ELSEVIER

Journal of Econometrics 90 (1999) 1–44

JOURNAL OF
Econometrics

Spurious regression and residual-based tests for cointegration in panel data

Chihwa Kao*

Center for Policy Research and Department of Economics, Syracuse University, Syracuse, NY 13244-1020, USA

Received 10 January 1995; received in revised form 9 December 1997

Abstract

In the first half of the paper I study spurious regressions in panel data. Asymptotic properties of the least-squares dummy variable (LSDV) estimator and other conventional statistics are examined. The asymptotics of LSDV estimator are different from those of the spurious regression in the pure time-series. This has an important consequence for residual-based cointegration tests in panel data, because the null distribution of residual-based cointegration tests depends on the asymptotics of LSDV estimator.

In the second half of the paper I study residual-based tests for cointegration regression in panel data. I study Dickey–Fuller (DF) tests and an augmented Dickey–Fuller (ADF) test to test the null of no cointegration. Asymptotic distributions of the tests are derived and Monte Carlo experiments are conducted to evaluate finite sample properties of the proposed tests.

JEL classification: C22; C23

Keywords: Panel data; Spurious regression; LSDV; Sequential limit theory; Residual-based tests; ADF; DF

1. Introduction

Although there is immense interest in testing for unit roots and cointegration in time-series data, not much attention has been paid to testing the unit roots and cointegration of panel data at either empirical or theoretical levels.

* E-mail: cdkao@maxwell.syr.edu.

PII: S0304-4076(98)00023-2

Apparently, the only theoretical studies in this area are Breitung and Meyer (1994), Quah (1994), Levin and Lin (1993), Im, Pesaran and Shin (1997), Kao and Chen (1995a,b), Kao and Chiang (1997), McCoskey and Kao (1998), Pedroni (1995), and Phillips and Moon (1997). 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 the unit roots tests for panel data that have simultaneous extensive cross-section and time-series variation. Levin and Lin (1993) recently derived the limiting 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. (1997) critiqued the Levin and Lin panel unit root statistics and proposed alternatives. Kao and Chen (1995a) and Kao and Chiang (1997) studied the asymptotic results for a least-squares dummy variable (LSDV) estimator, a fully modified estimator and a dynamic least-squares estimator in a cointegrated regression in panel data. Kao and Chen (1995b) proposed residual-based tests for cointegration under a set of restricted assumptions. McCoskey and Kao (1998) proposed further tests for the null hypothesis of cointegration in panel data. Another paper dealing with residual-based tests in the presence of spurious regression is the paper by Pedroni (1995). In it, Pedroni allows different assumptions of the homogeneity and heterogeneity of the panel data. In particular, Pedroni introduces a model which allows for different intercepts and different slopes and heterogeneous long-run variance covariance matrices. Phillips and Moon (1997) developed a sequential limit theory for nonstationary panel data. Pesaran and Smith (1995) are not directly concerned with cointegration but do touch on a number of related issues, including the potential problems of homogeneity misspecification for cointegrated panels.

The first half of this paper examines a spurious regression in panel data. Asymptotic properties of the LSDV estimator and other conventional statistics are examined. I show that the LSDV estimator, $\hat{\beta}$, is consistent for its true value, but the t-statistic, t_{β}, diverges so that inferences about the regression coefficient, β, are wrong with the probability that goes to one asymptotically. The asymptotics of $\hat{\beta}$ are also different from those of the spurious regression in the pure time-series. This has an important consequence for residual-based cointegration tests in panel data, because the null distribution of residual-based cointegration tests depends on the asymptotics of $\hat{\beta}$. The second half of the paper examines the asymptotic null distribution of residual-based cointegration tests in panel data. I study Dickey–Fuller (DF) tests and an augmented Dickey–Fuller (ADF) test to test the null of no cointegration. Asymptotic distributions of the tests are derived and Monte Carlo experiments are conducted to evaluate finite sample properties of the proposed tests. The simulation results suggest that the DF_{ρ}^{*} and DF_{t}^{*} tests have better size and power properties than the DF_{ρ},DF_{t}, and ADF tests when σ is small. However, when σ is large, ADF clearly dominates the rest.

Section 2 introduces the model and derives the asymptotic distributions of the LSDV estimator and various conventional statistics from the spurious regression in panel data. Section 3 derives the asymptotic distributions for the residual-based tests for the cointegration. Section 4 describes the estimation of long-run variances. Section 5 reports the simulation results. An extension to allow nonzero drifts for the regressors is discussed in Section 6. Concluding remarks are given in Section 7. The proofs in the text are collected in the Appendix.

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

2. The model and assumptions

Suppose that $w_{it} = (u_{it}, \varepsilon_{it})'$ is a bivariate process with zero mean vector and the long-run covariance matrix of w_{it}:

$$\Omega = \lim_{T\to\infty} \frac{1}{T}\mathrm{E}\left(\sum_{t=1}^{T} w_{it}\right)\left(\sum_{t=1}^{T} w_{it}\right)' = \Sigma + \Gamma + \Gamma' = \begin{bmatrix} \sigma_{0u}^2 & \sigma_{0u\varepsilon} \\ \sigma_{0u\varepsilon} & \sigma_{0\varepsilon}^2 \end{bmatrix}, \tag{1}$$

where

$$\Gamma = \lim_{T\to\infty} \frac{1}{T}\sum_{k=1}^{T-1}\sum_{t=k+1}^{T} \mathrm{E}(w_{it}w_{it-k}') = \begin{bmatrix} \Gamma_u & \Gamma_{u\varepsilon} \\ \Gamma_{\varepsilon u} & \Gamma_{\varepsilon} \end{bmatrix} \tag{2}$$

and

$$\Sigma = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathrm{E}(w_{it}w_{it}') = \begin{bmatrix} \sigma_u^2 & \sigma_{u\varepsilon} \\ \sigma_{u\varepsilon} & \sigma_{\varepsilon}^2 \end{bmatrix}. \tag{3}$$

Let $y_{it} = \sum_{s=1}^{t} u_{is}$ and $x_{it} = \sum_{s=1}^{t} \varepsilon_{is}$ in which $u_{i0} = \varepsilon_{i0} = \mathrm{O_p}(1)$. Suppose y_{it} and x_{it}, are incorrectly estimated by least squares for all i using panel data; the spurious LSDV regression model is

$$y_{it} = \alpha_i + \beta x_{it} + e_{it}, \quad i = 1,\ldots,N,\ t = 1,\ldots,T, \tag{4}$$

where e_{it} is $I(1)$. The LSDV estimator of β is

$$\hat{\beta}=\frac{\sum_{i=1}^{N}\sum_{t=1}^{T}y_{it}(x_{it}-\bar{x}_i)}{\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}-\bar{x}_i)^2}. \tag{5}$$

Assumption 1. The asymptotic theory employed in this paper is a sequential limit theory established by Phillips and Moon (1997) in which $T\to\infty$ followed by $N\to\infty$ sequentially.

Remark 1. One of main issues in deriving the asymptotic distribution of the double indexed model is how to treat the two indices, T and N. In general, there are three types of convergence: (i) sequential limit, $T\to\infty$ first and then $N\to\infty$, (ii) pairwise limit, assuming $N(T)$ where $N(T)\to\infty$ as $T\to\infty$, and (iii) joint limit, N and $T\to\infty$. The joint limit is the strongest concept among the three. Some of the results in this paper will continue to hold for a pairwise limit or a joint limit theory if additional assumptions are imposed, e.g., one needs to restrict the rate at which N grows relative to T so as to guarantee that the appropriate terms do not diverge as N grows large.

Assumption 2. The $(u_{it},\varepsilon_{it})'$ are assumed to be independent across i.

Remark 2. It is true that the assumption of independence across i is rather strong. However, the assumption of independence across i is needed in order to satisfy the requirement of the Lindeberg–Levy central limit theorem. Moreover, as pointed out by Quah (1994), modeling cross-sectional dependence is involved because individual observations in cross-section have no natural ordering. It is possible to extend the current model to allow a degree of dependency across i using the approaches adapted by Im et al. (1995) and Pedroni (1995). However, it goes beyond the scope of this paper.

Remark 3. Indeed, assuming constant variances across i is also restrictive. The real job is how to verify the Lindeberg condition. The Lemma 4.2 (a version of the central limit theorem for triangular arrays) presented by Levin and Lin (1993) is not ready to derive the asymptotics when the variances are not constant across i. The difficulty is that uniform integrability is not guaranteed from the moment conditions in their lemma. Moreover, their conditions in the lemma do not yield the Lindeberg condition. Hence, before technical issues can be resolved, I think the assumption of constant variance across i is an acceptable compromise for the purpose of this paper.

Assumption 3. For each i, we assume

1. $\mathrm{E}[w_{it}] = 0$,
2. $\sup_t \mathrm{E}\,|w_{it}|^p$ for some $2 \leqslant p < \infty$,
3. $\{w_{it}\}_1^\infty$ is strong mixing with mixing number η_m satisfying $\sum_{m=1}^{\infty} \eta_m^{1-2/p} < \infty$.

It follows that w_{it} satisfies the invariance principle so that

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} w_{it} \Rightarrow \Omega^{1/2} W_i(r) \quad \text{as } T \to \infty, \tag{6}$$

for all i, where $W_i(r)$ is a 2×1 dimensional standard Wiener process with

$$W_i(r) = \begin{bmatrix} V_i(r) \\ W_i(r) \end{bmatrix}.$$

All limits in Theorems 1–4 and Corollaries 1 and 2 are taken as $T \to \infty$ and followed by $N \to \infty$ sequentially by Assumption 1. The asymptotic properties of the LSDV estimator and other conventional statistics are given in the following theorem:

Theorem 1. Assume that Assumptions 1–3 are satisfied; *then*

(a) $\hat{\beta} \xrightarrow{p} \sigma_{0u\varepsilon}/\sigma_{0\varepsilon}^2$,
(b) $\sqrt{N}(\hat{\beta} - \sigma_{0u\varepsilon}/\sigma_{0\varepsilon}^2) \Rightarrow \mathrm{N}(0, 2\sigma_{0v}^2/5\sigma_{0\varepsilon}^2)$,
(c) $T^{-1/2} t_{\beta} - T^{-1/2} \sigma_{0\varepsilon}^{-2} \sigma_{0u\varepsilon} \sqrt{\sum_{i=1}^N \sum_{t=1}^T (x_{it} - \bar{x}_i)^2} / s \Rightarrow \mathrm{N}(0, 2/5)$,
(d) $R^2 \xrightarrow{p} \dfrac{\sigma_{0u\varepsilon}^2}{\sigma_{0v}^2 \sigma_{0\varepsilon}^2 + \sigma_{0u\varepsilon}^2}$,
(e) $DW = \mathrm{O_p}(T^{-1})$,
(f) $TDW \xrightarrow{p} 6$,

where

$$s = \sqrt{\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \{(y_{it} - \bar{y}_i) - \hat{\beta}(x_{it} - \bar{x}_i)\}^2}.$$

Remark 4. Theorem 1 shows that the LSDV estimator, $\hat{\beta}$, is consistent for its true value, $\sigma_{0u\varepsilon}/\sigma_{0\varepsilon}^2$. It is known that if a time-series regression for give i is performed in model (4), then the ordinary least squares (OLS) estimator of β is spurious (It is due to that the noise, e_{it}, is as strong as the signal, x_{it}, since both e_{it} and x_{it} are $I(1)$.) In the panel regression (4) with a large number of cross-section data, the strong noise of e_{it} is attenuated by pooling the data and a consistent estimate of the signal can be extracted. However, the t-statistic, t_β, diverges so that inferences about the regression coefficient, β, are wrong with the probability that goes to one asymptotically.

If we assume strong exogeneity and no serial correlation, then Theorem 1 is reduced to the following corollary which was investigated by Kao and Chen (1995b):

Corollary 1. If $u_{it} \sim iid(0, \sigma_u^2)$, $\varepsilon_{it} \sim iid(0, \sigma_\varepsilon^2)$, $\mathrm{E}(u_{it}^4) < \infty$, $\mathrm{E}(\varepsilon_{it}^4) < \infty$, *and* u_{it} *and* ε_{it} *are independent, then*

(a) $\hat{\beta} \xrightarrow{p} 0$,
(b) $\sqrt{N}\hat{\beta} \Rightarrow \mathrm{N}\left(0, \dfrac{2\sigma_u^2}{5\sigma_\varepsilon^2}\right)$ or $\hat{\beta} = \mathrm{O_p}(N^{-1/2})$,
(c) $T^{-1/2}t_\beta \Rightarrow \mathrm{N}(0,\frac{2}{5})$ or $t_\beta = \mathrm{O_p}(T^{1/2})$,
(d) $R^2 \xrightarrow{p} 0$,
(e) $DW = \mathrm{O_p}(T^{-1})$,
(f) $TDW \xrightarrow{p} 6$.

Remark 5. The results of Theorem 1 can also be extended to multiple regression, provided that $\{x_{it}\}$ are not cointegrated.

The asymptotics of $\hat{\beta}$ are different from those of a spurious regression in the pure time-series, and this difference has an important consequence for residual-based cointegration tests using panel data, because the null distribution of residual-based cointegration tests depends on the asymptotics of $\hat{\beta}$. This point is explained further in the next section.

3. Residual-based tests for cointegration

In this section we derive the limiting distributions of residual-based cointegration tests using DF tests and ADF when applied to the model (4) in Section 2.

3.1. Dickey–Fuller (DF) test

The DF test can be applied to the residuals using

$$\hat{e}_{it} = \rho\hat{e}_{it-1} + v_{it}, \tag{7}$$

where $\hat{e}_{it}$ is the estimate of e_{it} from Eq. (4). The OLS of ρ, $\hat{\rho}$, is

$$\hat{\rho} = \frac{\sum_{i=1}^{N}\sum_{t=2}^{T}\hat{e}_{it}\hat{e}_{it-1}}{\sum_{i=1}^{N}\sum_{t=2}^{T}\hat{e}_{it-1}^{2}}. \tag{8}$$

The null hypothesis that $\rho = 1$ in regression (7) is tested by

$$\sqrt{N}T(\hat{\rho}-1) = \frac{\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=2}^{T}\hat{e}_{it-1}\Delta\hat{e}_{it}}{\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T^2}\sum_{t=2}^{T}\hat{e}_{it-1}^{2}}$$

$$= \frac{\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=2}^{T}\hat{e}_{it-1}^{*}\Delta\hat{e}_{it}^{*}}{\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T^2}\sum_{t=2}^{T}(\hat{e}_{it-1}^{*})^{2}}$$

$$= \frac{\sqrt{N}\frac{1}{N}\sum_{i=1}^{N}\zeta_{3iT}}{\frac{1}{N}\sum_{i=1}^{N}\zeta_{4iT}}$$

$$= \frac{\sqrt{N}\xi_{3NT}}{\xi_{4NT}},$$

where $\hat{e}_{it}^{*}$ is defined in Eq. (A.3), $\zeta_{3iT} = (1/T)\sum_{t=2}^{T}\hat{e}_{it-1}^{*}\Delta\hat{e}_{it}^{*}$, $\zeta_{4iT} = (1/T^2)\sum_{t=2}^{T}\hat{e}_{it-1}^{*2}$, $\xi_{3NT} = (1/N)\sum_{i=1}^{N}\zeta_{3iT}$, and $\xi_{4NT} = (1/N)\sum_{i=1}^{N}\zeta_{4iT}$. The t-statistic to test $\rho = 1$, t_ρ, is

$$t_\rho = \frac{(\hat{\rho}-1)\sqrt{\sum_{i=1}^{N}\sum_{t=2}^{T}\hat{e}_{it-1}^{*2}}}{s_e},$$

where $s_e^2 = (1/NT)\sum_{i=1}^{N}\sum_{t=2}^{T}(\hat{e}_{it}^{*} - \hat{\rho}\hat{e}_{it-1}^{*})^2$.

Theorem 2. Suppose Assumptions 1–3 *hold. Then,*

$$\sqrt{N}T(\hat{\rho}-1) - \frac{\sqrt{N}\mu_{3T}}{\mu_{4T}} \Rightarrow \mathrm{N}\left(0,\ 3 + \frac{36}{5}\frac{\sigma_v^4}{\sigma_{0v}^4}\right)$$

and

$$t_\rho - \frac{\sqrt{N}\mu_{3T}}{s_e\sqrt{\mu_{4T}}} \Rightarrow \mathrm{N}\left(0, \frac{\sigma_{0v}^2}{2\sigma_v^2} + \frac{3\sigma_v^2}{10\sigma_{0v}^2}\right),$$

where $\mu_{3T} = \mathrm{E}[\zeta_{3iT}]$ *and* $\mu_{4T} = \mathrm{E}[\zeta_{4iT}]$.

Remark 6. Theorem 2 shows that the asymptotic distributions test statistics, $\sqrt{N}T(\hat{\rho}-1)$ and t_ρ depend on the nuisance parameters: μ_{3T}, μ_{4T}, σ_v^2 and σ_{0v}^2. First we note that μ_{3T} and μ_{4T} can be approximated by $\mu_3 = -\sigma_v^2/2$ and $\mu_4 = \sigma_{0v}^2/6$, respectively. It is necessary to construct new statistics whose limiting distributions are not dependent on $\mu_{3T}, \mu_{4T}, \sigma_v^2$ and σ_{0v}^2. We now define

$$DF_\rho^* = \frac{\sqrt{N}T(\hat{\rho}-1) + \dfrac{3\sqrt{N}\hat{\sigma}_v^2}{\hat{\sigma}_{0v}^2}}{\sqrt{3 + \dfrac{36\hat{\sigma}_v^4}{5\hat{\sigma}_{0v}^4}}},$$

and

$$DF_t^* = \frac{t_\rho + \dfrac{\sqrt{6N}\hat{\sigma}_v}{2\hat{\sigma}_{0v}}}{\sqrt{\dfrac{\hat{\sigma}_{0v}^2}{2\hat{\sigma}_v^2} + \dfrac{3\hat{\sigma}_v^2}{10\hat{\sigma}_{0v}^2}}},$$

where $\hat{\sigma}_v^2$ and $\hat{\sigma}_{0v}^2$ are consistent estimates of σ_v^2 and σ_{0v}^2. The limiting distributions of DF_ρ^* and DF_t^*, by construction, do not depend on σ_v^2 and σ_{0v}^2. It can be shown easily that $DF_\rho^* \Rightarrow \mathrm{N}(0,1)$ and $DF_\rho^* \Rightarrow \mathrm{N}(0,1)$ by the sequential limit theory.

Remark 7. Alternatively, we could define a bias-corrected serial correlation coefficient estimate,

$$\hat{\rho}^* = \frac{\sum_{i=1}^N \sum_{t=2}^T (\hat{e}_{it}\hat{e}_{it-1} - \mu_3)}{\sum_{i=1}^N \sum_{t=2}^T \hat{e}_{it-1}^2}$$

then it can be shown that the bias-corrected test statistics are

$$\sqrt{N}T(\hat{\rho}^* - 1) \Rightarrow \mathrm{N}(0,3)$$

and

$$t_\rho^* \Rightarrow \mathrm{N}\left(0, \frac{\sigma_{0v}^2}{2\sigma_v^2}\right)$$

by following Eqs. (C.8) and (C.14).

Note that with strong exogeneity and in the absence of serial correlation we have $\sigma_u^2 = \sigma_{0u}^2 = \sigma_v^2 = \sigma_{0v}^2$. This case was previously studied by Kao and Chen (1995b). We then have the following corollary:

Corollary 2. Suppose strong exogeneity holds and there is no serial correlation. Then,

$$\sqrt{N}T(\hat{\rho}-1) - \frac{\sqrt{N}\mu_{5T}}{\mu_{6T}} \Rightarrow \mathrm{N}\left(0, \frac{51}{5}\right)$$

and

$$\sqrt{5/4}\left\{t_\rho - \frac{\sqrt{N}\mu_{5T}}{\sqrt{\mu_{6T}}}\right\} \Rightarrow \mathrm{N}(0,1).$$

3.2. Augmented Dickey–Fuller (ADF) test

The DF test in Section 3.1 was based on a simple OLS regression of $\hat{e}_{it}$ on its own lagged value. Correction for serial correlation was made to the OLS and t-statistic. Alternatively, the lagged changes in the residuals can be added to the regression of (7):

$$\hat{e}_{it} = \rho\hat{e}_{it-1} + \sum_{j=1}^{p} \varphi_j \Delta\hat{e}_{it-j} + v_{itp}, \tag{9}$$

where p is chosen so that the residuals v_{itp} are serially uncorrelated. The ADF test statistic discussed in this section is the usual t-statistic of $\rho = 1$ in regression (9). Let X_{ip} and X_{ip}^* be the matrices of observation on the p regressors $(\Delta\hat{e}_{it-1}, \Delta\hat{e}_{it-2}, \ldots, \Delta\hat{e}_{it-p})$ and $(\Delta\hat{e}_{it-1}^*, \Delta\hat{e}_{it-2}^*, \ldots, \Delta\hat{e}_{it-p}^*)$ respectively, e_i and e_i^* equal the vector of observations of $\hat{e}_{it-1}$ and $\hat{e}_{it-1}^*$, respectively, v_i equal the vector of observations of v_{itp}, $Q_i = I - X_{ip}(X_{ip}'X_{ip})^{-1}X_{ip}'$, $Q_i^* = I - X_{ip}^*(X_{ip}^{*\prime}X_{ip}^*)^{-1}X_{ip}^{*\prime}$, and $s_v^2 = (1/NT)\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{v}_{itp}^2$, where $\hat{v}_{itp}$ is an estimate of v_{itp}. Under the null hypothesis of no cointegration, the *ADF* test takes the form

$$t_{ADF} = \frac{(\hat{\rho}-1)\left[\sum_{i=1}^{N}(e_i'Q_ie_i)\right]^{1/2}}{s_v},$$

where $\hat{\rho}$ is the OLS estimate of ρ. Note that

$$(\hat{\rho}-1) = \left[\sum_{i=1}^{N}(e_i'Q_ie_i)\right]^{-1}\left[\sum_{i=1}^{N}(e_i'Q_iv_i)\right].$$

Now,

$$
\begin{aligned}
\mathrm{t}_{ADF} &= \frac{\sum_{i=1}^{N}(e_i' Q_i v_i)}{s_v\sqrt{\sum_{i=1}^{N}(e_i' Q_i e_i)}} \\
&= \frac{\sqrt{N}\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}(e_i' Q_i v_i)}{s_v\sqrt{\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T^2}(e_i' Q_i e_i)}} \\
&= \frac{\sqrt{N}\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}(e_i^{*\prime} Q_i^* v_i)}{s_v\sqrt{\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T^2}(e_i^{*\prime} Q_i^* e_i^*)}} \\
&= \frac{\sqrt{N}\xi_{5NT}}{s_v\sqrt{\xi_{6NT}}},
\end{aligned}
$$

where $\xi_{5NT} = (1/N)\sum_{i=1}^{N}\zeta_{5iT}$, $\zeta_{5iT} = (1/T)(e_i^{*\prime} Q_i^* v_i)$, $\xi_{6NT} = (1/N)\sum_{i=1}^{N}\zeta_{6iT}$, and $\zeta_{6iT} = (1/T^2)(e_i^{*\prime} Q_i^* e_i^*)$.

Theorem 3. Suppose that the assumptions of Theorem 2 *are satisfied. Then, we have*

$$
t_{ADF} - \frac{\sqrt{N}\mu_{5T}}{s_v\sqrt{\mu_{6T}}} \Rightarrow \mathrm{N}\left(0, \frac{\sigma_{0v}^2}{2\sigma_v^2} + \frac{3\sigma_v^2}{10\sigma_{0v}^2}\right),
$$

provided $p \to \infty$ *as* $T \to \infty$ *such that* $p = \mathrm{o}(T^{1/3})$, *where* $\mu_{5T} = \mathrm{E}[\zeta_{5iT}]$ *and* $\mu_{6T} = E[\zeta_{6iT}]$.

Remark 8. Similarly, we can define

$$
ADF = \frac{t_{ADF} + \sqrt{6N}\hat{\sigma}_v/2\hat{\sigma}_{0u}}{\sqrt{\hat{\sigma}_{0v}^2/2\hat{\sigma}_v^2 + 3\hat{\sigma}_v^2/10\hat{\sigma}_{0v}^2}}
$$

which does not depend on the nuisance parameters. We thus have $ADF \Rightarrow$ N(0,1) by the sequential limit theory.

Remark 9. Alternatively, we could follow Pedroni (1995) to remove the nuisance parameters in the limiting distribution of t_{ADF} as follows:

$$\frac{\sigma_v}{s_v}\left(t_{ADF} - \frac{\sqrt{N}\lambda}{\sigma_{ov}\sqrt{\xi_{6NT}}} + \frac{\sqrt{6N}}{2}\right) \Rightarrow \mathrm{N}(0, \tfrac{4}{5}),$$

where $\lambda = (\sigma_{ov}^2 - \sigma_v^2)/2$.

Remark 10. This paper is related to recent work by Pedroni (1995) that came to my attention when this work was completed. Pedroni (1995) derived asymptotic distributions for residual-based tests of cointegration for both homogenous and heterogeneous panels. This paper discusses the fixed effect model, which Pedroni (1995) does not discuss. Using a pooled unstandardized version of the Phillips–Perron statistics, Pedroni finds that models with common slopes and homogeneous errors produce statistics which converge in distribution to normal random variables. This result is analogous to the results in Theorem 2, except that Pedroni's results (see his Proposition 2) are based on the overly restrictive assumption that the regressors are strictly exogenous.

4. Estimation of long-run parameters

The asymptotic distributions given in Theorems 2 and 3 depend on unknown parameters ($\sigma_u^2, \sigma_\varepsilon^2, \sigma_{u\varepsilon}, \sigma_{0\varepsilon}^2, \sigma_{0u}^2$ and $\sigma_{0u\varepsilon}$). Once the estimates of $w_{it} = (u_{it}, \varepsilon_{it})'$, $\hat{w}_{it}$, are obtained, we can use

$$\hat{\Sigma} = \begin{pmatrix} \hat{\sigma}_u^2 & \hat{\sigma}_{u\varepsilon} \\ \hat{\sigma}_{u\varepsilon} & \hat{\sigma}_\varepsilon^2 \end{pmatrix} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} \hat{w}_{it}\hat{w}_{it}' \tag{10}$$

to estimate Σ. Ω can be estimated by

$$\hat{\Omega} = \begin{pmatrix} \hat{\sigma}_{0u}^2 & \hat{\sigma}_{0u\varepsilon} \\ \hat{\sigma}_{0u\varepsilon} & \hat{\sigma}_{0\varepsilon}^2 \end{pmatrix} \tag{11}$$

$$= \frac{1}{N}\sum_{i=1}^{N}\left\{\frac{1}{T}\sum_{t=1}^{T} \hat{w}_{it}\hat{w}_{it}' + \frac{1}{T}\sum_{\tau=1}^{l} \varpi_{\tau l} \sum_{t=\tau+1}^{T} (\hat{w}_{it}\hat{w}_{it-\tau}' + \hat{w}_{it-\tau}\hat{w}_{it}')\right\} \tag{12}$$

where $\varpi_{\tau l}$ is a weight function or a kernel. Usually kernels are truncated by the bandwidth parameter l so that $\varpi_{\tau l} = 0$ for $\tau > l$. Using Phillips and Durlauf (1986), a law of large numbers and the sequential limit theory, $\hat{\Sigma}$ and $\hat{\Omega}$ can be shown to be consistent for Σ and Ω.

Remark 11. The distribution results for DF_{ρ}^{*}, DF_{t}^{*}, and ADF may require $N = o(T^{2\gamma})$ if a pairwise limit or a joint limit theory is used, where γ is the convergence rate of the spectral density estimator (e.g., $\gamma = 1/5$ for the Barlett kernel, 1/3 for the quadratic spectral kernel, and up to 1/2 for parametric kernel estimators). For example, for the quadratic spectral kernel with optimal selection of the bandwidth parameter, $\hat{\Omega}$ has bias of order $O(T^{-1/3})$. Since this bias is due to the declining weights assigned by the kernel to the true autocovariances, the bias does not shrink with the cross-sectional dimension N. Thus the asymptotic distributions presented in this paper will be appropriate if either one treats the asymptotics sequentially or one restricts the rate at which N grows relative to T so to guarantee that $\sqrt{N}(\hat{\Omega} - \Omega)$, for example, does not diverge as N grows large. If the sequential limit theory is used then the difference between $\hat{\Omega}$ and Ω will be eliminated in the first step when $T \to \infty$.

5. Some Monte Carlo results

5.1. Spurious regression

In this section, I report some results from Monte Carlo experiments that examined the finite-sample properties of the LSDV estimator, $\hat{\beta}$, t_{β}, R^2, and DW. The simulations were performed by a Sun SparcServer 1000 using the software GAUSS 3.27. The results I report are based on 10 000 replications. The data were generated by creating $T + 1000$ observations and discarding the first 1000 observations to remove the effect of the initial conditions.

The data generating process (DGP) was $y_{it} = y_{it-1} + u_{it}$ and $x_{it} = x_{it-1} + \varepsilon_{it}$ for $i = 1, \ldots, N$ and $t = 1, \ldots, T$, where the innovation sequences $(u_{it}, \varepsilon_{it})$ were generated from a bivariate normal with independence across both individuals and time periods, i.e.,

$$\begin{bmatrix} u_{it} \\ \varepsilon_{it} \end{bmatrix} \overset{\text{iid}}{\sim} \mathrm{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \right). \tag{13}$$

Eq. (13) assumes strong exogeneity and no serial correlation. Random numbers for $(u_{it}, \varepsilon_{it})$ were generated by the GAUSS procedure RNDNS. For each experiment, we calculated the mean values of $\hat{\beta}$, t_{β}, R^2, DW, the mean estimated standard error (ESE), $P(|t| > 1.96)$ (the rejection frequency of the t-test of H_0: $\beta = 0$), and the sample standard deviation (SSD) of $\hat{\beta}$ where

$$SSD(\hat{\beta}) = \left\{ \frac{1}{M-1} \sum_{i=1}^{M} [\hat{\beta}_i - \bar{\beta}]^2 \right\}^{1/2},$$

$\bar{\beta} = (1/M)\sum_{i=1}^{M}\hat{\beta}_i$, and M was the number of replications. For sample size, I considered different settings for T and N.

The Monte Carlo results are reported in Tables 1–3. In the case where $N = T = 30$, the probability of rejecting H_0, $P(|t_\beta| > 1.96)$ is approximately 0.5723, which is far from the 'conventional' value of 0.05. Similar discrepancies exist for all our other values of N and T. This means that even when the null hypothesis is true, we will wrongly reject it most of the time.

We know from Table 1 that the mean values of $\hat{\beta}$ almost always stay constant around zero, but $\bar{\beta} \pm 2 \times SSD$ converges to the mean value as panel size increases. Unlike in the pure time-series case, $\hat{\beta}$ converges to zero in probability instead of being a random variable. Table 1 also reveals some discrepancy between the SSD and the ESE. However, the discrepancy between these two measures is not as severe as the discrepancy in the pure time-series because both the SSD and the ESE diminish as T increases in panel data.

The mean value of the t-statistic changes little as $T(N)$ grows from 10 to 70, but the SSD increases rapidly. The likelihood of $|t_\beta| > 1.96$ becomes higher and higher so that the critical value at the 5% significance level increases as T increases. Therefore, as in a pure time-series, the spurious regression problem becomes worse as the panel size grows. This is confirmed by Table 1, which records the rejection frequency of the t-test for the sample sizes from 10 to 80. The rejection rate of the null hypothesis of no relationship between x_{it} and y_{it} increases steadily with T. At $N = T = 50$; the rejection rate is as high as 66%.

I also examine how the finite sample properties of the spurious LSDV regression change if N and T are different. I first fix N at 30 and change T. Then, I let T be fixed at 30 and change N. Table 2 summarizes the results for different time-series dimensions with $N = 30$. Table 3 summarizes the results for different cross-section dimensions with $T = 30$. We find that an increase in cross-section

Table 1
LSDV regression with $N = T$

$N(T)$	$\bar{\beta}$	$SSD(\hat{\beta})$	$ESE(\hat{\beta})$	$\bar{t}_\beta$	$SSD(t_\beta)$	$P(\lvert t_\beta\rvert > 1.96)$	$\bar{R}^2$	$\overline{DW}$
10	− 0.0028	0.2865	0.1495	− 0.0185	1.9603	0.3150	0.0384	0.6184
20	0.0024	0.2039	0.0727	0.0339	2.8375	0.4889	0.0200	0.3057
30	0.0012	0.1652	0.0479	0.0248	3.4693	0.5723	0.0133	0.2033
40	0.0006	0.1423	0.0358	0.0160	4.0065	0.6166	0.0100	0.1519
50	0.0001	0.1265	0.0286	0.0060	4.4392	0.6598	0.0079	0.1212
60	0.0004	0.1160	0.0237	0.0207	4.8987	0.6894	0.0066	0.1009
70	− 0.0006	0.1072	0.0203	− 0.0365	5.3014	0.7096	0.0057	0.0864
80	− 0.0001	0.1002	0.0178	− 0.0057	5.6369	0.7264	0.0049	0.0755

Note: (a) $\bar{\beta} = (1/M)\sum_{i=1}^{M}\hat{\beta}_i$. (b) $\bar{t}_\beta = (1/M)\sum_{i=1}^{M} t_{\beta_i}$. (c) $\overline{R^2} = (1/M)\sum_{i=1}^{M} R_i^2$. (d) $\overline{DW} = (1/M)\sum_{i=1}^{M} DW_i$. (e) The number of replications = 10,000.

Table 2
LSDV regression with $N = 30$

T	$\bar{\beta}$	$SSD(\hat{\beta})$	$ESE(\hat{\beta})$	$\bar{t}_\beta$	$SSD(t_\beta)$	$P(\lvert t_\beta\rvert > 1.96)$	$\bar{R}^2$	$\overline{DW}$
10	0.0006	0.1669	0.0861	0.0104	1.9546	0.3130	0.0136	0.5686
20	0.0022	0.1660	0.0593	0.0397	2.8192	0.4858	0.0134	0.2991
30	0.0012	0.1652	0.0479	0.0248	3.4693	0.5723	0.0133	0.2033
40	0.0004	0.1639	0.0413	0.0131	3.9855	0.6218	0.0131	0.1538
50	− 0.0001	0.1651	0.0368	− 0.0074	4.5078	0.6651	0.0133	0.1238
60	0.0013	0.1643	0.0336	0.0381	4.9167	0.6889	0.0131	0.1033
70	− 0.0012	0.1647	0.0310	− 0.0419	5.3318	0.7103	0.0132	0.0890
100	0.0026	0.1634	0.0259	0.1072	6.3427	0.7548	0.0130	0.0626
150	0.0002	0.1647	0.0211	0.0802	7.8426	0.8022	0.0132	0.0418

Note: (a) $\bar{\beta} = (1/M)\sum_{i=1}^{M}\hat{\beta}_i$. (b) $\overline{t_\beta} = (1/M)\sum_{i=1}^{M} t_{\beta_i}$. (c) $\overline{R^2} = (1/M)\sum_{i=1}^{M} R_i^2$. (d) $\overline{DW} = (1/M)\sum_{i=1}^{M} DW_i$.
(e) The number of replications = 10,000.

Table 3
LSDV Regression with $T = 30$

N	$\bar{\beta}$	$SSD(\hat{\beta})$	$ESE(\hat{\beta})$	$\bar{t}_\beta$	$SSD(t_\beta)$	$P(\lvert t_\beta\rvert > 1.96)$	$\bar{R}^2$	$\overline{DW}$
10	0.0029	0.2848	0.0832	0.0261	3.5000	0.5732	0.0378	0.2232
20	0.0037	0.2005	0.0588	0.0567	3.4577	0.5788	0.0195	0.2080
30	0.0012	0.1652	0.0479	0.0248	3.4693	0.5723	0.0133	0.2033
40	0.0009	0.1425	0.0415	0.0195	3.4532	0.5692	0.0099	0.2007
50	0.0001	0.1265	0.0371	0.0259	3.4306	0.5699	0.0109	0.1993
60	0.0017	0.1163	0.0339	0.0444	3.4478	0.5635	0.0067	0.1983
70	0.0020	0.1080	0.0313	0.0588	3.4613	0.5745	0.0058	0.1976

Note: (a) $\bar{\beta} = (1/M)\sum_{i=1}^{M}\hat{\beta}_i$. (b) $\overline{t_\beta} = (1/M)\sum_{i=1}^{M} t_{\beta_i}$. (c) $\overline{R^2} = (1/M)\sum_{i=1}^{M} R_i^2$. (d) $\overline{DW} = (1/M)\sum_{i=1}^{M} DW_i$.
(e) The number of replications = 10,000.

decreases the SSD of $\hat{\beta}$; varying cross-section size has little effect on the t-statistic; and the t-statistic is closely related to the time-series dimension. The reason why the t-statistic does not converge to the standard normal distribution is that the t-statistic does not have a zero mean even when T is fixed (at 30). Hence, increasing N will not help the t-statistic converge to a meaningful distribution (here a standard normal distribution) without some sort of normalization. When T grows, the SSD of the t-statistic and the rejection rate of the null hypothesis based on the conventional critical value increase significantly.

The results obtained from the pure time-series tell us that R^2 may be quite high in the spurious regression. However, this is not the case using panel data. Tables 1 and 2 reveal that R^2 remains very low and decreasing, which is consistent with the result in Corollary 1. Corollary 1 also demonstrates that the

$DW \to 0$ and $TDW \to 6$ as $T \to \infty$ and then $N \to \infty$. The results on the Durbin–Watson statistic are also what we expected. Clearly, as shown in Table 1, DW converges to zero in probability. Furthermore, TDW stays constant at about 6 on average.

5.2. Residual-based tests

To examine the finite sample properties of the proposed tests, I conducted Monte Carlo experiments which are similar to the design of Engle and Granger (1987) and are used in Gonzalo (1994) and Haug (1996). The data generating process (DGP) was

$$y_{it} - \alpha_i - \beta x_{it} = z_{it},$$

$$a_1 y_{it} - a_2 x_{it} = w_{it}, \qquad z_{it} = \rho z_{it-1} + e_{zit},$$

$$w_{it} = w_{it-1} + e_{wit}, \qquad e_{wit} = \phi_{it} + \pi \phi_{it-1},$$

and

$$\begin{bmatrix} e_{zit} \\ \phi_{it} \end{bmatrix} \overset{\text{iid}}{\sim} \mathrm{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \theta\sigma \\ \theta\sigma & \sigma^2 \end{bmatrix} \right).$$

The RNDN procedure in GAUSS was used to generate the random numbers. The data were generated by creating $T + 1000$ observations and discarding the first 1000 observations to remove the effect of the initial conditions. The results I report are based on 10 000 replications. I generate α_i from a uniform distribution, $U[0,10]$, and set $\beta = 2$.

For the purposes of the this paper, I consider the following values for the parameters in the DGP: $a_1 = 1$, $a_2 = -1$, $\rho = (0.95,1)$, $\sigma = (0.25,1,4)$, $\pi = (-0.8,0,0.8)$, and $\theta = (-0.5,0,0.5)$. When $\rho = 1$, y_{it} and x_{it} are not cointegrated, and when $|\rho| < 1$, they are cointegrated.

The estimate of the long-run covariance matrix

$$\hat{\Omega} = \begin{pmatrix} \hat{\sigma}^2_{0u} & \hat{\sigma}_{0u\varepsilon} \\ \hat{\sigma}_{0u\varepsilon} & \hat{\sigma}^2_{0\varepsilon} \end{pmatrix}$$

and the long-run conditional variance, $\hat{\sigma}^2_{0v} = \hat{\sigma}^2_{0u} - \hat{\sigma}^2_{0u\varepsilon}\hat{\sigma}^{-2}_{0\varepsilon}$, were obtained by using the procedure KERNEL in COINT 2.0 with a Bartlett window of lag length 5.

Define the following bias-corrected test statistics:

$$DF_{\rho} = \frac{\sqrt{N}T(\hat{\rho}-1)+3\sqrt{N}}{\sqrt{51/5}},$$

and

$$DF_t = \sqrt{\frac{5t_{\rho}}{4}} + \sqrt{\frac{15N}{8}}.$$

DF_{ρ}^{*}, DF_t^{*}, and ADF are defined in Remarks 5 and 6. We expect that DF_{ρ}^{*}, DF_t^{*}, and ADF converge to N(0,1) asymptotically. I also put DF_{ρ} and DF_t here for comparison and the two lags are selected for ADF.

Table 4 contains estimates of the size of tests at the 5% level when the asymptotic critical value, 1.645, from the N(0,1) distribution was used. Five different tests are considered – DF_{ρ}, DF_t, DF_{ρ}^{*}, DF_t^{*}, and ADF – for different sample sizes with $\theta = 0$, no moving average component ($\pi = 0.0$), $\sigma = 1$, and endogenous x_{it} ($a_1 = 1.0$). We note that all tests have a large size distortion when T is small (e.g., $T = 10$) even at $N = 300$, but the size distortion begins to disappear quickly when the T is increased to 25 for all N. The empirical sizes of DF_{ρ}^{*} and DF_t^{*} are close to 0.05 when T and N are both large. But the empirical size of the ADF test is greater than 0.09 at each sample size for T and N. Overall, it is found that DF_{ρ}^{*} and DF_t^{*} outperform the rest in terms of the size distortion.

Table 5 reports the unadjusted powers of the five tests when the asymptotic critical value is used and $\rho = 0.90$ for different sample sizes with 0 contemporaneous correlation ($\theta = 0$), no moving average component ($\pi = 0.0$), $\sigma = 1$ and endogenous x_{it} ($a_1 = 1.0$). All tests have little power when $T = 10$ and N is small, as one would expect. However, DF_{ρ}^{*} displays little power even when N is large. When T is increased to 25 and above, DF_{ρ}^{*} dominates DF_t^{*}, and the ADF tests for all N. Interestingly, DF_{ρ} and DF_t, though they are misspecified, perform pretty well in terms of size distortion and power.

To provide further insight into other specifications, I consider the size and power of these five tests for three different values of θ,π, and σ in Tables 6–8. With negative π, all five tests are severely distorted, as one would expect. In fact, with $\pi = -0.8$, all tests except ADF reject no cointegration 100% of the time for a nominal 5% level for sample sizes $T = 50$ and $N = 50$. The empirical sizes of the DF_{ρ} and DF_t tests are smaller than the nominal sizes of the tests (at 5%) when $\pi = 0.8$ at $N = T = 50$. For $\pi = 0.8$, the size distortion of all five tests begins to improve as σ is increased from 0.25 to 1 and 4. For $\pi = 0.0$, all five tests

Table 4
The empirical size at 5% for $a_1 = 1$, $\sigma = 1$, $\rho = 1$, $\pi = 0.0$, and $\theta = 0$

	$T = 10$					$T = 25$				
N	DF_ρ	DF_t	DF_ρ^*	DF_t^*	ADF	DF_ρ	DF_t	DF_ρ^*	DF_t^*	ADF
1	0.350	0.282	0.034	0.075	0.139	0.435	0.239	0.157	0.098	0.137
2	0.257	0.233	0.028	0.079	0.138	0.307	0.180	0.130	0.092	0.125
5	0.182	0.191	0.021	0.084	0.167	0.195	0.139	0.088	0.089	0.115
10	0.149	0.179	0.018	0.096	0.210	0.142	0.114	0.068	0.079	0.113
15	0.142	0.179	0.021	0.099	0.272	0.135	0.118	0.059	0.080	0.123
20	0.144	0.193	0.023	0.115	0.315	0.121	0.109	0.054	0.078	0.123
25	0.149	0.212	0.024	0.121	0.367	0.115	0.110	0.052	0.078	0.129
50	0.178	0.270	0.041	0.192	0.575	0.119	0.122	0.053	0.089	0.161
75	0.208	0.320	0.068	0.255	0.722	0.117	0.126	0.054	0.095	0.185
100	0.242	0.372	0.098	0.314	0.829	0.121	0.139	0.061	0.106	0.214
150	0.318	0.472	0.173	0.433	0.939	0.132	0.154	0.070	0.123	0.265
200	0.378	0.551	0.269	0.535	0.979	0.142	0.170	0.079	0.138	0.308
250	0.446	0.623	0.372	0.630	0.993	0.153	0.187	0.091	0.156	0.356
300	0.513	0.682	0.473	0.708	0.998	0.171	0.212	0.105	0.177	0.407
	$T = 50$					$T = 100$				
1	0.461	0.218	0.354	0.145	0.167	0.467	0.213	0.429	0.179	0.184
2	0.322	0.159	0.249	0.115	0.135	0.319	0.152	0.291	0.133	0.139
5	0.202	0.118	0.153	0.093	0.107	0.209	0.115	0.187	0.103	0.113
10	0.149	0.101	0.106	0.079	0.092	0.150	0.092	0.129	0.082	0.093
15	0.127	0.091	0.087	0.075	0.090	0.127	0.083	0.105	0.078	0.085
20	0.121	0.093	0.084	0.076	0.093	0.116	0.084	0.098	0.077	0.087
25	0.113	0.091	0.077	0.074	0.094	0.108	0.078	0.090	0.071	0.084
50	0.105	0.090	0.072	0.075	0.098	0.095	0.076	0.079	0.071	0.081
75	0.093	0.085	0.065	0.075	0.107	0.087	0.073	0.073	0.067	0.079
100	0.092	0.093	0.064	0.077	0.114	0.085	0.070	0.071	0.062	0.081
150	0.094	0.094	0.069	0.084	0.127	0.084	0.075	0.069	0.067	0.085
200	0.098	0.098	0.069	0.086	0.141	0.085	0.076	0.068	0.068	0.091
250	0.098	0.104	0.072	0.091	0.149	0.079	0.078	0.069	0.071	0.096
300	0.104	0.113	0.075	0.096	0.166	0.079	0.076	0.065	0.069	0.095

Note: (a) The number of replications = 10,000. (b) All rejection frequencies are based on the asymptotic critical value 1.645.

show little variation in size across the various values of θ and σ. Note that some of the results in Table 6 have extremely severe size distortions. Pedroni (1995) also observed this; it is probably due to more challenging experimental designs employed in this section.

For $\sigma \geqslant 1$, we note that the unadjusted power of tests is extremely high for all values of θ, and π. For $\sigma = 0.25$ and $\pi = 0$, the power of all tests is very low. When σ is increased to 1 and 4, the power increases dramatically.

Table 8 reports the size-adjusted powers with $\rho = 0.85$ such that each test has the same rejection frequency of 5% when the null hypothesis is true. For $\sigma = 0.25$ and $\pi = -0.8$, the power of all tests is low and the null

Table 5
The power at 5% for $a_1 = 1$, $\sigma = 1$, $\rho = 0.90$, $\pi = 0$, and $\theta = 0$

	$T=10$					$T=25$				
N	DF_ρ	DF_t	DF_ρ^*	DF_t^*	ADF	DF_ρ	DF_t	DF_ρ^*	DF_t^*	ADF
1	0.359	0.283	0.028	0.068	0.137	0.496	0.255	0.135	0.084	0.139
2	0.296	0.252	0.022	0.073	0.136	0.451	0.234	0.154	0.091	0.145
5	0.270	0.249	0.017	0.077	0.184	0.465	0.261	0.185	0.115	0.176
10	0.286	0.277	0.014	0.090	0.248	0.569	0.341	0.255	0.163	0.225
15	0.329	0.317	0.017	0.100	0.323	0.674	0.421	0.334	0.205	0.276
20	0.375	0.358	0.018	0.112	0.373	0.756	0.488	0.406	0.248	0.321
25	0.419	0.405	0.019	0.121	0.435	0.823	0.559	0.484	0.291	0.368
50	0.626	0.581	0.027	0.195	0.665	0.966	0.799	0.774	0.507	0.566
75	0.773	0.714	0.040	0.264	0.811	0.994	0.914	0.913	0.674	0.715
100	0.870	0.809	0.054	0.325	0.899	1.00	0.967	0.972	0.793	0.822
150	0.958	0.919	0.093	0.464	0.974	1.00	0.996	0.998	0.922	0.934
200	0.988	0.964	0.143	0.583	0.993	1.00	1.00	1.00	0.977	0.977
250	0.998	0.988	0.195	0.679	0.998	1.00	1.00	1.00	0.995	0.992
300	0.999	0.995	0.266	0.758	0.000	1.00	1.00	1.00	0.998	0.997
	$T=50$					$T=100$				
1	0.622	0.293	0.458	0.179	0.197	0.845	0.471	0.781	0.391	0.355
2	0.569	0.341	0.255	0.163	0.225	0.939	0.667	0.898	0.601	0.518
5	0.798	0.471	0.657	0.340	0.324	0.999	0.949	0.994	0.922	0.843
10	0.936	0.698	0.853	0.561	0.496	1.00	1.00	1.00	0.998	0.985
15	0.983	0.841	0.944	0.724	0.630	1.00	1.00	1.00	1.00	0.998
20	0.997	0.922	0.982	0.835	0.735	1.00	1.00	1.00	1.00	1.00
25	0.999	0.995	0.995	0.905	0.818	1.00	1.00	1.00	1.00	1.00
50	1.00	1.00	1.00	0.996	0.977	1.00	1.00	1.00	1.00	1.00
75	1.00	1.00	1.00	1.00	0.997	1.00	1.00	1.00	1.00	1.00
100	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
150	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
200	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
250	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
300	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Note: (a) The number of replications = 10,000. (b) All rejection frequencies are based on the asymptotic critical value 1.645.

hypothesis of no cointegration is often not rejected. With $\pi = 0$, ADF has the lowest power for all θ and, on the other hand, DF_ρ^* has the highest power. With $\pi = 0$, the power of all tests except ADF is high. When σ is increased to 1, the power of all tests increases to virtually 1 except when $\theta = -0.8$. In fact, the power of all tests remains very low, zeros for most cases when $\theta = -0.8$. Interestingly and to my surprise, ADF became very powerful when $\sigma = 4$ and $\theta \geqslant 0$.

The results in Tables 4–8 suggest that the DF_ρ^* and DF_t^* tests have better size and power properties than the DF_ρ, DF_t, and ADF tests when σ is small. However, when σ is large, ADF clearly dominates the rest.

Table 6
The empirical size of 5% of tests with the null hypothesis of no cointegration

	$\sigma = 0.25$			$\sigma = 1$			$\sigma = 4$		
θ	-0.5	0	0.5	-0.5	0	0.5	-0.5	0	0.5
$\pi = -0.8$									
DF_ρ	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
DF_t	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
DF^*_ρ	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
DF^*_t	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
ADF	0.814	0.038	0.062	1.00	0.985	0.998	0.993	0.672	0.466
$\pi = 0$									
DF_ρ	0.100	0.102	0.102	0.104	0.105	0.104	0.104	0.104	0.101
DF_t	0.089	0.089	0.089	0.087	0.090	0.088	0.089	0.088	0.088
DF^*_ρ	0.069	0.069	0.069	0.073	0.072	0.073	0.073	0.070	0.075
DF^*_t	0.072	0.072	0.072	0.077	0.090	0.088	0.074	0.076	0.073
ADF	0.097	0.096	0.096	0.099	0.098	0.100	0.099	0.099	0.100
$\pi = 0.8$									
DF_ρ	0.005	0.000	0.000	0.145	0.020	0.067	0.287	0.086	0.306
DF_t	0.014	0.000	0.000	0.113	0.028	0.064	0.206	0.078	0.215
DF^*_ρ	0.287	0.095	0.041	0.208	0.163	0.273	0.101	0.076	0.089
DF^*_t	0.161	0.070	0.048	0.149	0.116	0.174	0.098	0.078	0.088
ADF	0.519	0.615	0.416	0.157	0.317	0.269	0.049	0.113	0.044

Note: (a) The number of replications = 10,000. (b) $N = 50$ and $T = 50$. (c) All rejection frequencies are based on the asymptotic critical value 1.645. (d) $a_1 = 1$.

6. Extension to models with deterministic trends

It was assumed in the previous sections that y_{it} and x_{it} have zero drifts. It is known that many macroeconomic variables, such as GNP, could be described as $I(1)$ with drift, which is the sum of an $I(1)$ process with zero-mean increments and a linear trend. It is therefore important to study data generating processes which allow for some drift over time. In this section, I extend the results obtained in the previous sections to allow non-zero drifts for y_{it} and x_{it}. The treatment of nonzero drifts in this section follows Hansen (1992) and Hamilton (1994, pp. 591–601). Suppose that

$$\Delta x_{it} = \gamma_i + \varepsilon_{it}, \tag{14}$$

where $\gamma_i \neq 0$. Then $x_{it} = \gamma_i t + \sum_{s=1}^{t} \varepsilon_{is}$, where $\sum_{s=1}^{t} \varepsilon_{is}$ is in fact a trendless random walk.

Note that the stochastic part, $\sum_{s=1}^{t} \varepsilon_{is}$, is asymptotically dominated by the deterministic part, t. For example, the term $(x_{it} - \bar{x}_i)$ consists of the deterministic part, $\gamma_i(t - (T+1)/2)$, and the stochastic part, $\kappa_{it} - \bar{\kappa}_i$, i.e., $(x_{it} - \bar{x}_i) =$

Table 7
The power of 5% of tests with the null hypothesis of no cointegration and the alternative hypothesis of cointegration ($\rho = 0.90$)

	$\sigma = 0.25$			$\sigma = 1$			$\sigma = 4$		
θ	-0.5	0	0.5	-0.5	0	0.5	-0.5	0	0.5
$\pi = -0.8$									
DF_ρ	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
DF_t	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
DF_ρ^*	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
DF_t^*	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
ADF	0.198	0.021	0.012	0.882	0.781	0.877	0.986	0.991	0.999
$\pi = 0$									
DF_ρ	0.100	0.487	0.215	0.998	1.00	1.00	1.00	1.00	1.00
DF_t	0.044	0.298	0.165	0.907	1.00	1.00	1.00	1.00	1.00
DF_ρ^*	0.127	0.424	0.293	0.997	1.00	1.00	1.00	1.00	1.00
DF_t^*	0.054	0.251	0.206	0.812	0.996	1.00	0.999	1.00	1.00
ADF	0.095	0.259	0.224	0.813	0.977	0.997	0.988	0.998	0.998
$\pi = 0.8$									
DF_ρ	0.665	0.369	0.111	1.00	1.00	1.00	1.00	1.00	1.00
DF_t	0.324	0.186	0.076	1.00	1.00	1.00	1.00	1.00	1.00
DF_ρ^*	0.998	0.995	0.973	1.00	1.00	1.00	1.00	1.00	1.00
DF_t^*	0.888	0.855	0.738	1.00	1.00	1.00	1.00	1.00	1.00
ADF	0.844	0.993	0.996	0.987	1.00	1.00	0.982	0.999	0.988

Note: (a) The number of replications = 10,000. (b) $N = 50$ and $T = 50$. (c) All rejection frequencies are based on asymptotic critical value: 1.645. (d) $a_1 = 1$.

$\gamma_i(t - (T+1)/2) + (\kappa_{it} - \bar{\kappa}_i)$, where $\kappa_{it} = \sum_{s=1}^{t} \varepsilon_{is}$ and $\bar{\kappa}_i = (1/T)\sum_{t=1}^{T} \kappa_{it}$. For the deterministic part we note that

$$\gamma_i^2 \sum_{t=1}^{T} \left(t - \frac{T+1}{2} \right)^2 = \mathrm{O}(T^3).$$

For the stochastic part we know that $\sum_{t=1}^{T} (\kappa_{it} - \bar{\kappa}_i)^2 = \mathrm{O}(T^2)$. Thus within $\sum_{t=1}^{T} (x_{it} - \bar{x}_i)^2$ the deterministic part dominates the stochastic part.

x_{it} behaves asymptotically as if it were $\gamma_i t$. Thus the residuals from the LSDV estimation of Eq. (4) behave like the residuals from a regression of y_{it} on a time trend. Also, the dominance of the deterministic trend of x_{it} over the stochastic trend makes the correlation between u_{it} and ε_{it} irrelevant. It implies that the unit root test for $\hat{e}_{it}$ the residual from LSDV estimation of Eq. (4), has the same asymptotic distribution as the unit root test for testing $\rho = 1$ in the following regression:

$$y_{it} = \rho y_{it-1} + \alpha_i + \gamma_i t + \varepsilon_{it} \tag{15}$$

for an error term ε_{it}.

Table 8
The size-adjusted power of 5% of tests with the null hypothesis of no cointegration and the alternative hypothesis of cointegration ($\rho = 0.85$)

	$\sigma = 0.25$			$\sigma = 1$			$\sigma = 4$		
θ	-0.5	0	0.5	-0.5	0	0.5	-0.5	0	0.5
$\pi = -0.8$									
DF_ρ	0.001	0.027	0.012	0.000	0.000	0.000	0.000	0.001	0.010
DF_t	0.002	0.026	0.011	0.000	0.000	0.000	0.000	0.000	0.003
DF_ρ^*	0.000	0.012	0.004	0.000	0.000	0.000	0.000	0.000	0.000
DF_t^*	0.000	0.014	0.106	0.000	0.000	0.000	0.000	0.000	0.000
ADF	0.000	0.026	0.004	0.000	0.000	0.000	0.073	0.925	1.00
$\pi = 0$									
DF_ρ	0.297	0.732	0.508	1.00	1.00	1.00	1.00	1.00	1.00
DF_t	0.128	0.446	0.334	1.00	1.00	1.00	1.00	1.00	1.00
DF_ρ^*	0.388	0.751	0.728	1.00	1.00	1.00	1.00	1.00	1.00
DF_t^*	0.128	0.458	0.464	1.00	1.00	1.00	1.00	1.00	1.00
ADF	0.128	0.299	0.336	0.991	0.999	1.00	1.00	1.00	1.00
$\pi = 0.8$									
DF_ρ	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
DF_t	0.998	0.997	0.993	1.00	1.00	1.00	1.00	1.00	1.00
DF_ρ^*	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
DF_t^*	0.990	0.995	0.992	1.00	1.00	1.00	1.00	1.00	1.00
ADF	0.582	0.974	0.996	1.00	1.00	1.00	1.00	1.00	0.988

Note: (a) The number of replications = 10,000. (b) $N = 50$ and $T = 50$. (c) All rejection frequencies are based on the 5th percentile of the empirical distribution. (d) $a_1 = 1$.

Remark 12. Similarly, Eq. (15) will continue to hold when there is more than one regressor if all the regressors have nonzero drifts. Let us consider the case where there are two regressors, x_{1it} and x_{2it}, without loss of generality. Suppose that the regressors are generated by

$$\begin{bmatrix} \Delta x_{1it} \\ \Delta x_{2it} \end{bmatrix} = \begin{bmatrix} \gamma_{1i} \\ \gamma_{2i} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \end{bmatrix}. \tag{16}$$

Then Eq. (4) can be written as

$$y_{it} = \alpha_i + \beta_1 x_{1it} + \beta_2 x_{2it} + e_{it} \tag{17}$$

for an error term e_{it}. From Eqs. (16) and (17) we note the unit root test for $\hat{e}_{it}$, the residual from LSDV estimator of Eq. (17), again has the same asymptotic distribution as the unit root test for testing $\rho = 1$ in the following regression:

$$y_{it} = \rho y_{it-1} + \alpha_i + (\beta_1 \gamma_{1i} + \beta_2 \gamma_{2i})t + \varepsilon_{it}. \tag{18}$$

Obviously, Eq. (18) is identical to Eq. (15) with $\gamma_i = (\beta_1\gamma_{1i} + \beta_2\gamma_{2i})$.

Remark 13. If there is a drift in y_{it}, i.e., $\Delta y_{it} = \delta_i + u_{it}$. Then we simply replace y_{it} by $y_{it} - (\delta_i/\gamma_i)x_{it}$ in Eq. (15).

Remark 14. Testing $\rho = 1$ in Eq. (15) has been studied by Levin and Lin (1993) with i.i.d. error terms.

The asymptotic distributions of *DF* tests and an *ADF* test for Eq. (15) can be obtained analogously with Theorems 2 and 3.

Theorem 4. Suppose $\mathrm{E}(\Delta x_{it}) \neq 0$ *and Assumptions* 1–3 *are satisfied. Then*

$$\sqrt{N}\left[T(\hat{\rho}-1) - \frac{\mu_{7T}}{\mu_{8T}}\right] \Rightarrow \mathrm{N}\left(0, \frac{15}{4} + \frac{\sigma_u^4}{\sigma_{0u}^4}\frac{2475}{112}\right),$$

$$t_\rho - \frac{\sqrt{N}\mu_{7T}}{s_e\sqrt{\mu_{8T}}} \Rightarrow \mathrm{N}\left(0, \frac{1}{4}\frac{\sigma_{0u}^2}{\sigma_u^2} + \frac{165}{448}\frac{\sigma_u^2}{\sigma_{0u}^2}\right),$$

and

$$t_{ADF} - \frac{\sqrt{N}\mu_{7T}}{s_v\sqrt{\mu_8}} \Rightarrow \mathrm{N}\left(0, \frac{1}{4}\frac{\sigma_{0u}^2}{\sigma_u^2} + \frac{165}{448}\frac{\sigma_u^2}{\sigma_{0u}^2}\right),$$

where s_v *is defined in a similar fashion as in Section 3.2 and* s_e, μ_{7T} *and* μ_{8T} *are defined in Eqs.* (*E.10*), (*E.8*) *and* (*E.9*), *respectively.*

Remark 15. From Theorem 4 we find that inclusion of drifts in the regressors does not affect the asymptotic distributions of the tests of cointegration in panel data although the asymptotic variances and asymptotic biases are different from the ones in Theorems 2 and 3. That is, allowing drifts for the regressors only affects the precision of the testing and the main results given in Section 3 continue to hold.

7. Conclusion

The first half of this paper develops a framework for understanding the behavior of spurious panel regression. I have provided an asymptotic theory for the behavior of the LSDV estimator in a model which attempts to estimate the panel regression when the dependent variable and independent variable are actually independent $I(1)$ processes. In particular, the t-statistic diverges in spite of the fact that the LSDV estimator converges to its true value in probability. The asymptotic results and simulation results in this paper are useful beyond explaining the spurious effects on panel regression. Asymptotic distributions of residual-based tests depend on the LSDV estimator from the spurious regression because residual-based tests for cointegration in panel data take the spurious regression as the null hypothesis.

In the second half of this paper, I propose tests for the null hypothesis of no cointegration in panel data and derive asymptotic distributions for each test. The simulations show that the distributions of DF^*_ρ, DF^*_t, and ADF can be far different from N(0, 1) suggested by the theory when the underlying process contains an MA component. From the Monte Carlo results, it is apparent that the DF_ρ and DF_t tests are substantially robust despite the model misspecification. The results in Tables 4–8 suggest that the DF^*_ρ and DF^*_t tests have better size and power properties than the DF_ρ, DF_t, and ADF tests.

Acknowledgements

I thank Suzanne McCoskey, Biing-Shen Kuo, two referees, an associate editor and a coeditor, Richard Blundell, for helpful comments. I also thank participants of workshop seminars at Syracuse University, Hong Kong University of Science and Technology, Academia Sinica, and SUNY-Albany for helpful comments and suggestions. I thank Bangtian Chen for his research assistance on an earlier draft of this paper. Thanks also go to Jodi Woodson and Martha Bonney for correcting my English and carefully checking the manuscript to enhance its readability.

Appendix A. Lemma 1

Before we derive the asymptotics of Theorem 1 we need to summarize some limit theory for the LSDV estimator. Note that

$$\hat{\beta} = \frac{\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T^2}\sum_{t=1}^{T}(x_{it}-\bar{x}_i)y_{it}}{\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T^2}\sum_{t=1}^{T}(x_{it}-\bar{x}_i)^2} = \frac{\frac{1}{N}\sum_{i=1}^{N}\zeta_{1iT}}{\frac{1}{N}\sum_{i=1}^{N}\zeta_{2iT}} = \frac{\xi_{1NT}}{\xi_{2NT}}, \tag{A.1}$$

where $\bar{x}_i = (1/T)\sum_{t=1}^{T}x_{it}$, $\bar{y}_i = (1/T)\sum_{t=1}^{T}y_{it}$, $\zeta_{1iT} = (1/T^2)\sum_{t=1}^{T}(x_{it}-\bar{x}_i)y_{it}$, $\zeta_{2iT} = (1/T^2)\sum_{t=1}^{T}(x_{it}-x_i)^2$, $\xi_{1NT} = (1/N)\sum_{i=1}^{N}\zeta_{1iT}$, and $\xi_{2NT} = (1/N)\sum_{i=1}^{N}\zeta_{2iT}$.

Remark A.1. We need to require that ζ_{1iT} and ζ_{1i} (ζ_{2iT} and ζ_{2i}) are defined on the same probability space so that $\sum_{i=1}^{N}\zeta_{1i}(\sum_{i=1}^{N}\zeta_{2i})$ can be used to derive the sequential asymptotics. The requirement that ζ_{1iT} and ζ_{1i} are defined on the same probability space can be done by suitably enlarging the underlying space by using the Skorohod's theorem in Billingsley (1986, p. 399) and Phillips and Moon (1997, p. 8).

All limits in Lemma 1 are taken as $T \to \infty$, for all i, except (j) and (k). We use $V_i(x)$ to denote $\sum_{t=1}^{T}(x_{it} - \bar{x}_i)^2$, use $V(x)$ to denote $\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it} - \bar{x}_i)^2$, use $V_i(x,y)$ to denote $\sum_{t=1}^{T} y_{it}(x_{it} - \bar{x}_i)$, and use $V(x,y)$ to denote $\sum_{i=1}^{N}\sum_{t=1}^{T} y_{it}(x_{it} - \bar{x}_i)$ to simplify the excessive notation.

Lemma A.1. If Assumptions 1–3 hold, then

(a) $\zeta_{1iT} \Rightarrow \int W_i^* V_i^* - \int W_i^* \int V_i^* = \zeta_{1i}$,
(b) $\zeta_{2iT} \Rightarrow \int W_i^{*2} - (\int W_i^*)^2 = \zeta_{2i}$,
(c) $\mathrm{E}[\zeta_{1iT}] = \mu_{1T} = \mathrm{E}[\zeta_{1i}] = \mu_1 = 0$,
(d) $\mathrm{E}[\zeta_{2i}] = \mu_2 = \frac{1}{6}$,
(e) $\mathrm{E}[\zeta_{2iT}] = \mu_{2T} = \mu_2 + \mathrm{O}(T^{-1})$,
(f) $Var[\zeta_{1iT}] = \sigma_{1T}^2 = \frac{1}{90} + \mathrm{O}(T^{-1})$,
(g) $Var[\zeta_{2iT}] = \sigma_{2T}^2 = \frac{1}{45} + \mathrm{O}(T^{-1})$,
(h) $Var[\zeta_{1i}] = \sigma_1^2 = \frac{1}{90}$,
(i) $Var[\zeta_{2i}] = \sigma_2^2 = \frac{1}{45}$,
(j) $\xi_{1NT} \xrightarrow{\mathrm{p}} \mu_1 = 0$,
(k) $\xi_{2NT} \xrightarrow{\mathrm{p}} \mu_2 = \frac{1}{6}$.

Proof. Let

$$y_{it}^* = y_{it} - \sigma_{0u\varepsilon}\sigma_{0\varepsilon}^{-2} x_{it}, \tag{A.2}$$

$$x_{it}^* = \sigma_{0\varepsilon}^{-1} x_{it},$$

with

$$\hat{e}_{it}^* = y_{it}^* - \hat{\alpha}_i^* - \hat{\beta}^* x_{it}^*. \tag{A.3}$$

The residuals from LSDV estimation of Eq. (4) are identical to those from LSDV of Eq. (A.2) in that

$$\begin{aligned} y_{it} - \hat{\alpha}_i - \hat{\beta} x_{it} &= y_{it}^* - \hat{\alpha}_i^* - \hat{\beta}^* x_{it}^* \\ &= y_{it} - \sigma_{0u\varepsilon}\sigma_{0\varepsilon}^{-2} x_{it} - \hat{\alpha}_i^* - \hat{\beta}^*(\sigma_{0\varepsilon}^{-1/2} x_{it}) \\ &= y_{it} - \hat{\alpha}_i^* - (\hat{\beta}^* \sigma_{0\varepsilon}^{-1} + \sigma_{0u\varepsilon}\sigma_{0\varepsilon}^{-2}) x_{it} \end{aligned} \tag{A.4}$$

with $\hat{\alpha}_i = \hat{\alpha}_i^*$ and $\hat{\beta} = \hat{\beta}^* \sigma_{0\varepsilon}^{-1} + (\sigma_{0\varepsilon}^{-2})\sigma_{0u\varepsilon}$. This implies that $\hat{\beta}^* = \sigma_{0\varepsilon}\hat{\beta} - \sigma_{0\varepsilon}^{-1}\sigma_{0u\varepsilon}$. Define

$$\sigma_{0v}^2 = \sigma_{0u}^2 - \sigma_{0u\varepsilon}^2 \sigma_{0\varepsilon}^{-2} \tag{A.5}$$

and $\sigma_v^2 = \sigma_u^2 - \sigma_{u\varepsilon}^2 \sigma_\varepsilon^{-2}$. Notice that

$$\begin{bmatrix} y_{it}^*/\sigma_{0v} \\ x_{it}^* \end{bmatrix} = L' \begin{bmatrix} y_{it} \\ x_{it} \end{bmatrix},$$

where

$$L' = \begin{bmatrix} \dfrac{1}{\sigma_{0v}} & -\dfrac{1}{\sigma_{0v}} \sigma_{0u\varepsilon}\sigma_{0\varepsilon}^{-2} \\ 0 & \dfrac{1}{\sigma_{0\varepsilon}} \end{bmatrix}.$$

It follows that (e.g., Hamilton, 1994) $W_i^*(r) = L'\Omega^{1/2} W_i(r)$ is a Wiener process with an identity covariance, I, with

$$W_i^*(r) = \begin{bmatrix} V_i^*(r) \\ W_i^*(r) \end{bmatrix}.$$

Therefore

$$\begin{aligned} \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} L' w_{it} &= \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} \begin{bmatrix} \dfrac{1}{\sigma_{0v}} & -\dfrac{1}{\sigma_{0v}} \sigma_{0u\varepsilon}\sigma_{0\varepsilon}^{-2} \\ 0 & \sigma_{0\varepsilon}^{-1} \end{bmatrix} \begin{bmatrix} u_{it} \\ \varepsilon_{it} \end{bmatrix} \\ &= \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} \begin{bmatrix} \dfrac{u_{it}}{\sigma_{0v}} & -\dfrac{\varepsilon_{it}}{\sigma_{0v}} \sigma_{0u\varepsilon}\sigma_{0\varepsilon}^{-2} \\ \dfrac{\varepsilon_{it}}{\sigma_{0\varepsilon}} & \end{bmatrix} \\ &= \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} \begin{bmatrix} \dfrac{u_{it}^*}{\sigma_{0v}} \\ \varepsilon_{it}^* \end{bmatrix} \\ &\Rightarrow W_i^*(r) \quad \text{as } T \to \infty, \end{aligned}$$

where $u_{it}^* = u_{it} - \varepsilon_{it}\sigma_{0u\varepsilon}\sigma_{0\varepsilon}^{-2}$ and $\varepsilon_{it}^* = \varepsilon_{it}/\sigma_{0\varepsilon}$. Define

$$\sqrt{N}\frac{\hat{\beta}^*}{\sigma_{0v}} = \frac{\sqrt{N}\xi_{1NT}}{\xi_{2NT}},$$

where $\zeta_{1iT}=(1/T^2)\sum_{t=1}^{T}(x_{it}^*-\bar{x}_i^*)(y_{it}^*/\sigma_{0v})=(1/T^2\sigma_{0v})V_i(x^*,y^*)$, $\zeta_{2iT}=(1/T^2)\sum_{t=1}^{T}(x_{it}^*-\bar{x}_i^*)^2=(1/T^2)V_i(x^*)$, $\xi_{1NT}=(1/N)\sum_{i=1}^{N}\zeta_{1iT}$, and $\xi_{2NT}=(1/N)\sum_{i=1}^{N}\zeta_{4iT}$.

First we note that

$$\begin{bmatrix} T^{-3/2}\sum_{t=1}^{T}\frac{y_{it}^*}{\sigma_{0v}} \\ T^{-3/2}\sum_{t=1}^{T}x_{it}^* \end{bmatrix}=T^{-3/2}\sum_{t=1}^{T}L'\begin{bmatrix} y_{it} \\ x_{it} \end{bmatrix}\Rightarrow\int W_i^*, \tag{A.6}$$

and

$$\begin{bmatrix} \frac{1}{T^2}\sum_{t=1}^{T}\frac{(y_{it}^*)^2}{\sigma_{0v}^2} & \frac{1}{T^2}\sum_{t=1}^{T}\frac{y_{it}^*x_{it}^*}{\sigma_{0v}} \\ \frac{1}{T^2}\sum_{t=1}^{T}\frac{x_{it}^*y_{it}^*}{\sigma_{0v}} & \frac{1}{T^2}\sum_{t=1}^{T}(x_{it}^*)^2 \end{bmatrix}$$

$$=L'\left(\frac{1}{T^2}\sum_{t=1}^{T}\begin{bmatrix} y_{it} \\ x_{it} \end{bmatrix}[y_{it}\ x_{it}]\right)L\Rightarrow\int W_i^*W_i^{*\prime}$$

$$=\begin{bmatrix} \int V_i^{*2} & \int V_i^*W_i^* \\ \int W_i^*V_i^* & \int W_i^{*2} \end{bmatrix}. \tag{A.7}$$

It follows that

$$\xi_{2iT}\Rightarrow\int W_i^{*2}-\left\{\int W_i^*\right\}^2=\int \bar{W}_i^2,$$

for all i, where $\bar{W}_i$ is the demeaned standard Wiener process $\bar{W}_i=W_i^*-\int W_i^*$, proving (b). Recall that

$$\zeta_{1iT}=\frac{1}{T^2\sigma_{0v}}V_i(x^*,y^*)=\frac{1}{T^2}\sum_{t=1}^{T}x_{it}^*\frac{y_{it}^*}{\sigma_{0v}}-(T^{-1/2}\bar{x}_i^*)\left(T^{-1/2}\frac{\bar{y}_i^*}{\sigma_{0v}}\right).$$

Using Eqs. (A.6) and (A.7), we have

$$\frac{1}{T^2}\sum_{t=1}^{T}x_{it}^*\frac{y_{it}^*}{\sigma_{0v}}\Rightarrow\int W_i^*V_i^*,$$

$$T^{-1/2}\bar{x}_i^*\Rightarrow\int W_i^*,$$

and $T^{-1/2}(\bar{y}_i/\sigma_{0v}) \Rightarrow \int V_i^*$. Hence,

$$\zeta_{1iT} \Rightarrow \int W_i^* V_i^* - \int W_i^* \int V_i^* = \int \bar{W}_i \bar{V}_i,$$

for all i, where $\bar{V}_i = V_i^* - \int V_i^*$. This proves (a).

To prove (c), we use the fact that $\mathrm{E}[\int W_i^* V_i^*] = 0$ and $\mathrm{E}[\int W_i^*] = \mathrm{E}[\int V_i^*] = 0$, and, therefore, $\mathrm{E}[\zeta_{1i}] = \mu_1 = 0$. Also, $\mathrm{E}[\zeta_{1iT}] = \mu_{1T} = 0$ because of the independence of x_{it}^* and y_{it}^*.

Parts (d) and (g) can be shown easily as follows: $\mathrm{E}[\zeta_{2i}] = \frac{1}{6} = \mu_2$ and $Var[\zeta_{2i}] = \frac{1}{45} = \sigma_2^2$.

To prove (e), we note that

$$\begin{aligned}
\mu_{2T} &= \mathrm{E}[\zeta_{2iT}] \\
&= \frac{1}{T^2}\left\{\sum_{t=1}^{T} \mathrm{E}[x_{it}^*]^2 - \frac{1}{T}\mathrm{E}\left(\sum_{t=1}^{T} x_{it}^*\right)^2\right\} \\
&= \frac{1}{T^2}\left\{\frac{T(T+1)}{2} - \frac{T(T+1)(2T+1)}{6T}\right\} \\
&= \tfrac{1}{6} + \mathrm{O}(T^{-1}) = \mu_2 + \mathrm{O}(T^{-1}).
\end{aligned}$$

To prove (f), we first rewrite

$$\begin{aligned}
\mathrm{Var}(\zeta_{1iT}) &= \frac{1}{\sigma_{0v}^2} Var\left(\frac{1}{T^2} V_i(x^*, y^*)\right) \\
&= \frac{1}{\sigma_{0v}^2}\frac{1}{T^4}\mathrm{E}\left[\left(\sum_{t=1}^{T} x_{it}^* y_{it}^*\right)^2\right] \\
&\quad - \frac{1}{\sigma_{0v}^2}\frac{2}{T^5}\mathrm{E}\left[\left(\sum_{t=1}^{T} x_{it}^* y_{it}^*\right)\left(\sum_{t=1}^{T} x_{it}^*\right)\left(\sum_{t=1}^{T} y_{it}^*\right)\right] \\
&\quad + \frac{1}{\sigma_{0v}^2}\frac{1}{T^6}\mathrm{E}\left[\left(\sum_{t=1}^{T} x_{it}^*\right)^2\left(\sum_{t=1}^{T} y_{it}^*\right)^2\right].
\end{aligned}$$

After some tedious algebra, we can show that

$$\frac{1}{\sigma_{0v}^2}\frac{1}{T^4}\mathrm{E}\left[\left(\sum_{t=1}^{T} x_{it}^* y_{it}^*\right)^2\right] = \frac{1}{6} + \mathrm{O}(T^{-1}),$$

$$\frac{1}{\sigma_{0v}^2}\frac{2}{T^5}\mathrm{E}\left[\left(\sum_{t=1}^{T} x_{it}^* y_{it}^*\right)\left(\sum_{t=1}^{T} x_{it}^*\right)\left(\sum_{t=1}^{T} y_{it}^*\right)\right] = \frac{4}{15} + \mathrm{O}(T^{-1}),$$

and

$$\frac{1}{\sigma_{0v}^2}\frac{1}{T^6}\mathrm{E}\left[\left(\sum_{t=1}^{T} x_{it}^*\right)^2\left(\sum_{t=1}^{T} y_{it}^*\right)^2\right]=\frac{1}{9}+\mathrm{O}(T^{-1}).$$

It follows that

$$Var(\zeta_{1iT})=\frac{1}{\sigma_{0v}^2}\left(\frac{1}{6}-\frac{4}{15}+\frac{1}{9}\right)+\mathrm{O}(T^{-1})=\frac{1}{90}+\mathrm{O}(T^{-1}),$$

proving (f). Next, we prove (h). First,

$$\begin{aligned}\mathrm{Var}(\zeta_{1i})&=\frac{1}{\sigma_{0v}^2}\mathrm{E}\left[\left(\int W_i^*V_i^*-\int W_i^*\int V_i^*\right)^2\right]\\&=\frac{1}{\sigma_{0v}^2}\left\{\mathrm{E}\left[\left(\int W_i^*V_i^*\right)^2\right]+\frac{1}{\sigma_{0v}^2}\mathrm{E}\left[\left(\int W_i^*\int V_i^*\right)^2\right]\right.\\&\quad\left.-\frac{2}{\sigma_{0v}^2}\mathrm{E}\left[\left(\int W_i^*V_i^*\right)\left(\int W_i^*\int V_i^*\right)\right]\right\}.\end{aligned}$$

Again we can show, after tedious algebra,

$$\begin{aligned}&\frac{1}{\sigma_{0v}^2}\mathrm{E}\left[\left(\int W_i^*V_i^*\right)^2\right]=\frac{1}{6},\\&\frac{1}{\sigma_{0v}^2}E\left[\left(\int W_i^*\int V_i^*\right)^2\right]=\frac{1}{9},\end{aligned}$$

and

$$\frac{1}{\sigma_{0v}^2}\mathrm{E}\left[\left(\int W_i^*V_i^*\right)\left(\int W_i^*\int V_i^*\right)\right]=\frac{2}{15}.$$

Hence, $Var(\zeta_{1i})=\frac{1}{90}$, establishing (h).

Part (i) is easy. Finally, by a law of large numbers $(1/N)\sum_{i=1}^{N}\zeta_{1i}\xrightarrow{\mathrm{p}}\mu_1$ and $(1/N)\sum_{i=1}^{N}\zeta_{2i}\xrightarrow{\mathrm{p}}\mu_2$ as $N\to\infty$. It follows that

$$\xi_{1NT}=\frac{1}{N}\sum_{i=1}^{N}\zeta_{1iT}\xrightarrow{\mathrm{p}}\mu_1=0$$

and

$$\xi_{2NT} = \frac{1}{N} \sum_{i=1}^{N} \zeta_{2iT} \xrightarrow{p} \mu_2 = \frac{1}{6}$$

as required for (j) and (k) by the sequential limit theory. □

Appendix B. Proof of Theorem 1

Proof. Using (j) and (k) in Lemma A.1 we know that $\hat{\beta}^* \xrightarrow{p} 0$. Recall that $\hat{\beta} = \hat{\beta}^* \sigma_{0\varepsilon}^{-1} + \sigma_{0\varepsilon}^{-2} \sigma_{0u\varepsilon}$. Then we obtain $\hat{\beta} \xrightarrow{p} \sigma_{0u\varepsilon}/\sigma_{0\varepsilon}^2$, proving (a).

We note that, unlike the pure time-series case, $\hat{\beta}$ does converge in probability to its true value here. Next we show that $\sqrt{N}\hat{\beta}^* \Rightarrow \mathrm{N}(0, 2\sigma_{0v}^2/5)$. First, we note that $\mathrm{E}[\zeta_{1i}] = 0$ and $Var[\zeta_{1i}] = \frac{1}{90}$ from Lemma 1. Then from the Lindeberg–Levy central limit theorem and combing this the limit of $(1/N)\sum_{i=1}^{N} \zeta_{2i}$ we have

$$\frac{\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \zeta_{1i}}{\frac{1}{N} \sum_{i=1}^{N} \zeta_{2i}} \Rightarrow \mathrm{N}\left(0, \frac{2}{5}\right).$$

Then we obtain

$$\sqrt{N}\hat{\beta}^* - \sigma_{0v} \frac{\sqrt{N}\xi_{1NT}}{\xi_{2NT}} \Rightarrow \mathrm{N}\left(0, \frac{2}{5}\sigma_{0v}^2\right)$$

by the sequential limit theory. Hence

$$\sqrt{N}\left(\hat{\beta} - \frac{\sigma_{0u\varepsilon}}{\sigma_{0\varepsilon}^2}\right) \Rightarrow \mathrm{N}\left(0 \frac{2\sigma_{0v}^2}{5\sigma_{0\varepsilon}^2}\right) \tag{B.1}$$

establishes (b). Thus, after rescaling by $\sqrt{N}$, $\hat{\beta}$ converges in distribution to a normal random variable for all i as $T \to \infty$. Note that

$$\begin{aligned} \frac{1}{\sigma_{0v}^2} \frac{1}{T} s^2 &= \frac{1}{N} \sum_{i=1}^{N} \frac{1}{T^2} \sum_{t=1}^{T} \frac{1}{\sigma_{0v}^2} \hat{e}_{it}^2 \\ &= \frac{1}{N} \sum_{i=1}^{N} \frac{1}{T^2} \sum_{t=1}^{T} \frac{1}{\sigma_{0v}^2} \hat{e}_{it}^{*2} \end{aligned}$$

$$= \frac{1}{N} \sum_{i=1}^{N} \left[\frac{1}{T^2} \sum_{t=1}^{T} \left(\frac{y_{it}^* - \bar{y}_i^*}{\sigma_{0v}} \right)^2 - (\hat{\beta}^*)^2 \frac{1}{\sigma_{0v}^2} \frac{1}{T^2} (x_{it}^* - \bar{x}_i^*)^2 \right]$$

$$\xrightarrow{p} \frac{1}{6}, \tag{B.2}$$

where $\hat{e}_{it}^* = y_{it}^* - \hat{\alpha}_i^* - \hat{\beta}^* x_{it}^*$, $(1/N)\sum_{i=1}^{N} \sum_{t=1}^{T} 1/T^2 (x_{it}^* - \bar{x}_i^*)^2 = \mathrm{O_p}(1)$, and $\hat{\beta}^* = \mathrm{o_p}(1)$.

Observe that

$$T^{-1/2} t_\beta = \frac{T^{-1/2} \hat{\beta} V(x)}{s}$$

$$= \frac{\sqrt{N} \hat{\beta} \dfrac{1}{\sigma_{0v}} \sqrt{\dfrac{1}{N} \sum_{i=1}^{N} \dfrac{1}{T^2} V_i(x)}}{\sqrt{\dfrac{1}{\sigma_{0v}^2} \dfrac{1}{T} s^2}}$$

$$= \frac{\sqrt{N} (\hat{\beta}^* \sigma_{0\varepsilon}^{-1} + \sigma_{0\varepsilon}^{-2} \sigma_{0u\varepsilon}) \dfrac{\sigma_{0\varepsilon}}{\sigma_{0v}} \sqrt{\dfrac{1}{N} \sum_{i=1}^{N} \dfrac{1}{T^2} V_i(x^*)}}{\sqrt{\dfrac{1}{\sigma_{0v}^2} \dfrac{1}{T} s^2}}$$

$$= \frac{\dfrac{\sqrt{N}}{\sigma_{0v}} \left(\hat{\beta}^* + \dfrac{\sigma_{0u\varepsilon}}{\sigma_{0\varepsilon}} \right) \sqrt{\dfrac{1}{N} \sum_{i=1}^{N} \dfrac{1}{T^2} V_i(x^*)}}{\sqrt{\dfrac{1}{\sigma_{0v}^2} \dfrac{1}{T} s^2}}.$$

It follows that

$$T^{-1/2} t_\beta - \frac{T^{-1/2} \sigma_{0\varepsilon}^{-2} \sigma_{0u\varepsilon} V(x)}{s}$$

$$= \frac{T^{-1/2} \hat{\beta} \sqrt{V(x)} - T^{-1/2} \sigma_{0\varepsilon}^{-2} \sigma_{0u\varepsilon} \sqrt{V(x)}}{s}$$

$$= \frac{T^{-1/2} (\hat{\beta} - \sigma_{0\varepsilon}^{-2} \sigma_{0u\varepsilon}) \sqrt{V(x)}}{s}$$

$$= \frac{T^{-1/2} (\hat{\beta}^* \sigma_{0\varepsilon}^{-1}) \sqrt{V(x)}}{s}$$

$$= \frac{\dfrac{\sqrt{N}}{\sigma_{0v}}(\hat{\beta}^*)\sqrt{\dfrac{1}{N}\sum_{i=1}^{N}\dfrac{1}{T^2}V(x^*)}}{\sqrt{\dfrac{1}{\sigma_{0v}^2}\dfrac{1}{T}s^2}}$$

$$\Rightarrow \frac{\mathrm{N}(0,2/5)\cdot\sqrt{1/6}}{\sqrt{1/6}} = \mathrm{N}(0,\tfrac{2}{5}),$$

proving (c). It is easy to show that

$$\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T^2}V_i(x) = \sigma_{0\varepsilon}^2\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T^2}V_i(x^*) = \sigma_{0\varepsilon}^2\frac{1}{N}\sum_{i=1}^{N}\zeta_{4iT}\xrightarrow{\mathrm{p}}\frac{\sigma_{0\varepsilon}^2}{6}, \tag{B.3}$$

and

$$\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T^2}V_i(y)\\
&= \frac{1}{N}\sum_{i=1}^{N}\frac{1}{T^2}\sum_{t=1}^{T}\left(y_{it}^* + \frac{\sigma_{0u\varepsilon}}{\sigma_{0\varepsilon}}x_{it}^* - \bar{y}_{it}^* - \frac{\sigma_{0u\varepsilon}}{\sigma_{0\varepsilon}}\bar{x}_{it}^*\right)^2\\
&= \frac{1}{N}\sum_{i=1}^{N}\frac{1}{T^2}\sum_{t=1}^{T}\left[(y_{it}^* - \bar{y}_{it}^*)^2 + \left(\frac{\sigma_{0u\varepsilon}}{\sigma_{0\varepsilon}}\right)^2(x_{it}^* - \bar{x}_{it}^*)^2\right.\\
&\qquad\left. - 2\frac{\sigma_{0u\varepsilon}}{\sigma_{0\varepsilon}}(y_{it}^* - \bar{y}_{it}^*)(x_{it}^* - \bar{x}_{it}^*)\right]\\
&\qquad\xrightarrow{\mathrm{p}}\frac{\sigma_{0v}^2}{6} + \left(\frac{\sigma_{0u\varepsilon}}{\sigma_{0\varepsilon}}\right)^2\frac{1}{6}.
\end{aligned} \tag{B.4}$$

To prove (d), we use Eqs. (B.3) and (B.4) to obtain

$$\begin{aligned}
R^2 &= \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}(\hat{y}_{it} - \bar{y}_i)^2}{\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - \bar{y}_i)^2}\\
&= \frac{\hat{\beta}^2\dfrac{1}{N}\sum_{i=1}^{N}\dfrac{1}{T^2}V_i(x)}{\dfrac{1}{N}\sum_{i=1}^{N}\dfrac{1}{T^2}V_i(y)}
\end{aligned}$$

$$\xrightarrow{\text{p}} \frac{\left(\dfrac{\sigma_{0u\varepsilon}}{\sigma_{0\varepsilon}^2}\right)^2 \dfrac{\sigma_{0\varepsilon}^2}{6}}{\dfrac{\sigma_{0v}^2}{6} + \left(\dfrac{\sigma_{0u\varepsilon}}{\sigma_{0\varepsilon}}\right)^2 \dfrac{1}{6}} = \frac{\sigma_{0u\varepsilon}^2}{\sigma_{0v}^2\sigma_{0\varepsilon}^2 + \sigma_{0u\varepsilon}^2}.$$

Again using Eq. (A.4) gives

$$DW = \frac{\sum_{i=1}^N \sum_{t=2}^T (\hat{e}_{it} - \hat{e}_{it-1})^2}{\sum_{i=1}^N \sum_{t=1}^T \hat{e}_{it}^2} = \frac{\sum_{i=1}^N \sum_{t=2}^T (\hat{e}_{it}^* - \hat{e}_{it-1}^*)^2}{\sum_{i=1}^N \sum_{t=1}^T \hat{e}_{it}^{*\,2}}$$

$$= \frac{1}{T} \frac{\dfrac{1}{N} \sum_{i=1}^N \dfrac{1}{T} \sum_{t=2}^T (u_{it}^* - \hat{\beta}\varepsilon_{it}^*)^2}{\dfrac{1}{N} \sum_{i=1}^N \dfrac{1}{T^2} \sum_{t=1}^T \hat{e}_{it}^{*2}}.$$

Now,

$$\frac{1}{N}\sum_{i=1}^N \frac{1}{T}\sum_{t=2}^T (u_{it}^* - \hat{\beta}^* \varepsilon_{it}^*)^2$$

$$= \frac{1}{N}\sum_{i=1}^N \left\{ \frac{1}{T}\sum_{t=2}^T u_{it}^{*2} - \hat{\beta}^* \frac{1}{T}\sum_{t=2}^T u_{it}^*\varepsilon_{it}^* + \frac{1}{T}(\hat{\beta}^*)^2 \sum_{t=2}^T \varepsilon_{it}^{*2} \right\}$$

$$= \frac{1}{N}\sum_{i=1}^N \left\{ \frac{1}{T}\sum_{t=2}^T u_{it}^{*2} \right\} - \hat{\beta}^* \frac{1}{N}\sum_{i=1}^N \left\{ \frac{1}{T}\sum_{t=2}^T u_{it}^*\varepsilon_{it}^* \right\} + \hat{\beta}^{*2} \frac{1}{N}\sum_{i=1}^N \frac{1}{T}\sum_{t=2}^T \varepsilon_{it}^{*2}$$

$$\xrightarrow{\text{p}} \sigma_{0v}^2$$

because

$$\frac{1}{N}\sum_{i=1}^N \left\{ \frac{1}{T}\sum_{t=2}^T u_{it}^{*2} \right\} \xrightarrow{\text{p}} \sigma_{0v}^2,$$

$\hat{\beta}^* = \mathrm{o_p}(1)$, $(1/N)\sum_{i=1}^N \{(1/T)\sum_{t=2}^T u_{it}^*\varepsilon_{it}^*\} = \mathrm{o_p}(1)$, and $(1/N)\sum_{i=1}^N (1/T)\sum_{t=2}^T \varepsilon_{it}^{*2} = \mathrm{O_p}(1)$.

From Eq. (B.2),

$$\frac{1}{N}\sum_{i=1}^N \frac{1}{T^2}\sum_{t=1}^T \hat{e}_{it}^{*2} \xrightarrow{\text{p}} \frac{\sigma_{0v}^2}{6}.$$

Thus, $DW = \mathrm{O_p}(T^{-1})$ or $DW \xrightarrow{\text{p}} 0$. It also follows that $TDW \xrightarrow{\text{p}} 6$. Therefore, we establish parts (e) and (f). □

Appendix C. Proof of Theorem 2

Proof. First, we note that

$$\hat{e}^*_{it-1} = y^*_{it-1} - \hat{\alpha}^*_i - \hat{\beta}^* x^*_{it-1} = y^*_{it-1} - \bar{y}^*_i - \hat{\beta}^*(x^*_{it-1} - \bar{x}^*_i)$$

and $\Delta\hat{e}^*_{it} = u^*_{it} - \hat{\beta}^*\varepsilon^*_{it}$, where $u^*_{it} = u_{it} - (\sigma_{0u\varepsilon}/\sigma^2_{0\varepsilon})\varepsilon_{it}$ and $\varepsilon^*_{it} = \varepsilon_{it}/\sigma_{0\varepsilon}$. It follows that

$$\begin{aligned}
\zeta_{4NT} &= \frac{1}{N}\sum_{i=1}^{N}\left\{\frac{1}{T^2}\sum_{t=2}^{T} \hat{e}^{*2}_{it-1}\right\} \\
&= \frac{1}{N}\sum_{i=1}^{N}\left\{\frac{1}{T^2}\sum_{t=2}^{T}(y^*_{it-1} - \bar{y}^*_i)^2\right\} - \hat{\beta}^{*2}\frac{1}{N}\sum_{i=1}^{N}\left\{\frac{1}{T^2}\sum_{t=2}^{T}(x^*_{it-1} - \bar{x}^*_i)^2\right\} \\
&= \frac{1}{N}\sum_{i=1}^{N}\left\{\frac{1}{T^2}\sum_{t=2}^{T}(y^*_{it-1} - \bar{y}^*_i)^2\right\} + o_p(1), \qquad \text{(C.1)}
\end{aligned}$$

since $(1/N)\sum_{i=1}^{N}\{(1/T^2)\sum_{t=2}^{T}(x^*_{it-1} - \bar{x}^*_i)^2\} = O_p(1)$ and $\hat{\beta}^{*2} = o_p(1)$. We also note that

$$\frac{1}{T^2}\sum_{t=2}^{T}(y^*_{it-1} - \bar{y}^*_i)^2 \Rightarrow \sigma^2_{0v}\left[\int V^{*2}_i - \left\{\int V^*_i\right\}^2\right] = \zeta_{4i}, \qquad \text{(C.2)}$$

$$E[\zeta_{4i}] = \frac{\sigma^2_{0v}}{6} = \mu_4,$$

and $Var[\zeta_{4i}] = \sigma^4_{0v}/45 = \sigma^2_4$. Thus,

$$\frac{1}{N}\sum_{i=1}^{N}\left\{\frac{1}{T^2}\sum_{t=2}^{T}(y^*_{it} - \bar{y}^*_i)^2\right\} \xrightarrow{p} \frac{\sigma^2_{0v}}{6} = \mu_4 \qquad \text{(C.3)}$$

and $\zeta_{4NT} \xrightarrow{p} \sigma^2_{0v}/6 = \mu_4$ by a law of large numbers and the sequential limit theory. Moreover,

$$\begin{aligned}
&\sqrt{N}\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=2}^{T}\hat{e}^*_{it-1}\Delta\hat{e}^*_{it} \\
&\quad = \sqrt{N}\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=2}^{T}[y^*_{it-1} - \bar{y}^*_i - \hat{\beta}^*(x^*_{it-1} - \bar{x}^*_i)][u^*_{it} - \hat{\beta}^*\varepsilon^*_{it}]
\end{aligned}$$

$$= \sqrt{N}\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=2}^{T}[(y^*_{it-1}-\bar{y}^*_i)u^*_{it}-\hat{\beta}^*(x^*_{it-1}-\bar{x}^*_i)u^*_{it}-\hat{\beta}^*(y^*_{it-1}-\bar{y}^*_i)\varepsilon^*_{it}$$

$$+\hat{\beta}^{*2}(x^*_{it-1}-\bar{x}^*_i)]\varepsilon^*_{it} = \sqrt{N}\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=2}^{T}(y^*_{it-1}-\bar{y}^*_i)u^*_{it}+\mathrm{o_p}(1),$$

since

$$\frac{1}{N}\sum_{i=1}^{N}\left\{\frac{1}{T}\sum_{t=2}^{T}(x^*_{it-1}-\bar{x}^*_i)\varepsilon^*_{it}\right\}\xrightarrow{\mathrm{p}}-\frac{1}{2},$$

$$\frac{1}{N}\sum_{i=1}^{N}\left\{\frac{1}{T}\sum_{t=2}^{T}(y^*_{it-1}-\bar{y}^*_i)\varepsilon^*_{it}\right\}=\mathrm{o_p}(1),$$

and

$$\frac{1}{N}\sum_{i=1}^{N}\left\{\frac{1}{T}\sum_{t=2}^{T}(x^*_{it-1}-\bar{x}^*_i)u^*_{it}\right\}=\mathrm{o_p}(1)$$

by a law of large numbers and the sequential limit theory. Hence,

$$\sqrt{N}T(\hat{\rho}-1)=\frac{\sqrt{N}\xi_{3NT}}{\xi_{4NT}}$$

$$=\frac{\sqrt{N}\frac{1}{N}\sum_{i=1}^{N}\left\{\frac{1}{T}\sum_{t=2}^{T}(y^*_{it-1}-\bar{y}^*_i)u^*_{it}\right\}}{\frac{1}{N}\sum_{i=1}^{N}\left\{\frac{1}{T^2}\sum_{t=2}^{T}(y^*_{it-1}-\bar{y}^*_i)^2\right\}}+\mathrm{o_p}(1). \tag{C.4}$$

Next, we derive the asymptotic distribution of

$$\frac{\sqrt{N}\frac{1}{N}\sum_{i=1}^{N}\left\{\frac{1}{T}\sum_{t=2}^{T}(y^*_{it-1}-\bar{y}^*_i)u^*_{it}\right\}}{\frac{1}{N}\sum_{i=1}^{N}\left\{\frac{1}{T^2}\sum_{t=2}^{T}(y^*_{it-1}-\bar{y}^*_i)^2\right\}}=\frac{\sqrt{N}\xi'_{3NT}}{\xi_{4NT}}=\frac{\sqrt{N}\frac{1}{N}\sum_{i=1}^{N}\zeta'_{3iT}}{\frac{1}{N}\sum_{i=1}^{N}\zeta_{4iT}},$$

where $\zeta'_{3iT} = (1/T)\sum_{t=2}^{T}(y^*_{it-1} - \bar{y}^*_i)u^*_{it}$ and $\xi'_{3NT} = (1/N)\sum_{i=1}^{N}\zeta'_{3iT}$. But

$$\frac{1}{T}\sum_{t=2}^{T}(y^*_{it-1} - \bar{y}^*_i)u^*_{it} = T^{-1}\sum_{t=2}^{T} y^*_{it-1}u^*_{it} - \left(T^{-3/2}\sum_{t=2}^{T} y^*_{it}\right)\left(T^{-1/2}\sum_{t=2}^{T} u^*_{it}\right),$$

$$T^{-1}\sum_{t=2}^{T} y^*_{it-1}u^*_{it} \Rightarrow \frac{\sigma^2_{0v}}{2}\left[V^*_i(1)^2 - \frac{\sigma^2_v}{\sigma^2_{0v}}\right],$$

$$T^{-3/2}\sum_{t=2}^{T} y^*_{it} \Rightarrow \sigma_{0v}\int V^*_i,$$

and $T^{-1/2}\sum_{t=2}^{T}u^*_{it} \Rightarrow \sigma_{0v}V^*_i(1)$. Clearly,

$$\frac{1}{T}\sum_{t=2}^{T}(y^*_{it-1} - \bar{y}^*_i)u^*_{it} \Rightarrow \frac{\sigma^2_{0v}}{2}\left[V^*_i(1)^2 - \frac{\sigma^2_v}{\sigma^2_{0v}}\right] - \sigma^2_{0v}V^*_i(1)\int V^*_i = \zeta_{3i}, \tag{C.5}$$

$$\mathrm{E}[\zeta_{3i}] = -\sigma^2_v/2 = \mu_3,$$

and $Var[\zeta_{3i}] = \sigma^4_{0v}/12 = \sigma^2_3$, since $\mathrm{E}[V^*_i(1)\int V^*_i] = \frac{1}{2}$ and $\mathrm{E}[V^*_i(1)^2] = 1$.

Let $\mu_{3T} = \mathrm{E}[\zeta'_{3iT}] = \mu_3 + \mathrm{O}(T^{-1})$ and $\mu_{4T} = \mathrm{E}[\zeta_{4iT}] = \mu_4 + \mathrm{O}(T^{-1})$. The next step is to find an appropriate normalization of $(\xi'_{3NT}/\xi_{4NT} - \mu_{3T}/\mu_{4T})$ to make sure it converges to a proper random variable. For this purpose, we note the following standardization:

$$\frac{\sqrt{N}\xi'_{3NT}}{\xi_{4NT}} - \frac{\sqrt{N}\mu_{3T}}{\mu_{4T}} = \frac{\sqrt{N}(\zeta'_{3NT} - \mu_{3T})}{\xi_{4NT}} + \xi'_{3NT}\sqrt{N}\left[\frac{1}{\xi_{4NT}} - \frac{1}{\mu_{4T}}\right]. \tag{C.6}$$

Since $\{\xi_{3i}\}$ satisfies the conditions of the Lindeberg–Levy central limit theorem, we can readily see that the

$$\frac{1}{\sqrt{N}}\sum_{i=2}^{N}(\xi_{3i} - \mu_3) \Rightarrow \mathrm{N}(0, \sigma^2_3)$$

and the $(1/N)\sum_{i=2}^{N}\xi_{4i}$ will converge to μ_4 in probability as $N \to \infty$. Using the Slutsky theorem, we obtain

$$\frac{\frac{1}{\sqrt{N}}\sum_{i=2}^{N}(\xi_{3i} - \mu_3)}{\frac{1}{N}\sum_{i=2}^{N}\xi_{4i}} \Rightarrow \mathrm{N}\left(0, \frac{\sigma^2_3}{\mu^2_4}\right). \tag{C.7}$$

It follows that

$$\frac{\sqrt{N}(\zeta'_{3NT}-\mu_{3T})}{\xi_{4NT}} \Rightarrow \mathrm{N}\left(0, \frac{\sigma_3^2}{\mu_4^2}\right) \tag{C.8}$$

by the sequential limit theory. Similarly,

$$\xi'_{3NT}\sqrt{N}\left[\frac{1}{\zeta_{4NT}}-\frac{1}{\mu_{4T}}\right] \Rightarrow \mathrm{N}\left(0, \frac{\mu_3^2}{\mu_4^4}\sigma_4^2\right), \tag{C.9}$$

since it can be shown that $cov(\zeta'_{3iT}, \zeta_{4iT}) = \mathrm{O}(T^{-1})$ and $\sqrt{N}[\xi'_{3NT}-\mu_{3T}]$ and $\sqrt{N}[\xi_{4NT}-\mu_{4T}]$ are asymptotically uncorrelated. We, therefore, conclude

$$\sqrt{N}\left[\frac{\sqrt{N}\xi'_{3NT}}{\xi_{4NT}}-\frac{\mu_{3T}}{\mu_{4T}}\right] \Rightarrow \mathrm{N}\left(0, \frac{\sigma_3^2}{\mu_4^2}+\frac{\mu_3^2}{\mu_4^4}\sigma_4^2\right), \tag{C.10}$$

where

$$\frac{\sigma_3^2}{\mu_4^2} = \frac{\sigma_{0v}^4/12}{(\sigma_{0v}^2/6)^2} = 3 \quad \text{and} \quad \frac{\mu_3^2}{\mu_4^4}\sigma_4^2 = \frac{(\sigma_v^2/2)^2}{(\sigma_{0v}^2/6)^4}\frac{\sigma_{0v}^4}{45} = \frac{36}{5}\frac{\sigma_v^4}{\sigma_{0v}^4}.$$

Thus,

$$\sqrt{N}T(\hat{\rho}-1)-\frac{\sqrt{N}\mu_{3T}}{\mu_{4T}} \Rightarrow \mathrm{N}\left(0, 3+\frac{36\sigma_v^4}{5\sigma_{0v}^4}\right). \tag{C.11}$$

The t-statistic is

$$t_\rho = \frac{(\hat{\rho}-1)\sqrt{\sum_{i=1}^N\sum_{t=2}^T \hat{e}_{it-1}^{*2}}}{s_e} = \frac{\sqrt{N}T(\hat{\rho}-1)\sqrt{\frac{1}{N}\sum_{i=1}^N\frac{1}{T^2}\sum_{t=2}^T\hat{e}_{it-1}^{*2}}}{s_e}$$

$$= \frac{\sqrt{N}T(\hat{\rho}-1)\sqrt{\zeta_{4NT}}}{s_e} = \frac{\dfrac{\sqrt{N}\xi_{3NT}}{\zeta_{4NT}}\sqrt{\xi_{4NT}}}{s_e}$$

$$= \frac{\sqrt{N}\xi_{3NT}}{s_e\sqrt{\xi_{4NT}}} = \frac{\sqrt{N}\xi'_{3NT}}{s_e\sqrt{\xi_{4NT}}}+\mathrm{o}(1), \tag{C.12}$$

where $s_e^2 = (1/NT)\sum_{i=1}^{N}\sum_{t=2}^{T}(\hat{e}_{it}^{*} - \hat{\rho}\hat{e}_{it-1}^{*})^2$. It follows that

$$s_e^2 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}\{(u_{it}^{*} - \hat{\beta}^{*}\varepsilon_{it}^{*})\}^2 + o_p(1)$$
$$= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}(u_{it}^{*2} - 2\hat{\beta}^{*}u_{it}^{*}\varepsilon_{it}^{*} + \hat{\beta}^{*2}\varepsilon_{it}^{*2}) + o_p(1) \xrightarrow{p} \sigma_v^2.$$

That is, s_e^2 is a consistent estimator of σ_v^2. Next, we derive the asymptotic distribution of $\sqrt{N}\xi'_{3NT}/s_e\sqrt{\xi_{4NT}}$.

We use the same approach for the asymptotic distribution of $\sqrt{N}\xi'_{3NT}/s\sqrt{\xi_{4NT}}$ as we used for $\sqrt{N}T(\hat{\rho}-1)$. We make the following normalization:

$$\frac{N\xi'_{3NT}}{s_e\sqrt{\xi_{4NT}}} - \frac{N\mu_{3T}}{s_e\mu_{4T}}$$
$$= \frac{N\xi'_{3NT}}{s_e\xi_{4NT}} - \frac{N\mu_{3T}}{s_e\mu_{4T}} = \frac{N[\xi'_{3NT} - \mu_{3T}]}{s_e\xi_{4NT}} + \frac{\mu_{3T}}{s_e}N\left[\frac{1}{\xi_{4NT}} - \frac{1}{\mu_{4T}}\right]. \tag{C.13}$$

Since

$$\frac{N[\xi'_{3NT} - \mu_{3T}]}{s_e\xi_{4NT}} \Rightarrow N\left(0, \frac{\sigma_3^2}{\mu_4\sigma_v^2}\right), \tag{C.14}$$

$$\frac{\mu_{3T}}{s_e}\sqrt{N}[\sqrt{1/\xi_{4NT}} - \sqrt{1/\mu_{4T}}] \Rightarrow N\left(0, \frac{1}{4}\frac{\mu_3^2\sigma_4^2}{\mu_4^3\sigma_v^2}\right),$$

and $\sqrt{N}[\xi'_{3NT} - \mu_{3T}]$ and $\sqrt{N}[\xi_{4NT} - \mu_{4T}]$ are asymptotically uncorrelated, we note that

$$\frac{N\xi'_{3NT}}{s_e\xi_{4NT}} - \frac{N\mu_{3T}}{s_e\sqrt{\mu_{4T}}} \Rightarrow N\left(0, \frac{\sigma_3^2}{\mu_4\sigma_v^2} + \frac{1}{4}\frac{\mu_3^2\sigma_4^2}{\mu_4^3\sigma_v^2}\right)$$

where

$$\frac{\sigma_3^2}{\mu_4\sigma_v^2} = \frac{\frac{\sigma_{0v}^4}{12}}{\frac{\sigma_{0v}^2}{6}\sigma_v^2} = \frac{\sigma_{0v}^2}{2\sigma_v^2} \quad \text{and} \quad \frac{1}{4}\frac{\mu_3^2\sigma_4^2}{\mu_4^3\sigma_v^2} = \frac{1}{4}\frac{\left(\frac{\sigma_v^2}{2}\right)^2\frac{\sigma_{0v}^4}{45}}{\left(\frac{\sigma_{0v}^2}{6}\right)^3\sigma_v^2} = \frac{3\sigma^2 v}{10\sigma_{0v}^2}.$$

Thus,

$$t_\rho - \frac{\sqrt{N}\mu_{3T}}{s_e\sqrt{\mu_{4T}}} \Rightarrow \mathrm{N}\left(0, \frac{\sigma_{0v}^2}{2\sigma_v^2} + \frac{3\sigma_v^2}{10\sigma_{0v}^2}\right).$$

Therefore, we established Theorem 3. □

Appendix D. Proof of Theorem 3

Proof. Observe that

$$\begin{aligned}\xi_{6NT} = \frac{1}{N}\sum_{i=1}^{N} \zeta_{6iT} = \frac{1}{N}\sum_{i=1}^{N} \frac{1}{T^2}(e_i^{*\prime}Q_i^* e_i^*) \\ = \frac{1}{N}\sum_{i=1}^{N} \frac{1}{T^2}(e_i^{*\prime}e_i^*) + \mathrm{o_p}(1) = \frac{1}{N}\sum_{i=1}^{N} \frac{1}{T^2}\sum_{t=2}^{T} (\hat{e}_{it-1}^{*2}) \\ + \mathrm{o_p}(1) = \frac{1}{N}\sum_{i=1}^{N} \left\{\frac{1}{T^2}\sum_{t=2}^{T} (y_{it}^* - \bar{y}_i^*)^2\right\} + \mathrm{o_p}(1).\end{aligned}$$

We also note that

$$\frac{1}{T^2}\sum_{t=2}^{T} (y_{it}^* - \bar{y}_i^*)^2 \Rightarrow \sigma_{0v}^2\left[\int V_i^{*2} - \left\{\int V_i^*\right\}^2\right] = \zeta_{6i}, \tag{D.1}$$

$$\mathrm{E}[\zeta_{6i}] = \frac{\sigma_{0v}^2}{6} = \mu_6,$$

and

$$Var[\zeta_{6i}] = \sigma_{0v}^4/45 = \sigma_6^2.$$

Under the null hypothesis that $\rho = 1$, expression (9) can be written as

$$d(L)\Delta\hat{e}_{it}^* = v_{it}, \tag{D.2}$$

where $d(L) = (1 - \varphi_1 L - \varphi_2 L^2 - \cdots + \varphi_p L^p)$ and L is the backshift operator. Moreover,

$$\sqrt{N}\xi_{5NT} = \sqrt{N}\frac{1}{N}\sum_{i=1}^{N} \zeta_{5NT} = \sqrt{N}\frac{1}{N}\sum_{i=1}^{N} \frac{1}{T}(e_i^{*\prime}Q_i^* v_i)$$

$$= \sqrt{N}\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}(e_i^{*\prime}v_i) + \mathrm{o_p}(1) = \sqrt{N}\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=2}^{T}(\hat{e}_{it-1}^{*}v_{it})$$

$$+ \mathrm{o_p}(1) = \sqrt{N}\frac{1}{N}\sum_{i=1}^{N}\left\{\frac{1}{T}\sum_{t=2}^{T}(y_{it}^{*} - \bar{y}_i^{*})d(L)u_{it}^{*}\right\} + \mathrm{o_p}(1).$$

Since

$$\sqrt{N}\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{T=2}^{T}\hat{e}_{it-1}^{*}v_{it}$$

$$= \sqrt{N}\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\left[\sum_{t=2}^{T}\hat{e}_{it-1}^{*}(1 - \varphi_1 L - \varphi_2 L^2 - \cdots - \varphi_p L^p)\Delta\hat{e}_{it}^{*}\right]$$

$$= \sqrt{N}\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\left[\sum_{t=2}^{T}\hat{e}_{it-1}^{*}\Delta\hat{e}_{it}^{*} - \varphi_1\hat{e}_{it-1}^{*}\Delta\hat{e}_{it-1}^{*}\right.$$

$$\left. - \varphi_2\hat{e}_{it-1}^{*}\Delta\hat{e}_{it-2}^{*} - \cdots - \varphi_p\hat{e}_{it-1}^{*}\Delta\hat{e}_{it-p}^{*}\right]$$

$$= \frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\sum_{t=2}^{T}[(y_{it-1}^{*} - \bar{y}_i^{*})(u_{it} - \varphi_1 u_{it-1}$$

$$- \varphi_2 u_{it-2} - \cdots - \varphi_p u_{it-p})] + \mathrm{o_p}(1),$$

it follows that

$$\sqrt{N}\xi_{5NT} = \sqrt{N}\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=2}^{T}\hat{e}_{it-1}^{*}v_{it} + \mathrm{o_p}(1)$$

$$= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}(y_{it-1}^{*} - \bar{y}_i^{*})(u_{it} - \varphi_1 u_{it-1} - \varphi_2 u_{it-2} - \cdots$$

$$- \varphi_p u_{it-p}) + \mathrm{o_p}(1).$$

Hence,

$$t_{ADF} = \frac{\sqrt{N}\xi_{5NT}}{s_v\sqrt{\xi_{6NT}}} = \frac{\sqrt{N}\frac{1}{N}\sum_{i=1}^{N}\left\{\frac{1}{T}\sum_{t=2}^{T}(y_{it-1} - \bar{y}_i)d(L)u_{it}^{*}\right\}}{s_v\sqrt{\xi_{6NT}}} + \mathrm{o_p}(1)$$

$$= \frac{\sqrt{N}\xi_{5NT}'}{s_v\sqrt{\xi_{6NT}}} + \mathrm{o_p}(1),$$

where

$$\zeta'_{5iT} = \frac{1}{T}\sum_{t=2}^{T} (y^*_{it-1} - \bar{y}^*_i) d(L) u^*_{it} \quad \text{and} \quad \xi'_{5NT} = \frac{1}{N}\sum_{i=1}^{N} \zeta'_{5iT}.$$

Assume that the sequence $\{u^*_{it}\}$ satisfies condition (C2) in Phillips and Ouliaris (1990) so that

$$\frac{1}{T}\sum_{t=2}^{[rT]} d(L)u^*_{it} \Rightarrow d(1)\sigma_{0v} V^*_i(r)$$

for all i. Consequently,

$$\frac{1}{T}\sum_{t=2}^{T} y^*_{it-1} d(L) u^*_{it} \Rightarrow d(1)\sigma^2_{0v} \int V^*_i \, \mathrm{d}V^*_i$$

and

$$\frac{1}{T}\sum_{t=2}^{T} \bar{y}^*_i d(L) u^*_{it} \Rightarrow d(1)\sigma^2_{0v} V^*_i(1) \int V^*_i.$$

for all i. Clearly,

$$\frac{1}{T}\sum_{t=2}^{T} (y^*_{it-1} - \bar{y}^*_i)\, \mathrm{d}(L) u^*_{it} \Rightarrow d(1)\sigma^2_{0v} \int V^*_i \, \mathrm{d}V^*_i - d(1)\sigma^2_{0v} V^*_i(1) \int V^*_i = \zeta_{5i}$$

for all i, with $\mathrm{E}(\zeta_{5i}) = -d(1)\sigma^2_{0v}/2$ and $Var(\zeta_{5i}) = d^2(1)\sigma^4_{0v}/12$. Let

$$s^2_v = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T} \{(\hat{e}^*_{it} - \hat{\rho}^* \hat{e}_{it-1}) - (\hat{\varphi}_1 L + \hat{\varphi}_2 L^2 + \cdots + \hat{\varphi}_p L^p)\Delta\hat{e}^*_{it}\}^2.$$

Observe that $\{\hat{e}^*_{it}\}$ are linear combinations of $u^*_{it}$and ε^*_{it}. Those linear combinations are in general ARMA processes. Therefore, we need to have $p \to \infty$ as $T \to \infty$ such that $p = o(T^{1/3})$ to capture the true structure of v_{it}. As a result of $p \to \infty$, $\hat{\rho} \xrightarrow{\mathrm{p}} 1$ and the estimated coefficients $\{\hat{\varphi}_i\}$ may well represent $\{\varphi_i\}$:

$$\begin{aligned} s^2_v &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T} \{(1 - \varphi_1 L - \varphi_2 L^2 - \cdots - \varphi_p L^p)(u^*_{it} - \hat{\beta}^* \varepsilon^*_{it})\}^2 + o_p(1) \\ &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T} d^2(L)(u^{*2}_{it} - 2\hat{\beta}^* u^*_{it}\varepsilon^*_{it} + \hat{\beta}^{*2}\varepsilon^{*2}_{it}) + o_p(1) \xrightarrow{\mathrm{p}} d^2(1)\sigma^2_v = \sigma^2 \end{aligned}$$

by the sequential limit theory. We are now in a position to derive the asymptotic distribution of $\sqrt{N}\xi'_{5NT}/s_v\sqrt{\xi_{6NT}}$. Define

$$\mu_{5T} = \mathrm{E}(\zeta'_{5iT}) = \mathrm{E}(\zeta'_{5i}) + \mathrm{o_p}(1)$$

and

$$\mu_{6T} = \mathrm{E}(\zeta_{6iT}) = \mathrm{E}(\zeta_{6i}) + \mathrm{o_p}(1).$$

Then, we make the following normalization:

$$t_{ADF} - \frac{\sqrt{N}\mu_{5T}}{s_v\mu_{6T}} = \frac{\sqrt{N}\xi'_{5NT}}{s_v\sqrt{\xi_{6NT}}} - \frac{\sqrt{N}\mu_{5T}}{s_v\sqrt{\mu_{6T}}} = \frac{\sqrt{N}(\xi'_{5NT}-\mu_{5T})}{s_v\sqrt{\xi_{6NT}}} + \frac{\mu_{5T}}{s_v}\sqrt{N}(\sqrt{1/\zeta_{6NT}} - \sqrt{1/\mu_{6T}}).$$

Since

$$\frac{\sqrt{N}(\xi'_{5NT}-\mu_{5T})}{s_v\sqrt{\xi_{6NT}}} \Rightarrow \mathrm{N}\left(0, \frac{\frac{d^2(1)\sigma_{0v}^4}{12}}{\frac{\sigma^2\sigma_{0v}^2}{6}}\right) = \mathrm{N}\left(0, \frac{d^2(1)\sigma_{0v}^2}{2\sigma^2}\right)$$

$$\frac{\mu_{5T}}{s_v}\sqrt{N}\left(\sqrt{\frac{1}{\xi_{6NT}}} - \sqrt{\frac{1}{\mu_{6T}}}\right) \Rightarrow \mathrm{N}\left(0, \frac{1}{4}\frac{\left(-d(1)\frac{\sigma_v^2}{2}\right)^2\frac{\sigma_{0v}^4}{45}}{\sigma^2\left(\frac{\sigma_{0v}^2}{6}\right)^3}\right) = \mathrm{N}\left(0, \frac{3d^2(1)\sigma_v^4}{10\sigma^2\sigma_{0v}^2}\right)$$

and $\sqrt{N}(\xi_{5NT}-\mu_{5T})$ and $\sqrt{N}(\xi_{6NT}-\mu_{6T})$ are asymptotically uncorrelated, we conclude that

$$t_{ADF} - \frac{\sqrt{N}\mu_{5T}}{s_v\sqrt{\mu_{6T}}} \Rightarrow \mathrm{N}\left(0, \frac{d^2(1)\sigma_{0v}^2}{2d^2(1)\sigma_v^2} + \frac{3d^2(1)\sigma_v^4}{10d^2(1)\sigma_v^2\sigma_{0v}^2}\right) = \mathrm{N}\left(0, \frac{\sigma_{0v}^2}{2\sigma_v^2} + \frac{3\sigma_v^2}{10\sigma_{0v}^2}\right). \qquad \square$$

Appendix E. Proof of Theorem 4

Proof. The OLS of ρ of Eq. (15) is

$$\hat{\rho} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{y}_{it}\tilde{y}_{it-1}}{\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{y}_{it-1}^{2}}, \tag{E.1}$$

where $\tilde{y}_{it} = y_{it} - \bar{y}_i - d_i$, $d_i = \sum_{t=1}^{T}(t-\bar{t})y_{it}/\sum_{t=1}^{T}(t-\bar{t})^2$, $\bar{t} = (T+1)/2$, and $\bar{y}_i = (1/T)\sum_{t=1}^{T}y_{it}$. Under the null of no cointegration, $\rho = 1$, we have

$$\sqrt{N}T(\hat{\rho}-1) = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{y}_{it-1}\varepsilon_{it}}{\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{y}_{it-1}^{2}} = \frac{\sqrt{N}\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T}\tilde{y}_{it-1}\varepsilon_{it}}{\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T^2}\sum_{t=1}^{T}\tilde{y}_{it-1}^{2}}. \tag{E.2}$$

Consider the denominator in Eq. (E.2) from Phillips and Perron (1988):

$$\frac{1}{T^2}\sum_{t=1}^{T}\tilde{y}_{it-1}^{2} \Rightarrow \sigma_{0u}^{2}\int\tilde{V}^2, \tag{E.3}$$

where $\tilde{V}(r)$ is the standard demeaned and detrended Wiener process

$$\tilde{V}(r) = V(r) + (6r-4)\int V + (-12r+6)\int sV. \tag{E.4}$$

Note that $\mathrm{E}[\sigma_{0u}^{2}\int\tilde{V}^2] = \sigma_{0u}^{2}/15 = \mu_8$ and $Var[\sigma_{0u}^{2}\int\tilde{V}^2] = 11\sigma_{0u}^{4}/6300$. It follows that

$$\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T^2}\sum_{t=1}^{T}\tilde{y}_{it-1}^{2} \xrightarrow{p} \frac{\sigma_{0u}^{2}}{15}$$

by a law of large numbers and the sequential limit theory. Next consider the numerator from Phillips and Perron (1988):

$$\frac{1}{T}\sum_{t=1}^{T}\tilde{y}_{it-1}\varepsilon_{it} \Rightarrow \sigma_{0u}^{2}\int\tilde{V}\,\mathrm{d}V + \lambda \tag{E.5}$$

where $\lambda = \frac{1}{2}(\sigma_{0u}^{2} - \sigma_{u}^{2})$. It can be shown that

$$\mathrm{E}\left[\int\tilde{V}\,dV\right] = -\frac{\sigma_{0u}^{2}}{2} + \lambda = -\frac{\sigma_{u}^{2}}{2} = \mu_9 \tag{E.6}$$

and $Var[\int \tilde{V}\,\mathrm{d}V] = \sigma_{0u}^4/60$. Then

$$\sqrt{N}\left[T(\hat{\rho}-1) - \frac{\mu_{7T}}{\mu_{8T}}\right] \Rightarrow \mathrm{N}\left(0, \frac{15}{4} + \frac{\sigma_u^4}{\sigma_{0u}^4}\frac{2475}{112}\right), \tag{E.7}$$

by the Lindeberg–Levy central limit theorem and the sequential limit theory, where

$$\mu_{7T} = \mathrm{E}\left[\frac{1}{T}\sum_{t=1}^{T} \tilde{y}_{it-1}\varepsilon_{it}\right] \tag{E.8}$$

and

$$\mu_{8T} = \mathrm{E}\left[\frac{1}{T^2}\sum_{t=1}^{T} \tilde{y}_{it-1}^2\right]. \tag{E.9}$$

The t-statistic to test $\rho = 1$, t_ρ, is

$$t_\rho = \frac{(\hat{\rho}-1)\sqrt{\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{y}_{it-1}^2}}{s_e},$$

where

$$s_e^2 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}(\hat{\varepsilon}_{it})^2. \tag{E.10}$$

The asymptotic distribution of t_ρ can be found using the same approach as in Eqs. (E.2), (E.3), (E.4), (E.5), (E.6), (E.7), (E.8) and (E.9)

$$t_\rho - \frac{\sqrt{N}\mu_{7T}}{s_e\sqrt{\mu_{8T}}} \Rightarrow \mathrm{N}\left(0, \frac{1}{4}\frac{\sigma_{0u}^2}{\sigma_u^2} + \frac{165}{448}\frac{\sigma_u^2}{\sigma_{0u}^2}\right). \tag{E.11}$$

Finally, it can be shown that the ADF test, t_{ADF}, has the same asymptotic distribution as the t_ρ using the same approach as in Section 3.2. Therefore, we have established Theorem 4. □

References

Billingsley, P., 1986. Probability and Measure. John Wiley, New York.

Breitung, J., Meyer, W., 1994. Testing for unit roots in panel data: Are wages on different bargaining levels cointegrated? Applied Economics 26, 353–361.

Engle, R.F., Granger, C.W.J., 1987. Co-integration and error correction: Representation, estimation, and testing. Econometrica 55, 251–276.
Gonzalo, J., 1994. Five alternative methods of estimating long-run equilibrium relationships. Journal of Econometrics 60, 203–233.
Hansen, B., 1992. Efficient estimation and testing of cointegrating vectors in the presence of deterministic trends. Journal of Econometrics 53, 87–121.
Haug, A., 1996. Tests for cointegration: A Monte Carlo comparison. Journal of Econometrics 71, 89–115.
Im K., Pesaran, H., Shin, Y., 1997. Testing for unit roots in heterogeneous panels. Manuscript, Department of Applied Economics, University of Cambridge.
Kao, C., Chen, B., 1995a. On the estimation and inference of a cointegrated regression in panel data when the cross-section and time-series dimensions are comparable. Manuscript, Department of Economics, Syracuse University.
Kao, C., Chen, B., 1995b. Residual-based tests for cointegration in panel data when the cross-section and time-series dimensions are comparable. Manuscript, Department of Economics, Syracuse University.
Kao, C., Chiang, M.-H., 1997. On the estimation and inference of a cointegrated regression in panel data. Econometric Theory, Under Revision.
Levin, A., Lin, C.-F., 1993. Unit root tests in panel data: New results. Discussion Paper, Department of Economics, UC-San Diego.
McCoskey, S., Kao, C., 1998. A residual-based test of the null of cointegration in panel data. Econometric Reviews 17, 57–84.
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., Smith, R., 1995. Estimating long-run relationships from dynamic heterogeneous panels. Journal of Econometrics 68, 79–113.
Phillips, P.C.B., Moon, H., 1997. Linear regression limit theory for nonstationary panel data. Cowles Foundation Discussion Paper.
Phillips, P.C.B., Ouliaris, S., 1990. Asymptotic properties of residual based test for cointegration. Econometrica 58, 165–193.
Phillips, P.C.B., Perron, P., 1988. Testing for unit root in time series regression. Biometrika 75, 335–346.
Quah, D., 1994. Exploiting cross section variation for unit root inference in dynamic data. Economics Letters 44, 9–19.