta}_D^*$ *are free of nuisance parameters.*

5 Hypothesis Testing

We now consider a linear hypothesis that involves the elements of the coefficient vector β. We show that hypothesis tests constructed using the FM and DOLS estimators have asymptotic chi-squared distributions. The null hypothesis has the form:

$$H_0 : R\beta = r, \tag{16}$$

where r is a $m \times 1$ known vector and R is a known $m \times k$ matrix describing the restrictions. A natural test statistic of the Wald test using $\widehat{\beta}_{FM}$ or $\widehat{\beta}_D$ for homogeneous panels is

$$W = \frac{1}{6} NT^2 \left(R\widehat{\beta}_{FM} - r \right)' \left[R\widehat{\Omega}_{\varepsilon}^{-1} \widehat{\Omega}_{u.\varepsilon} R' \right]^{-1} \left(R\widehat{\beta}_{FM} - r \right). \tag{17}$$

For the heterogeneous panels, a natural statistic using $\widehat{\beta}^*_{FM}$ or $\widehat{\beta}^*_D$ to test the null hypothesis is

$$W^* = \frac{1}{6} NT^2 \left(R\widehat{\beta}^*_{FM} - r \right)' \left[RR' \right]^{-1} \left(R\widehat{\beta}^*_{FM} - r \right). \tag{18}$$

It is clear that W and W^* converge in distribution to a chi-squared random variable with k degrees of freedom, χ^2_k, as $N \to \infty$ and $T \to \infty$ under the null hypothesis. Hence, we establish the following theorem:

Theorem 6 *If Assumptions* $1-8$ *hold, then under the null hypothesis (16),*

(a) $W \xrightarrow{d} \chi^2_k$,

(b) $W^* \xrightarrow{d} \chi^2_k$.

Remark 11 *Because the FM and the DOLS estimators have the same asymptotic distribution, it is easy to verify that the Wald statistics based on the FM estimator share the same limiting distributions with those based on the DOLS estimator.*

6 Monte Carlo Simulations

To compare the performance of OLS, FM, and DOLS estimators, we conducted Monte Carlo experiments based on the design similar to Phillips and Hansen (1990) and Phillips and Loretan (1991). The data generating process (DGP) was

$$y_{it} = \alpha_i + \beta x_{it} + u_{it}$$

and

$$x_{it} = x_{it-1} + \varepsilon_{it}$$

for $i = 1, ..., N$, $t = 1, ...T$, where

$$\begin{pmatrix} u_{it} \\ \varepsilon_{it} \end{pmatrix} = \begin{pmatrix} u^*_{it} \\ \varepsilon^*_{it} \end{pmatrix} + \begin{pmatrix} 0.3 & -0.4 \\ \theta_{21} & 0.6 \end{pmatrix} \begin{pmatrix} u^*_{it-1} \\ \varepsilon^*_{it-1} \end{pmatrix}$$

with

$$\begin{pmatrix} u^*_{it} \\ \varepsilon^*_{it} \end{pmatrix} \overset{iid}{\sim} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \sigma_{21} \\ \sigma_{21} & 1 \end{bmatrix} \right).$$

We generated α_i from a uniform distribution, $U[0,10]$, and set $\beta = 2$. We allowed θ_{21} and σ_{21} to vary and considered values of $\{0.8, 0.4, 0.0, -0.8\}$ for θ_{21} and $\{-0.8, -0.4, 0.4\}$ for σ_{21}. Random numbers for $(u_{it}^*, \varepsilon_{it}^*)$ were generated by the GAUSS procedure RNDNS. At each replication, we generated $N(T+1000)$ length of random numbers and then split it into N series so that each series had the same mean and variance. The first $1,000$ observations were discarded for each series. $\{u_{it}\}$ and $\{\varepsilon_{it}\}$ were constructed with $u_{i0} = 0$ and $\varepsilon_{i0} = 0$.

Once the estimates of w_{it}, $\widehat{w}_{it}$ were estimated, we used

$$\widehat{\Sigma} = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \widehat{w}_{it} \widehat{w}_{it}^{'} \tag{19}$$

to estimate Σ. Ω was estimated by

$$\widehat{\Omega} = \frac{1}{N} \sum_{i=1}^{N} \left\{ \frac{1}{T} \sum_{t=1}^{T} \widehat{w}_{it} \widehat{w}_{it}^{'} + \frac{1}{T} \sum_{\tau=1}^{l} \varpi_{\tau l} \sum_{t=\tau+1}^{T} \left(\widehat{w}_{it} \widehat{w}_{it-\tau}^{'} + \widehat{w}_{it-\tau} \widehat{w}_{it}^{'} \right) \right\}, \tag{20}$$

where $\varpi_{\tau l}$ is a weight function or a kernel. Using Phillips and Durlauf (1986) and the law of large numbers for triangular arrays, $\widehat{\Sigma}$ and $\widehat{\Omega}$ can be shown to be consistent for Σ and Ω. The estimate of the long-run covariance matrix in (20) was obtained by using the procedure KERNEL in COINT 2.0 with a Bartlett window of lag length five. Results with other kernels, such as Parzen and QS kernels, are not reported, because no essential differences were found for most cases.

Next, we recorded the results from our Monte Carlo experiments that examined the finite-sample properties of the OLS estimator, $\widehat{\beta}_{OLS}$; the FM estimator, $\widehat{\beta}_{FM}$; and the DOLS estimator, $\widehat{\beta}_{D}$. The simulations were performed by a Sun SparcServer 1,000. GAUSS 3.2.31 and COINT 2.0 were used to perform the simulations. The results we report are based on $10,000$ replications and are summarized in Tables 1 and 2. The FM estimator was obtained by using a Bartlett window of lag length five as in (20). Four lags and two leads were used for the DOLS estimator.

Table 1 reports the Monte Carlo means and standard deviations (in parentheses) of $\left(\widehat{\beta}_{OLS} - \beta\right)$, $\left(\widehat{\beta}_{FM} - \beta\right)$, and $\left(\widehat{\beta}_{D} - \beta\right)$ for sample sizes $T(N) = (20, 40, 60)$. The biases of the OLS estimator, $\widehat{\beta}_{OLS}$, decrease at a rate of T. For example, with $\sigma_{21} = -0.8$ and $\theta_{21} = 0.8$, the bias at $T = 20$ is -0.201 and at $T = 40$ is -0.104. Also, the biases increase in θ_{21} (if $\theta_{21} > 0$) and decrease in σ_{21}.

In general, the FM estimator, $\widehat{\beta}_{FM}$, presents the same degree of difficulty with bias as does the OLS estimator, $\widehat{\beta}_{OLS}$. For example, while the FM estimator, $\widehat{\beta}_{FM}$, reduces the bias substantially and outperforms $\widehat{\beta}_{OLS}$ when $\theta_{21} > 0$ and $\sigma_{21} < 0$, the opposite is true when $\theta_{21} > 0$ and $\sigma_{21} > 0$, the opposite is true. Likewise, when $\theta_{21} = -0.8$, $\widehat{\beta}_{FM}$ is less biased than $\widehat{\beta}_{OLS}$ for values of $\sigma_{21} = -0.8$. Yet, for values of $\sigma_{21} = -0.4$, the bias in $\widehat{\beta}_{OLS}$ is less than the bias in $\widehat{\beta}_{FM}$. There seems to be little to choose between $\widehat{\beta}_{OLS}$ and $\widehat{\beta}_{FM}$ when

$\theta_{21} < 0$. This is probably due to the failure of the semiparametric correction procedure in the presence of a negative serial correlation of the errors, i.e., a negative MA value, $\theta_{21} < 0$. Finally, for the cases where $\theta_{21} = 0.0$, $\widehat{\beta}_{FM}$ outperforms $\widehat{\beta}_{OLS}$ when $\sigma_{21} < 0$. On the other hand, $\widehat{\beta}_{FM}$ is more biased than $\widehat{\beta}_{OLS}$ when $\sigma_{21} > 0$.

In contrast, the results in Table 1 show that the DOLS, $\widehat{\beta}_D$, is distinctly superior to the OLS and FM estimators for all cases in terms of the mean biases. Clearly, the DOLS out-performed both the OLS and FM estimators.

While the limiting theory depends on the assumption that the cross-section and time-series dimensions are comparable in magnitude, the actual panel data have a wide variety of cross-section and time-series dimensions. It is important to know the effects of the variations in panel dimensions. Table 2 considers 20 different settings for N and T, each ranging from 20 to 120. First, we notice that the cross-section dimension has significant effect on the biases of $\widehat{\beta}_{OLS}$, $\widehat{\beta}_{FM}$, and $\widehat{\beta}_D$ when N is increased from 1 to 20. However, when N is increased from 20 to 40 and on, there is little effect on the biases of $\widehat{\beta}_{OLS}$, $\widehat{\beta}_{FM}$, and $\widehat{\beta}_D$. The results in Table 2 also confirm the superiority of the DOLS.

7 Conclusion

This paper derives limiting distributions for the OLS, FM, and DOLS estimators in a cointegrated regression and shows they are asymptotically normal.

We also investigated the finite sample proprieties of the OLS, FM, and DOLS estimators. Our findings are summarized as follows:

1. The OLS estimator has a non-negligible bias in finite samples.

2. The FM estimator does not improve over the OLS estimator in general.

3. The DOLS estimator may be more promising than OLS or FM estimators in estimating the cointegrated panel regressions.

Appendix

A Proof of Lemma 1

Proof. (a) and (b) are taken from the literature, i.e.,

$$
\begin{aligned}
\zeta_{2iT} &= \tfrac{1}{T^2}\sum_{t=1}^{T}(x_{it}-\overline{x}_i)(x_{it}-\overline{x}_i)' \\
&\xrightarrow{d} \Omega_{\varepsilon}^{1/2}\left(\int \widetilde{W}_i\widetilde{W}_i{}'\right)\Omega_{\varepsilon}^{1/2}=\zeta_{2i},
\end{aligned} \tag{21}
$$

where

$$\widetilde{W}_i = W_i - \int W_i,$$

and

$$
\begin{aligned}
&\zeta_{1iT} = \tfrac{1}{T}\sum_{t=1}^{T}(x_{it}-\overline{x}_i)u_{it} \\
&\xrightarrow{d} \Omega_{\varepsilon}^{1/2}\left(\int \widetilde{W}_i dV_i\right)\Omega_{u.\varepsilon}^{1/2} + \Omega_{\varepsilon}^{1/2}\left(\int \widetilde{W}_i dW_i'\right)\Omega_{\varepsilon}^{-1/2}\Omega_{\varepsilon u} + \Delta_{\varepsilon u} \\
&= \zeta_{1i},
\end{aligned}
$$

where $\triangle_{\varepsilon u} = \Sigma_{\varepsilon u} + \Gamma_{\varepsilon u}$.

The row i, column i diagonal element of $\int \widetilde{W}_i\widetilde{W}_i{}'$ is $\int W_i^2 - \left[\int W_i\right]^2$, and the row i, column j off-diagonal element is

$$\int W_iW_j - \left[\int W_i\right]\left[\int W_j\right].$$

Using

$$E\left[\int W_i^2 - \left[\int W_i\right]^2\right] = \frac{1}{6}$$

and

$$Var\left[\int W_i^2 - \left[\int W_i\right]^2\right] = \frac{1}{45},$$

the corresponding generalization is

$$E\left[\int \widetilde{W}_i\widetilde{W}_i{}'\right] = \frac{1}{6}\mathbf{I}_k,$$

where $\mathbf{I}_K$ is a $k \times k$ identity matrix. It then follows that

$$E[\zeta_{2i}] = \frac{1}{6}\Omega_{\varepsilon}^{1/2}\mathbf{I}_k\Omega_{\varepsilon}^{1/2} \tag{22}$$

$$= \frac{1}{6}\Omega_{\varepsilon} \tag{23}$$

establishing (d). To prove (c), we use the fact that

$$
\begin{aligned}
&E\left[\left(\int \widetilde{W}_i dV_i\right)\right] \\
&= 0
\end{aligned}
$$

and

$$
\begin{aligned}
&E\left[\int \widetilde{W}_i dW_i'\right] \\
&= -\tfrac{1}{2}\mathbf{I}_k
\end{aligned}
$$

resulting in

$$E\left[\zeta_{1i}\right] \quad = \quad -\tfrac{1}{2}\Omega_{\varepsilon u} + \Delta_{\varepsilon u}.$$

Finally,

$$\xi_{1NT} \equiv \frac{1}{N}\sum_{i=1}^{N}\zeta_{1iT} \xrightarrow{p} -\frac{1}{2}\Omega_{\varepsilon u} + \Delta_{\varepsilon u}$$

and

$$\xi_{2NT} \equiv \frac{1}{N}\sum_{i=1}^{N}\zeta_{2iT} \xrightarrow{p} \frac{1}{6}\Omega_{\varepsilon} \tag{24}$$

as required for (e) and (f) by the law of large numbers for triangular arrays. ■

B Proof of $Var\left[\Omega_{\varepsilon}^{1/2}\left(\int \widetilde{W}_i dV_i\right)\Omega_{u.\varepsilon}^{1/2}\right] = \frac{1}{6}\Omega_{u.\varepsilon}\Omega_{\varepsilon}$

Proof. Note that

$$\begin{aligned}
Var\left[\Omega_{\varepsilon}^{1/2}\left(\int \widetilde{W}_i dV_i\right)\Omega_{u.\varepsilon}^{1/2}\right] &= E\left[\Omega_{\varepsilon}^{1/2}\left(\int \widetilde{W}_i dV_i\right)\Omega_{u.\varepsilon}^{1/2}\right]\left[\Omega_{\varepsilon}^{1/2}\left(\int \widetilde{W}_i dV_i\right)\Omega_{u.\varepsilon}^{1/2}\right]^{'} \\
&= \Omega_{u.\varepsilon}\Omega_{\varepsilon}^{1/2}E\left[\left(\int \widetilde{W}_i dV_i\right)\left(\int \widetilde{W}_i dV_i\right)^{'}\right]\Omega_{\varepsilon}^{1/2} \\
&= \Omega_{u.\varepsilon}\Omega_{\varepsilon}^{1/2}\frac{1}{6}\mathbf{I}_k\Omega_{\varepsilon}^{1/2} \\
&= \frac{1}{6}\Omega_{u.\varepsilon}\Omega_{\varepsilon}
\end{aligned}$$

using $E\left[\left(\int \widetilde{W}_i dV_i\right)\left(\int \widetilde{W}_i dV_i\right)^{'}\right] = \frac{1}{6}\mathbf{I}_k$. ■

C Proof of Theorem 1

Proof. (a) is immediately evident from Lemma 1.

Recall that

$$\begin{aligned}
&\sqrt{N}T\left(\widehat{\beta}_{OLS} - \beta\right) - \left[\tfrac{1}{N}\textstyle\sum_{i=1}^{N}\tfrac{1}{T^2}\sum_{t=1}^{T}\left(x_{it} - \overline{x}_i\right)\left(x_{it} - \overline{x}_i\right)^{'}\right]^{-1}\sqrt{N}\tfrac{1}{N}\left[\textstyle\sum_{i=1}^{N}\Omega_{\varepsilon}^{1/2}\left(\int \widetilde{W}_i dW_i^{'}\right)\Omega_{\varepsilon}^{-1/2}\Omega_{\varepsilon u} + \Delta_{\varepsilon u}\right] \\
&= \left[\tfrac{1}{N}\textstyle\sum_{i=1}^{N}\zeta_{2iT}\right]^{-1}\left[\sqrt{N}\tfrac{1}{N}\textstyle\sum_{i=1}^{N}\left\{\zeta_{1iT} - \Omega_{\varepsilon}^{1/2}\left(\int \widetilde{W}_i dW_i^{'}\right)\Omega_{\varepsilon}^{-1/2}\Omega_{\varepsilon u} + \Delta_{\varepsilon u}\right\}\right] \\
&= \left[\xi_{2NT}\right]^{-1}\sqrt{N}\xi_{1NT}^{*}.
\end{aligned}$$

Note that the sequence $\left\{\zeta_{1iT} - \Omega_{\varepsilon}^{1/2}\left(\int \widetilde{W}_i dW_i^{'}\right)\Omega_{\varepsilon}^{-1/2}\Omega_{\varepsilon u} + \Delta_{\varepsilon u}\right\}$ in

$$\sqrt{N}\left(\xi_{1NT}^{*}\right) \quad = \quad \tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^{N}\left\{\zeta_{1iT} - \Omega_{\varepsilon}^{1/2}\left(\int \widetilde{W}_i dW_i^{'}\right)\Omega_{\varepsilon}^{-1/2}\Omega_{\varepsilon u} + \Delta_{\varepsilon u}\right\}$$

is a triangular array sequence; thus, a central limit theorem of a triangular array is needed. Recall that the variances are assumed to be the same across i,which implies that the Lindeberg condition will hold. We can readily see that the $\sqrt{N}\xi_{1NT}^{*}$ will converge to a normal variable with an appropriate normalization and that ξ_{2NT} will converge to $\frac{1}{6}\Omega_{\varepsilon}$ in probability by a law of large number for a triangular array. First, we note from Lemma 1 that

$$\zeta_{1iT}-\Omega_{\varepsilon}^{1/2}\left(\int\widetilde{W}_{i}dW_{i}^{'}\right)\Omega_{\varepsilon}^{-1/2}\Omega_{\varepsilon u}+\Delta_{\varepsilon u}\xrightarrow{d}\Omega_{\varepsilon}^{1/2}\left(\int\widetilde{W}_{i}dV_{i}\right)\Omega_{u.\varepsilon}^{1/2}.$$

It follows that

$$\frac{1}{N}\sum_{i=1}^{N}\left\{\zeta_{1iT}-\Omega_{\varepsilon}^{1/2}\left(\int\widetilde{W}_{i}dW_{i}^{'}\right)\Omega_{\varepsilon}^{-1/2}\Omega_{\varepsilon u}+\Delta_{\varepsilon u}\right\}\xrightarrow{p}E\left[\Omega_{\varepsilon}^{1/2}\left(\int\widetilde{W}_{i}dV_{i}\right)\Omega_{u.\varepsilon}^{1/2}\right]=0$$

and

$$Var\left[\Omega_{\varepsilon}^{1/2}\left(\int\widetilde{W}_{i}dV_{i}\right)\Omega_{u.\varepsilon}^{1/2}\right]=\frac{1}{6}\Omega_{u.\varepsilon}\Omega_{\varepsilon}$$

from Appendix B. Using the Slutsky theorem, we obtain

$$\left[\xi_{2NT}\right]^{-1}\sqrt{N}\xi_{1NT}^{*}\xrightarrow{d}N\left(0,6\Omega_{\varepsilon}^{-1}\Omega_{u.\varepsilon}\right).$$

.

Hence,

$$\begin{aligned}&\sqrt{N}T\left(\widehat{\beta}_{OLS}-\beta\right)-\sqrt{N}\delta_{NT}\\&\xrightarrow{d}N\left(0,6\Omega_{\varepsilon}^{-1}\Omega_{u.\varepsilon}\right),\end{aligned}\tag{25}$$

proving (b), where

$$\delta_{NT}=\left[\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T^{2}}\sum_{t=1}^{T}\left(x_{it}-\overline{x}_{i}\right)\left(x_{it}-\overline{x}_{i}\right)^{'}\right]^{-1}\frac{1}{N}\left[\sum_{i=1}^{N}\Omega_{\varepsilon}^{1/2}\left(\int\widetilde{W}_{i}dW_{i}^{'}\right)\Omega_{\varepsilon}^{-1/2}\Omega_{\varepsilon u}+\Delta_{\varepsilon u}\right].\tag{26}$$

Therefore, we establish Theorem 1. ■

D Proof of Theorem 2

The FM estimator of β can be rewritten as follows

$$\begin{aligned}\widehat{\beta}_{FM}&=\left[\sum_{i=1}^{N}\sum_{t=1}^{T}\left(x_{it}-\overline{x}_{i}\right)\left(x_{it}-\overline{x}_{i}\right)^{'}\right]^{-1}\left[\sum_{i=1}^{N}\left(\sum_{t=1}^{T}\left(x_{it}-\overline{x}_{i}\right)\widehat{y}_{it}^{+}-T\widehat{\Delta}_{\varepsilon u}^{+}\right)\right]\\&=\beta+\left[\sum_{i=1}^{N}\sum_{t=1}^{T}\left(x_{it}-\overline{x}_{i}\right)\left(x_{it}-\overline{x}_{i}\right)^{'}\right]^{-1}\left[\sum_{i=1}^{N}\left(\sum_{t=1}^{T}\left(x_{it}-\overline{x}_{i}\right)\widehat{u}_{it}^{+}-T\widehat{\Delta}_{\varepsilon u}^{+}\right)\right].\end{aligned}\tag{27}$$

First, we note that

$$\begin{aligned}\sqrt{N}T\left(\widehat{\beta}_{FM}-\beta\right) &= \left[\tfrac{1}{N}\sum_{i=1}^{N}\tfrac{1}{T^2}\sum_{t=1}^{T}\left(x_{it}-\overline{x}_i\right)\left(x_{it}-\overline{x}_i\right)^{'}\right]\sqrt{N}\tfrac{1}{N}\sum_{i=1}^{N}\tfrac{1}{T}\sum_{t=1}^{T}\left[\left(x_{it}-\overline{x}_i\right)\widehat{u}_{it}^{+}-\widehat{\Delta}_{\varepsilon u}^{+}\right] \\ &= \left[\tfrac{1}{N}\sum_{i=1}^{N}\zeta_{2iT}\right]^{-1}\left[\sqrt{N}\tfrac{1}{N}\sum_{i=1}^{N}\zeta_{1iT}^{**}\right] \\ &= \left[\xi_{2NT}\right]^{-1}\sqrt{N}\xi_{1NT}^{**},\end{aligned} \tag{28}$$

where $\zeta_{1iT}^{**}=\frac{1}{T}\sum_{t=1}^{T}\left[\left(x_{it}-\overline{x}_i\right)\widehat{u}_{it}^{+}-\widehat{\Delta}_{\varepsilon u}^{+}\right]$, and $\xi_{1NT}^{**}=\frac{1}{N}\sum_{i=1}^{N}\zeta_{1iT}^{**}$.

Let $w_{it}^{+}=\begin{pmatrix} u_{it}^{+} & \varepsilon_{it}^{'}\end{pmatrix}^{'}$ and we have

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]}w_{it}^{+}\overset{d}{\to}\begin{bmatrix} B_{ui}^{+} \\ B_{\varepsilon i}\end{bmatrix}\equiv BM\left(\Omega^{+}\right)\ \text{ as } T\to\infty, \tag{29}$$

where

$$\Omega^{+}=\begin{bmatrix}\Omega_{u.\varepsilon} & 0 \\ 0 & \Omega_{\varepsilon}\end{bmatrix}=\Sigma^{+}+\Gamma^{+}+\Gamma^{+'}$$

and

$$\begin{bmatrix} B_{ui}^{+} \\ B_{\varepsilon i}\end{bmatrix}=\begin{bmatrix} I & -\Omega_{u\varepsilon}\Omega_{\varepsilon}^{-1} \\ 0 & I\end{bmatrix}\begin{bmatrix} B_{ui} \\ B_{\varepsilon i}\end{bmatrix}.$$

Let

$$\Delta^{+}=\Sigma^{+}+\Gamma^{+}$$

and

$$\zeta_{1iT}^{+}=\frac{1}{T}\sum_{t=1}^{T}\left(x_{it}-\overline{x}_i\right)\widehat{u}_{it}^{+}.$$

From Lemma 1 and the consistency of $\widehat{\Omega}_{u\varepsilon}$ and $\widehat{\Omega}_{\varepsilon}^{-1}$ we note that

$$\zeta_{1iT}^{+}\overset{d}{\to}\int\widetilde{B}_{\varepsilon i}dB_{ui}^{+}+\Delta_{\varepsilon u}^{+},$$

where

$$\begin{aligned}\Delta_{\varepsilon u}^{+} &= \begin{pmatrix}\Delta_{\varepsilon u} & \Delta_{\varepsilon}\end{pmatrix}\begin{pmatrix} 1 \\ -\Omega_{\varepsilon}^{-1}\Omega_{\varepsilon u}\end{pmatrix} \\ &= \Delta_{\varepsilon u}-\Delta_{\varepsilon}\Omega_{\varepsilon}^{-1}\Omega_{\varepsilon u}.\end{aligned}$$

It follows in terms of standard Wiener processes that

$$\zeta_{1iT}^{+}\overset{d}{\to}\Omega_{\varepsilon}^{1/2}\left(\int\widetilde{W}_i dV_i\right)\Omega_{u.\varepsilon}^{1/2}+\Delta_{\varepsilon u}^{+}.$$

Now let

$$\zeta_{1iT}^{**} = \zeta_{1iT}^{+} - \widehat{\Delta}_{\varepsilon u}^{+}.$$

Clearly, from Lemma 1 we know that

$$E\left[\zeta_{1iT}^{**}\right] = 0$$

and

$$Var\left[\zeta_{1iT}^{**}\right] = \frac{1}{6}\Omega_{u.\varepsilon}\Omega_{\varepsilon}.$$

A similar argument to the proof of Theorem 1 yields

$$\sqrt{N}T\left(\widehat{\beta}_{FM} - \beta\right) \xrightarrow{d} N\left(0, 6\Omega_{\varepsilon}^{-1}\Omega_{u.\varepsilon}\right)$$

as required. ■

E Proof of Theorem 3

First we write (10) in vector from:

$$\begin{aligned} y_i &= e\alpha_i + x_i\beta + Z_{iq}C + v_i \\ &= x_i\beta + Z_iD + v_i \text{ (say)}, \end{aligned}$$

where y_i is a $T \times 1$ vector of y_{it}; e a is $T \times 1$ unit vector, Z_{iq} is the $T \times (q_1 + q_2)$ matrix of observations on the $q_1 + q_2$ regressors $\triangle x_{it-q_1}, \cdots, \triangle x_{it+q_2}$; x_i is a vector of $T \times k$ of x_{it}; C is a $(q_1 + q_2) \times 1$ vector of c_{ij}; v_i is a $T \times 1$ vector of v_{it}; Z_i is a $T \times (q_1 + q_2 + 1)$ matrix, $Z_i = (e, Z_{iq})$; and D is a $(q_1 + q_2 + 1) \times 1$ vector of parameters. Let $Q_i = I - Z_i(Z_i'Z_i)^{-1}Z_i'$. It follows that

$$\left(\widehat{\beta}_D - \beta\right) = \left[\sum_{i=1}^{N}\left(x_i'Q_ix_i\right)\right]^{-1}\left[\sum_{i=1}^{N}\left(x_i'Q_iv_i\right)\right].$$

We then write

$$\begin{aligned} &\sqrt{N}T\left(\widehat{\beta}_D - \beta\right) \\ &= \left[\tfrac{1}{N}\sum_{i=1}^{N}\tfrac{1}{T^2}\left(x_i'Q_ix_i\right)\right]^{-1}\left[\sqrt{N}\tfrac{1}{N}\sum_{i=1}^{N}\tfrac{1}{T}\left(x_i'Q_iv_i\right)\right] \\ &= \left[\tfrac{1}{N}\sum_{i=1}^{N}\zeta_{6iT}\right]^{-1}\left[\sqrt{N}\tfrac{1}{N}\sum_{i=1}^{N}\zeta_{5iT}\right] \\ &= \left[\xi_{6NT}\right]^{-1}\left[\sqrt{N}\xi_{5NT}\right], \end{aligned}$$

where $\xi_{5NT} = \frac{1}{N}\sum_{i=1}^{N}\zeta_{5iT}$, $\zeta_{5iT} = \frac{1}{T}\left(x_i'Q_iv_i\right)$, $\xi_{6NT} = \frac{1}{N}\sum_{i=1}^{N}\zeta_{6iT}$, and $\zeta_{6iT} = \frac{1}{T^2}\left(x_i'Q_ix_i\right)$.

Observe that from Saikkonen (1991)

$$
\begin{aligned}
\xi_{6NT} &= \frac{1}{N}\sum_{i=1}^{N}\zeta_{6iT} \\
&= \frac{1}{N}\sum_{i=1}^{N}\frac{1}{T^2}\left(x_i'Q_ix_i\right) \\
&= \frac{1}{N}\sum_{i=1}^{N}\frac{1}{T^2}\left(x_i^{'}W_Tx_i\right) + o_p(1) \\
&= \frac{1}{N}\sum_{i=1}^{N}\frac{1}{T^2}\sum_{t=q_1+1}^{T-q_2}\left(x_{it}-\overline{x}_i\right)\left(x_{it}-\overline{x}_i\right)^{'} + o_p(1),
\end{aligned}
$$

$$
\begin{aligned}
\sqrt{N}\xi_{5NT} &= \sqrt{N}\frac{1}{N}\sum_{i=1}^{N}\zeta_{5iT} \\
&= \sqrt{N}\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\left(x_i'Q_iv_i\right) \\
&= \sqrt{N}\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\left(x_i^{'}W_Tv_i\right) + o_p(1) \\
&= \sqrt{N}\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=q_1+1}^{T-q_2}\left(x_{it}-\overline{x}_i\right)v_{it} + o_p(1),
\end{aligned}
$$

and

$$
\frac{1}{T}\sum_{t=q_1+1}^{T-q_2}\left(x_{it}-\overline{x}_i\right)v_{it} \xrightarrow{d} \int \widetilde{B}_{\varepsilon i}dB_{ui}^{+},
$$

where $W_T = I_T - \frac{1}{T}ee^{'}$.

Using arguments similar to those in Theorem 2, we establish the following:

$$
\sqrt{N}T\left(\widehat{\beta}_D - \beta\right) \xrightarrow{d} N\left(0, 6\Omega_{\varepsilon}^{-1}\Omega_{u.\varepsilon}\right)
$$

as required. ■

References

[1] Billingsley, P. (1986), *Probability and Measure*, John Wiley, New York.

[2] Breitung, J., and Meyer, W. (1994), "Testing for Unit Roots in Panel Data: Are Wages on Different Bargaining Levels Cointegrated?" *Applied Economics*, 26, 353-361.

[3] Chen, B., McCoskey, S., and Kao, C. (1996), "Estimation and Inference of a Cointegrated Regression in Panel Data: A Monte Carlo Study," Manuscript, Center for Policy Research, Syracuse University.

[4] Im, K., Pesaran, H., and Shin, Y. (1995), "Testing for Unit Roots in Heterogeneous Panels," Manuscript, University of Cambridge.

[5] Kao, C. (1996), "Spurious Regression and Residual-Based Tests for Cointegration in Panel Data when the Cross-Section and Time-Series Dimensions are Comparable," Manuscript, Center for Policy Research, Syracuse University.

[6] Kao, C., and Chen, B. (1995), "On the Estimation and Inference for Cointegration in Panel Data when the Cross-Section and Time-Series Dimensions are Comparable," Manuscript, Center for Policy Research, Syracuse University.

[7] Levin, A., and Lin, C.-F. (1993), "Unit Root Tests in Panel Data: New Results," Discussion Paper, Department of Economics, UC-San Diego.

[8] Maddala, G. S., and Wu, S. (1996), "A Comparative Study of Unit Root Tests with Panel Data and a New Simple Test: Evidence From Simulations and the Bootstrap," Manuscript, Department of Economics, Ohio State University..

[9] McCoskey, S., and Kao, C. (1996), "Residuals-Based Tests of the Null of Cointegration in Panel Data," Manuscript, Center for Policy Research, Syracuse University.

[10] Park, J., and Ogaki, M. (1991), "Seemingly Unrelated Canonical Cointegrating Regressions," The Rochester Center for Economic Research, Working Paper No. 280.

[11] Pedroni, P. (1995), "Panel Cointegration: Asymptotics 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.

[12] Pedroni, P. (1996), "Fully Modified OLS for Heterogeneous Cointegrated Panels and the Case of Purchasing Power Parity," Indiana University Working Papers in Economics, No. 96-20.

[13] Phillips, P. C. B. (1995), "Fully Modified Least Squares and Vector Autoregression," *Econometrica*, 63, 1023-1078.

[14] Phillips, P. C. B., and Durlauf, S. N. (1986), "Multiple Time Series Regression with Integrated Processes," *Review of Economic Studies*, 53, 473-495.

[15] Phillips, P. C. B., and Hansen, B. E. (1990), "Statistical Inference in Instrumental Variables Regression with I(1) Processes," *Review of Economic Studies*, 57, 99-125.

[16] Phillips, P. C. B., and Loretan, M. (1991), "Estimating Long-Run Economic Equilibria," *Review of Economic Studies*, 58, 407-436.

[17] Phillips, P. C. B., and Solo, V. (1992), "Asymptotics for linear processes," *Annals of Statistics*, 20, 971-1001.

[18] Quah, D. (1994), "Exploiting Cross Section Variation for Unit Root Inference in Dynamic Data," *Economics Letters*, 44, 9-19.

[19] Saikkonen, P. (1991), "Asymptotically Efficient Estimation of Cointegrating Regressions," *Econometric Theory*, 58, 1-21.

[20] Stock, J., and Watson, M. (1993), "A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems," *Econometrica*, 61, 783-820.

Table 1: Means Biases and Standard Deviations of OLS, FM, and DOLS Estimators

	$\sigma_{21} = -0.8$			$\sigma_{21} = -0.4$			$\sigma_{21} = 0.8$		
	$\widehat{\beta}_{OLS}-\beta$	$\widehat{\beta}_{FM}-\beta$	$\widehat{\beta}_{D}-\beta$	$\widehat{\beta}_{OLS}-\beta$	$\widehat{\beta}_{FM}-\beta$	$\widehat{\beta}_{D}-\beta$	$\widehat{\beta}_{OLS}-\beta$	$\widehat{\beta}_{FM}-\beta$	$\widehat{\beta}_{D}-\beta$
$\theta_{21}= 0.8$									
T = 20	-.201	-.189	-.001	-.097	-.117	-.002	-.022	-.069	-.009
	(.049)	(.047)	(.040)	(.032)	(.035)	(.033)	(.011)	(.016)	(.009)
T = 40	-.104	-.099	-.000	-.049	-.060	-.001	-.011	-.034	-.004
	(.019)	(.017)	(.013)	(.012)	(.013)	(.011)	(.004)	(.006)	(.003)
T = 60	-.070	-.067	-.000	-.033	-.041	-.000	-.007	-.023	-.003
	(.010)	(.009)	(.007)	(.007)	(.007)	(.006)	(.002)	(.003)	(.002)
$\theta_{21}= 0.4$									
T = 20	-.132	-.079	-.001	-.082	-.075	-.002	-.014	-.068	-.003
	(.038)	(.026)	(.027)	(.030)	(.029)	(.031)	(.013)	(.018)	(.013)
T = 40	-.066	-.040	-.001	-.041	-.038	-.001	-.007	-.034	-.001
	(.014)	(.009)	(.027)	(.011)	(.011)	(.009)	(.005)	(.006)	(.004)
T = 60	-.044	-.027	.-.000	-.027	-.025	-.001	-.005	-.023	-.001
	(.007)	(.005)	(.005)	(.006)	(.006)	(.005)	(.002)	(.003)	(.002)
$\theta_{21}= 0.0$									
T = 20	-.079	-.016	.001	-.059	-.027	.002	.005	-.062	.006
	(.027)	(.015)	(.017)	(.026)	(.022)	(.026)	(.016)	(.019)	(.017)
T = 40	-.039	-.008	.001	-.059	-.027	.002	.001	-.031	.003
	(.009)	(.005)	(.005)	(.026)	(.022)	(.026)	(.003)	(.007)	(.005)
T = 60	-.026	-.005	.000	-.019	-.014	-.001	.001	-.021	.002
	(.005)	(.003)	(.003)	(.005)	(.008)	(.008)	(.003)	(.004)	(.003)
$\theta_{21}= -0.8$	.000								
T = 20	-.029	.029	.007	-.019	.029	.007	.114	.029	.000
	(.016)	(.012)	(.008)	(.017)	(.015)	(.014)	(.034)	(.027)	(.031)
T = 40	-.017	.015	.003	-.009	.015	.003	.057	.016	-.000
	(.006)	(.003)	(.002)	(.006)	(.005)	(.004)	(.012)	(.009)	(.009)
T = 60	-.009	.010	.002	-.007	.010	.002	.038	.011	.000
	(.003)	(.002)	(.001)	(.003)	(.002)	(.002)	(.007)	(.005)	(.005)

Note:
(a) N =T.
(b) The lag length 5 of the Bartlett windows is used for the FM estimator.
(c) The 4 lags and 2 leads are used for the DOLS estimator.

Table 2: Means Biases and Standard Deviations of OLS, FM, and DOLS Estimators for Different N and T

(N,T)	$\widehat{\beta}_{OLS}-\beta$	$\widehat{\beta}_{FM}-\beta$	$\widehat{\beta}_{D}-\beta$
(1,20)	-.135	-.122	-.007
	(.184)	(.189)	(.297)
(1,40)	-.070	-.065	-.001
	(.093)	(.092)	(.106)
(1,60)	-.047	-.043	-.001
	(.063)	(.061)	(.064)
(1,120)	-.024	-.022	-.001
	(.032)	(.031)	(.029)
(20,20)	-.082	-.075	-.002
	(.030)	(.029)	(.031)
(20,40)	-.042	-.039	.-.001
	(.016)	(.015)	(.014)
(20,60)	-.028	-.026	- .000
	(.010)	(.009)	(.009)
(20,120)	-.014	-.013	- .000
	(.005)	(.005)	(.005)
(40,20)	-.081	-.073	-.001
	(.022)	(.021)	(.022)
(40,40)	-.041	-.038	-.001
	(.011)	(.011)	(.009)
(40,60)	-.028	-.025	-.001
	(.007)	(.007)	(.007)
(40,120)	-.014	-.013	-.000
	(.004)	(.003)	(.003)
(60,20)	-.080	-.073	-.002
	(.017)	(.017)	(.018)
(60,40)	-.041	-.038	-.001
	(.009)	(.009)	(.008)
(60,60)	-.027	-.025	-.001
	(.006)	(.006)	(.005)
(60,120)	-.014	-.012	-.000
	(.003)	(.003)	(.003)
(120,20)	-.079	-.072	-.002
	(.012)	(.012)	(.012)
(120,40)	-.041	-.037	-.001
	(.006)	(.006)	(.006)
(120,60)	-.027	-.025	-.001
	(.004)	(.004)	(.004)
(120,120)	-.014	-.013	-.000
	(.002)	(.002)	(.002)

Note:
(a) The lag length 5 of the Bartlett windows is used for the FM estimator.
(b) The 4 lags and 2 leads are used for the DOLS estimator.