ELSEVIER

Journal of Econometrics 108 (2002) 1–24

JOURNAL OF Econometrics

www.elsevier.com/locate/econbase

Unit root tests in panel data: asymptotic and finite-sample properties

Andrew Levin[a], Chien-Fu Lin[b], Chia-Shang James Chu[b,*]

[a]*Federal Reserve Board of Governors, Washington, DC 20551, USA*
[b]*Department of Economics, National Taiwan University, No. 21 Hsu Chou Road, Taipei, Taiwan, ROC*

Received 19 May 1992; revised 4 June 2001; accepted 13 July 2001

Abstract

We consider pooling cross-section time series data for testing the unit root hypothesis. The degree of persistence in individual regression error, the intercept and trend coefficient are allowed to vary freely across individuals. As both the cross-section and time series dimensions of the panel grow large, the pooled t-statistic has a limiting normal distribution that depends on the regression specification but is free from nuisance parameters. Monte Carlo simulations indicate that the asymptotic results provide a good approximation to the test statistics in panels of moderate size, and that the power of the panel-based unit root test is dramatically higher, compared to performing a separate unit root test for each individual time series. © 2002 Elsevier Science B.V. All rights reserved.

JEL classification: C12; C23

Keywords: ADF regression; Nonstationary panel; Panel unit root test; Pooled t-statistics

1. Introduction

A large body of literature during the past two decades has considered the impact of integrated time series in econometric research (cf. surveys by Diebold and Nerlove, 1990; Campbell and Perron, 1991). In univariate analysis, the Box–Jenkins (Box and Jenkins, 1970) approach of studying difference-stationary ARMA models requires a consistent and powerful test

* Corresponding author.

PII: S0304-4076(01)00098-7

for the presence of unit roots. In general, such tests have non-standard limiting distributions; for example, the original Dickey–Fuller test (Dickey and Fuller, 1979) and the subsequent augmented Dickey–Fuller (ADF) test statistic (Dickey and Fuller, 1981) converge to a function of Brownian motion under quite general conditions (Said and Dickey, 1984). The critical values of the empirical distribution were first tabulated by Dickey (cf. Fuller, 1976). Semi-parametric test procedures have also been proposed (i.e. Phillips, 1987; Phillips and Perron, 1988), with improved empirical size and power properties under certain conditions (cf. Diebold and Nerlove, 1990).

In finite samples, these unit root test procedures are known to have limited power against alternative hypotheses with highly persistent deviations from equilibrium. Simulation exercises also indicate that this problem is particularly severe for small samples (see Campbell and Perron, 1991). This paper considers pooling cross-section time series data as a means of generating more powerful unit root tests. The test procedures are designed to evaluate the null hypothesis that each individual in the panel has integrated time series versus the alternative hypothesis that all individuals time series are stationary. The pooling approach yields higher test power than performing a separate unit root test for each individual.

Some earlier work has analyzed the properties of panel-based unit root tests under the assumption that the data is identically distributed across individuals. Quah (1990, 1994) used random field methods to analyze a panel with i.i.d. disturbances, and demonstrated that the Dickey–Fuller test statistic has a standard normal limiting distribution as both the cross-section and time series dimensions of the panel grow arbitrarily large. Unfortunately, the random field methodology does not allow either individual-specific effects or aggregate common factors (Quah, 1990, p. 17). Breitung and Meyer (1991) have derived the asymptotic normality of the Dickey–Fuller test statistic for panel data with an arbitrarily large cross-section dimension and a small fixed time series dimension (corresponding to the typical microeconomic panel data set). Their approach allows for time-specific effects and higher-order serial correlation, as long as the pattern of serial correlation is identical across individuals, but cannot be extended to panel with heterogeneous errors. More recent advances in nonstationary panel analysis include Im et al. (1995), Harris and Tzavalis (1996) and Phillips and Moon (1999), among others.

What type of the asymptotics considered in the panel unit root test is a delicate issue. Earlier work by Anderson and Hsiao (1982) consider a stationary panel with fixed time series observations while letting the cross sectional units grow arbitrarily large. Similar asymptotic method is used in nonstationary panel by Breitung and Meyer (1991) and Harris and Tzavalis (1996). Im et al. (1995) and Quah (1990, 1994) explore the case of joint limit in which both time series and cross sectional dimension approach infinity with certain restrictions. The precise meaning regarding in what way the cross sectional and time series dimension approach infinity has been clearly defined in Phillips and Moon (1999). This article is one of the earlier research that consider the joint limit asymptotics in which both N and T approach infinity subjecting to certain conditions such as $\sqrt{N}/T \rightarrow 0$ in some model, and $N/T \rightarrow 0$ in others.

The panel-based unit root test proposed in this article allows for individual-specific intercepts and time trends. Moreover, the error variance and the pattern of higher-order serial correlation are also permitted to vary freely across individuals. Our asymptotic analysis in Section 3 indicates that the proposed test statistics have an interesting mixture of the asymptotic properties of stationary panel data and the asymptotic properties of integrated time series data. In contrast to the non-standard distributions of unit root test statistics for a single time series, the panel test statistics have limiting normal distributions, as in the case of stationary panel data (cf. Hsiao, 1986). However, in contrast to the results for stationary panel data, the convergence rate of the test statistics is higher with respect to the number of time periods (referred to as "super-consistency" in the time series literature) than with respect to the number of individuals in the sample. Furthermore, whereas regression t-statistics for stationary panel data converge to the standard normal distribution, we find that the asymptotic mean and variance of the unit root test statistics vary under different specification of the regression equation (i.e. the inclusion of individual-specific intercepts and time trends).

For practical purposes, the panel based unit root tests suggested in this paper are more relevant for panels of moderate size. If the time series dimension of the panel is very large then existing unit root test procedures will generally be sufficiently powerful to be applied separately to each individual in the panel, though pooling a small group of individual time series can be advantageous in handling more general patterns of correlation across individuals (cf. Park, 1990; Johansen, 1991). On the other hand, if the time series dimension of the panel is very small, and the cross-section dimension is very large, then existing panel data procedures will be appropriate (cf. MaCurdy, 1982; Hsiao, 1986; Holtz-Eakin et al., 1988; Breitung and Meyer, 1991). However, panels of moderate size (say, between 10 and 250 individuals, with 25–250 time series observations per individual) are frequently encountered in industry-level or cross-country econometric studies. For panels of this size, standard multivariate time series and panel data procedures may not be computationally feasible or sufficiently powerful, so that the unit root test procedures outlined in this paper will be particularly useful.

The remainder of this paper is organized as follows: Section 2 specifies the assumptions and outlines the panel unit root test procedure. Readers who are interested mainly in empirical application can skip the rest of the paper. Section 3 analyzes the limiting distributions of the panel test statistics. Section 4 briefly discusses the Monte Carlo simulations. Concluding remarks regarding the limitations of the proposed panel unit root test are offered in Section 5. All proofs are deferred to the appendix.

2. A panel unit root test

2.1. Model specifications

We observe the stochastic process $\{y_{it}\}$ for a panel of individuals $i=1,\dots,N$, and each individual contains $t=1,\dots,T$ time series observations. We wish to determine whether $\{y_{it}\}$ is integrated for each individual in the panel. As in the case of a single time series, the individual regression may include an intercept and time trend. We assume that all individuals in the panel have identical first-order partial autocorrelation, but all other parameters in the error process are permitted to vary freely across individuals.

Assumption 1.

(a) Assume that $\{y_{it}\}$ is generated by one of the following three models:

Model 1: $\Delta y_{it} = \delta y_{it-1} + \zeta_{it}$.
Model 2: $\Delta y_{it} = \alpha_{0i} + \delta y_{it-1} + \zeta_{it}$.
Model 3: $\Delta y_{it} = \alpha_{0i} + \alpha_{1i} t + \delta y_{it-1} + \zeta_{it}$, where $-2 < \delta \leqslant 0$ for $i=1,\dots,N$.

(b) The error process ζ_{it} is distributed independently across individuals and follows a stationary invertible ARMA process for each individual,

$$\zeta_{it} = \sum_{j=1}^{\infty} \theta_{ij} \zeta_{it-j} + \varepsilon_{it}.$$

(c) For all $i=1,\dots,N$ and $t=1,\dots,T$,

$$\mathrm{E}(\zeta_{it}^4) < \infty;\ \mathrm{E}(\varepsilon_{it}^2) \geqslant B_\varepsilon > 0;\ \text{and}\ \mathrm{E}(\zeta_{it}^2) + 2\sum_{j=1}^{\infty} \mathrm{E}(\zeta_{it}\zeta_{it-j}) < B_\zeta < \infty.$$

Assumption 1(a) includes three data generating processes. In Model 1, the panel unit root test procedure evaluates the null hypothesis H_0: $\delta = 0$ against the alternative H_1: $\delta < 0$. The series $\{y_{it}\}$ has an individual-specific mean in Model 2, but does not contain a time trend. In this case, the panel test procedure evaluates the null hypothesis that H_0: $\delta = 0$ and $\alpha_{0i} = 0$, for all i, against H_1: $\delta < 0$ and $\alpha_{0i} \in R$. Finally, under Model 3, the series $\{y_{it}\}$ has an individual-specific mean and time trend. In this case, the panel test procedure evaluates the null hypothesis that H_0: $\delta = 0$ and $\alpha_{1i} = 0$, for all i, against the alternative H_1: $\delta < 0$ and $\alpha_{1i} \in R$.

As in the case of a single time series, if a deterministic element (e.g. an intercept or time trend) is present but not included in the regression procedure, the unit root test will be inconsistent. On the other hand, if a deterministic element is included in the regression procedure but is not present in the observed data, the statistical power of the unit root test will be reduced. See Johansen's (1992) discussion on the interactions of the unit root test and various deterministic specifications. Campbell and Perron (1991) outline a method of ascertaining which deterministic elements should be included in the test procedure. For notational simplicity, d_{mt} is used to indicate the vector of deterministic variables and α_m is used to indicate the corresponding vector of coefficients for a particular model $m = 1,2,3$. Thus, $d_{1t} = \emptyset$ (the empty set); $d_{2t} = \{1\}$ and $d_{3t} = \{1, t\}$.

Assumption 1(b) is standard; individual time series may exhibit serial correlations. The finite-moment conditions of Assumption 1(c) correspond to the conditions for the weak convergence in Phillips (1987) and Phillips-Perron's (Phillips and Perron, 1988) unit root tests. In our panel unit root test we define a ratio of the long-run variance to innovation variance (cf. Eq. (6)), the boundedness conditions in Assumption 1(c) ensure this ratio remains finite for every individual in the panel as the cross-section N becomes arbitrarily large.

2.2. Test procedures

Our maintain hypothesis is

$$\Delta y_{it} = \delta y_{it-1} + \sum_{L=1}^{P_i} \theta_{iL} \Delta y_{it-L} + \alpha_{mi} d_{mt} + \varepsilon_{it}, \quad m = 1,2,3. \tag{1}$$

However, since p_i is unknown, we therefore suggest a three-step procedure to implement our test. In step 1 we carry out separate ADF regressions for each individual in the panel, and generate two orthogonalized residuals. Step 2 requires estimating the ratio of long run to short run innovation standard deviation for each individual. In the final step we compute the pooled t-statistics.

2.2.1. Step 1: Perform ADF regressions and generate orthogalized residuals

For each individual i, we implement the ADF regression

$$\Delta y_{it} = \delta_i y_{it-1} + \sum_{L=1}^{P_i} \theta_{iL} \Delta y_{it-L} + \alpha_{mi} d_{mt} + \varepsilon_{it}, \quad m = 1,2,3. \tag{1$'$}$$

The lag order p_i is permitted to vary across individuals. Campbell and Perron (1991) recommend the method proposed by Hall (1990) for selecting the appropriate lag order: for a given sample length T, choose a maximum lag order $p_{\max}$, and then use the t-statistics of $\hat{\theta}_{iL}$ to determine if a smaller lag order is preferred. (These t-statistics have a standard normal distribution under the null hypothesis ($\theta_{iL}=0$), both when $\delta_i=0$ and when $\delta_i<0$.

Having determined autoregression order p_i in ($1'$), we run two auxiliary regressions to generate orthogalized residuals. Regress Δy_{it} and y_{it-1} against Δy_{it-L} ($L=1,\dots,p_i$) and the appropriate deterministic variables, d_{mt}, then save the residuals $\hat{e}_{it}$ and $\hat{v}_{it-1}$ from these regressions. Specifically,

$$\hat{e}_{it}=\Delta y_{it}-\sum_{L=1}^{P_i}\hat{\pi}_{iL}\Delta y_{it-L}-\hat{\alpha}_{mi}d_{mt} \tag{2}$$

and

$$\hat{v}_{it-1}=y_{it-1}-\sum_{L=1}^{P_i}\tilde{\pi}_{iL}\Delta y_{it-L}-\tilde{\alpha}_{mi}d_{mt}.$$

To control for heterogeneity across individuals, we further normalize $\hat{e}_{it}$ and $\hat{v}_{it-1}$ by the regression standard error form Eq. ($1'$).

$$\tilde{e}_{it}=\frac{\hat{e}_{it}}{\hat{\sigma}_{\varepsilon i}},\qquad \tilde{v}_{it-1}=\frac{\hat{v}_{it-1}}{\hat{\sigma}_{\varepsilon i}}, \tag{3}$$

where $\hat{\sigma}_{\varepsilon i}$ is the regression standard error in ($1'$). Equivalently, it can also be calculated from the regression of $\hat{e}_{it}$ against $\hat{v}_{it-1}$

$$\hat{\sigma}^2_{\varepsilon i}=\frac{1}{T-p_i-1}\sum_{t=p_i+2}^{T}(\hat{e}_{it}-\hat{\delta}_i\hat{v}_{it-1})^2. \tag{4}$$

2.2.2. Step 2: Estimate the ratio of long-run to short-run standard deviations

Under the null hypothesis of a unit root, the long-run variance for Model 1 can be estimated as follows:

$$\hat{\sigma}^2_{yi}=\frac{1}{T-1}\sum_{t=2}^{T}\Delta y^2_{it}+2\sum_{L=1}^{\bar{K}}w_{\bar{K}L}\left[\frac{1}{T-1}\sum_{t=2+L}^{T}\Delta y_{it}\Delta y_{it-L}\right]. \tag{5}$$

For Model 2, we replace Δy_{it} in (5) with $\Delta y_{it}-\overline{\Delta y_{it}}$, where $\overline{\Delta y_{it}}$ is the average value of Δy_{it} for individual i. If the data include a time trend (Model 3), then the trend should be removed before estimating the long-run variance. The truncation lag parameter $\bar{K}$ can be data dependent. Andrews (1991) suggests a procedure to determine $\bar{K}$ to ensure the consistency of $\hat{\sigma}^2_{yi}$. The sample covariance weights $w_{\bar{K}L}$ depend on the choice of kernel. For example, if the Bartlett kernel is used,

$$w_{\bar{K}L}=1-\frac{L}{\bar{K}+1}.$$

Now for each individual i, we define the ratio of the long-run standard deviation to the innovation standard deviation,

$$s_i = \sigma_{yi}/\sigma_{\varepsilon i}. \tag{6}$$

Denote its estimate by $\hat{s}_i = \hat{\sigma}_{yi}/\hat{\sigma}_{\varepsilon i}$. Let the average standard deviation ratio be $S_N = (1/N)\sum_{i=1}^{N} s_i$, and its estimator $\hat{S}_N = (1/N)\sum_{i=1}^{N} \hat{s}_i$. This important statistic will be used to adjust the mean of the t-statistic later in step 3, see Eq. (12).

Before presenting step 3, we make two remarks. First, as one referee pointed out, a natural estimate of σ_{yi}^2 under the null hypothesis is $\hat{\sigma}_{\varepsilon i}^2/(1 - \sum_{i=1}^{p_i} \hat{\theta}_{iL})^2$, and since $\hat{\sigma}_{\varepsilon i}^2$ is a consistent estimator for $\hat{\sigma}_{\varepsilon i}^2$ under the null hypothesis, $\hat{s}_i$ may well be estimated by $|1 - \sum_{i=1}^{p_i} \hat{\theta}_{iL}|$. Secondly, it is important to emphasize that both the size and the power properties of the panel unit root test are enhanced by using first-differences (or demeaned first-differences) to estimate the long-run variance. Under the null hypothesis of a unit root, Schwert (1989) found that the long-run variance estimate based on first-differences had much smaller bias in finite samples than the long-run variance estimate based on the residuals in level regression, and the same advantage occurs here too.

2.2.3. Step 3: Compute the panel test statistics

Pool all cross sectional and time series observations to estimate

$$\tilde{e}_{it} = \delta \tilde{v}_{it-1} + \tilde{\varepsilon}_{it}, \tag{7}$$

based on a total of $N\tilde{T}$ observations, where $\tilde{T} = T - \bar{p} - 1$ is the average number of observations per individual in the panel, and $\bar{p} \equiv \frac{1}{N}\sum_{i=1}^{N} p_i$ is the average lag order for the individual ADF regressions. The conventional regression t-statistic for testing $\delta = 0$ is given by

$$t_\delta = \frac{\hat{\delta}}{STD(\hat{\delta})}, \tag{8}$$

where

$$\hat{\delta} = \frac{\sum_{i=1}^{N} \sum_{t=2+p_i}^{T} \tilde{v}_{it-1}\tilde{e}_{it}}{\sum_{i=1}^{N} \sum_{t=2+p_i}^{T} \tilde{v}_{it-1}^2}, \tag{9}$$

$$STD(\hat{\delta}) = \hat{\sigma}_{\tilde{\varepsilon}} \left[\sum_{i=1}^{N} \sum_{t=2+p_i}^{T} \tilde{v}_{it-1}^2 \right]^{-1/2}, \tag{10}$$

$$\hat{\sigma}_{\tilde{\varepsilon}}^2 = \left[\frac{1}{N\tilde{T}} \sum_{i=1}^{N} \sum_{t=2+p_i}^{T} (\tilde{e}_{it} - \hat{\delta}\tilde{v}_{it-1})^2 \right]. \tag{11}$$

Under the hypothesis H_0: $\delta = 0$, the asymptotic results in the next section indicates that the regression t-statistic (t_δ) has a standard normal limiting distribution in Model 1, but diverges to negative infinity for Models 2 and 3. Nevertheless, it is easy to calculate the following adjusted t-statistic:

$$t_\delta^* = \frac{t_\delta - N\tilde{T}\hat{S}_N \hat{\sigma}_{\tilde{\varepsilon}}^{-2} STD(\hat{\delta})\mu_{m\tilde{T}}^*}{\sigma_{m\tilde{T}}^*}, \tag{12}$$

where the mean adjustment $\mu_{m\tilde{T}}^*$ and standard deviation adjustment $\mu_{m\tilde{T}}^*$ can be found in Table 2 for a given deterministic specification ($m = 1, 2, 3$) and time series dimension $\tilde{T}$. (Table 2 also includes quick-and-dirty choices of lag truncation parameter $\bar{K}$ for each time series dimension $\tilde{T}$.) We show in Section 3 that this adjusted t-statistic t_δ^* obeys the standard normal distribution, asymptotically.

3. Asymptotic properties

Define the following sample statistics for each individual:

$$\xi_{1iT} = \frac{1}{\sigma_{\varepsilon i}^2 (T - p_i - 1)} \sum_{t=p_i+2}^{T} \hat{v}_{i,t-1}\hat{e}_{it}, \tag{13}$$

$$\xi_{2iT} = \frac{1}{\sigma_{\varepsilon i}^2 (T - p_i - 1)^2} \sum_{t=p_i+2}^{T} \hat{v}_{i,t-1}^2, \tag{14}$$

$$\xi_{3iT} = \frac{1}{\sigma_{\varepsilon i}^2 (T - p_i - 1)} \sum_{t=p_i+2}^{T} \hat{e}_{it}^2. \tag{15}$$

Next, define the following two ratios for each individual:

$$\gamma_{1iT} = \frac{(T - p_i - 1)\sigma_{\varepsilon i}^2}{\tilde{T}\hat{\sigma}_{\varepsilon i}^2}, \tag{16}$$

$$\gamma_{2iT} = \frac{(T - p_i - 1)^2 \sigma_{\varepsilon i}^2}{\tilde{T}^2 \hat{\sigma}_{\varepsilon i}^2}. \tag{17}$$

Given the above definitions, the statistics of interest in (9), (11) and (8) can be rewritten, respectively, as

$$\hat{\delta} = \frac{N^{-1}\sum_{i=1}^{N} \gamma_{1iT}\xi_{1iT}}{\tilde{T}N^{-1}\sum_{i=1}^{N} \gamma_{2iT}\xi_{2iT}}, \tag{18}$$

$$\hat{\sigma}_{\tilde{\varepsilon}}^2 = \frac{1}{N}\left[\sum_{i=1}^{N} \gamma_{1iT}\xi_{3iT} - 2\hat{\delta}\sum_{i=1}^{N} \gamma_{1iT}\xi_{1iT} + \tilde{T}\hat{\delta}^2\sum_{i=1}^{N} \gamma_{2iT}\xi_{2iT}\right], \tag{19}$$

$$t_\delta = \frac{N^{-1/2}\sum_{i=1}^{N} \gamma_{1iT}\xi_{1iT}}{\hat{\sigma}_{\tilde{\varepsilon}}[N^{-1}\sum_{i=1}^{N} \gamma_{2iT}\xi_{2iT}]^{1/2}}. \tag{20}$$

In contrast to Harris and Tzavalis (1996) and Im et al. (1995), our unit root test is based on the t_δ statistic, obtained from a pooled regression (7). Harris and Tzavalis (1996) work with pooled least squares estimator $\hat{\delta}$ directly. Moreover, the asymptotic property of $\hat{\delta}$ is investigated rather differently in that the time series dimension is fixed while the cross sectional dimension increases to infinity. Im et al. (1995) also base their unit root test on the t-statistics, but it is the average of the t-statistics from individual ADF regressions rather than the pooled t-statistics considered here.

It is easy to determine the limiting behaviors of ξ_{3iT}, γ_{1iT} and γ_{2iT}. Under the null, $\Delta y_{it} = \zeta_{it}$, is an invertible ARMA process. As $T \to \infty$, and as long as the ADF lag order p_i in (1) increases at an appropriate rate T^α, $0 < \alpha \leqslant \frac{1}{4}$, $\hat{e}_{it}$ should well approximate ε_{it}. Hence $(T - p_i - 1)^{-1}\sum \hat{e}_{it}^2 \to \sigma_{\varepsilon i}^2$ in probability. It follows that ξ_{3iT}, γ_{1iT} and γ_{2iT} all converge to one in probability. For easier reference in the sequel, we state the following theorem without proof.

Theorem 1. Given assumption 1 *and* $p_i = 0\ (T^\alpha)$, $0 < \alpha \leqslant \frac{1}{4}$, *for all i. Under* $\mathrm{H}_0, \xi_{3iT} \to 1$, $\gamma_{1iT} \to 1$, *and* $\gamma_{2iT} \to 1$ *in probability, as* $T \to \infty$.

Weak convergence of ξ_{1iT} and ξ_{2iT} are well documented in the literature (e.g. Said and Dickey, 1984; Phillips and Perron, 1988). Results are summarized in the following theorem.

Theorem 2. Given conditions in Theorem 1, $\xi_{1iT} \to s_i W_{m1i}$; $\xi_{2iT} \to s_i^2 W_{m2i}$, *where* $m = 1,2,3$, *is the model index, and* $W_{m1i} = \int_0^1 U_{mi}(r)\,\mathrm{d}B_i(r)$, $W_{m2i} = \int_0^1 U_{mi}^2(r)\,\mathrm{d}B_i(r)$, $B_i(r)$ *is independent Brownian motion*; $U_{1i} = B_i$; U_{2i} *and* U_{3i} *are demeaned and detrended Brownian motions, respectively.*

Note that distributions of W_{m1i} and W_{m2i} depend on the deterministic specification but not on the particular values of model parameters. All moments of W_{m1i} and W_{m2i} exist. Specifically, the mean and variance of W_{m1i} and

Table 1
Asymptotic moments

Model (m)	μ_{m1}	σ^2_{m1}	μ_{m2}	σ^2_{m2}
1. No intercepts or trends	0	1/2	1/2	1/3
2. Individual-specific intercepts	−1/2	1/12	1/6	1/45
3. Individual-specific intercepts & trends	−1/2	1/60	1/15	11/6300

W_{m2i} can be computed following tedious yet straightforward algebra, and are displayed in Table 1.

For each of the three deterministic specifications ($m=1,2,3$), the sample statistics ξ_{1iT} and ξ_{2iT} are independent across individuals i, for each time series dimension T. Those interesting sample statistics defined in (18)–(20) are of the form $N^{-1}\sum_{i=1}^{n}\gamma_{hit}\xi_{hiT}$, $h=1,2$, which is asymptotically equivalent to $N^{-1}\sum_{i=1}^{n}\xi_{hiT}$ due to Theorem 1. It is natural to consider applying a LLN to each panel average $N^{-1}\sum_{i=1}^{n}\xi_{hiT}$ and a CLT to the normalized sum $N^{-1/2}\sum_{i=1}^{n}\xi_{hiT}$ and then use these results to derive the limiting distributions of the panel unit root test statistics, t^*_δ and t_δ.

To gain more intuitions, consider the sequence limit of $\hat{q}_\delta = N^{-1/2}\sum_{i=1}^{n}\xi_{1iT}$ and $\hat{H}_{\delta\delta}=N^{-1}\sum_{i=1}^{n}\xi_{2iT}$. Under model 1, $\hat{H}_{\delta\delta}^{-1/2}\hat{q}_\delta$ is a Studentized LM test for the unit root, which is asymptotically equivalent to the t_δ statistic. Suppose that we first let $T\to\infty$. From Theorem 2, we have $\hat{q}_\delta \to N^{-1/2}\sum_{i=1}^{N}s_iW_{11i}$. Since W_{11i} is a zero mean, independent sequence across individuals with $\mathrm{E}(W_{11i})=0.5$, we can expect to derive the asymptotic normality provided the Lindeberg condition holds. Hence, $\hat{q}_\delta = \mathrm{N}(0,0.5V)$, as $N\to\infty$, where $V=\lim_{N\to\infty}N^{-1}\sum_{i=1}^{N}s_i^2$. A LLN is also expected to work for $\hat{H}_{\delta\delta}$. Since $\mathrm{E}(W_{12i})=0.5$, we have $\hat{H}_{\delta\delta}\to 0.5V$. Combining these results, we deduce that $\hat{H}_{\delta\delta}^{-1/2}\hat{q}_\delta =^d t_\delta \Rightarrow \mathrm{N}(0,1)$. There are some complications when applying the above arguments to model 2, however. This is owing to the fact that $\mathrm{E}(W_{21i})=-\frac{1}{2}$ and $\mathrm{E}(W_{22i})=\frac{1}{6}$. Hence t_δ does not converge to a standard normal distribution. Similar problems arise in Model 3 too. As a result, some sort of adjustment is required. For this, we define $\mathrm{E}(\xi_{1iT})\equiv\mu_{1iT}$ and its consistent estimator as

$$\hat{\mu}_{1iT}=\frac{\hat{s}_i}{\gamma_{1iT}}\mu^*_{m\tilde{T}}. \tag{21}$$

Then (12) can be rewritten as

$$t^*_\delta=\frac{N^{-1/2}\sum_{i=1}^{N}\gamma_{1iT}(\xi_{1iT}-\hat{\mu}_{1iT})}{\sigma^*_{m\tilde{T}}\hat{\sigma}_{\tilde{\varepsilon}}[N^{-1}\sum_{i=1}^{N}\gamma_{2iT}\xi_{2iT}]^{1/2}}. \tag{22}$$

As proved later in Theorem 5, this adjusted t-statistics, t^*_δ converges to a N(0,1). The mean and standard deviation adjustment, $\mu^*_{m\tilde{T}}$ and $\sigma^*_{m\tilde{T}}$, to be used in empirical work are determined from Monte Carlo simulations, as discussed in Section 4 below. Formal definitions are given as follows.

Definition. Let y_t^* be a standard Gaussian random walk, i.e. $y_t^* = y_{t-1}^* + \varepsilon_t^*$, where $\varepsilon_t^* \sim$ i.i.d. N(0,1). Following the test procedure described in Section 2 for a given deterministic specification ($m = 1,2,3$), let $\tilde{T}$ observation on y_t^* be used to generate ξ_{m1T}^* as defined in (13). Then define $\mu_{m\tilde{T}}^* \equiv \mathrm{E}(\xi_{m1T}^*)$, and $\sigma_{m\tilde{T}}^*$ as an arbitrary positive sequence which converges σ_{m1}/μ_{m2} as $\tilde{T} \to \infty$.

By the construction of $\mathrm{E}(\xi_{m1T}^*)$, the mean adjustment $\mu_{m\tilde{T}}^*$ converges to μ_{m1} as $\tilde{T} \to \infty$. We did not specify particular values of $\sigma_{m\tilde{T}}^*$ for finite $\tilde{T} \to \infty$, because only the limit of the sequence is relevant for the asymptotic analysis.

While the foregoing discussions help to speculate the asymptotic results, they do not handle the case in which both T and N approach infinity jointly. Recently, there is growing research interest in studying the asymptotics of nonstationary panel with both N and T infinity simultaneously, e.g. Quah (1990) and Im et al. (1995), among others. Technically, when N and T goes to infinity, asymptotic analysis should be based on the LLN and CLT for triangular arrays. Theorems 3, 4 and 5 below are proved along this line. By triangular arrays we mean that the cross sectional dimension N is an arbitrary monotonically increasing function of T. We shall use the notation N_T instead of N to emphasize the fact that N_T increases with respect to T.

Theorem 3. Suppose that
(a) $\Delta y_{it} = \zeta_{it}$,
(b) *assumption* 1(b) *and* (c) *hold,*
(c) $P_i = 0(T^{\alpha})$, $0 < \alpha \leqslant \frac{1}{4}$, *for all* i,
(d) $\lim_{N_T \to \infty} N_T^{-1} \sum_{i=1}^{N_T} s_i^2 = V$ *exists. Then,* $N_T^{-1/2} \sum_{i=1}^{N_T} \xi_{1iT} =^d N_T^{-1/2} \sum_{i=1}^{N_T} \gamma_{1iT} \xi_{1iT} \Rightarrow \mathrm{N}(0, 0.5V)$, *as* $\sqrt{N_T}/T \to 0$. *Suppose further that*
(e) $\lim_{N_T \to \infty} N_T^{-1} \sum_{i=1}^{N_T} s_i^4$ *exists. Then* $N_T^{-1} \sum_{i=1}^{N_T} \gamma_{2iT} \xi_{2iT} \to^p 0.5V$, *and* $t_\delta \Rightarrow \mathrm{N}(0,1)$.

Theorem 3 states that the asymptotic normality holds as long as $\sqrt{N_T}/T \to 0$. This is particularly relevant for a typical microeconomic panel date set because the time series dimension T is allowed to grow slower than the cross sectional dimension N_T. Other divergence speeds such as $N_T/T \to 0$, and $N_T/T \to$ constant, are sufficient, but obviously not necessary. The panel with no individual specific effects studied by Quah (1990) falls within the territory of Theorem 3. Our result that $\sqrt{N_T}/T \to 0$ is an improvement over Quah's condition that $N_T/T \to$ constant.

When individual-specific effects are introduced in Model 2 and 3, we require faster growth rate in T to achieve the asymptotic normality, as the following theorem shows.

Theorem 4. Suppose that
(a) y_{it} *generated from Model* 2 *or* 3,
(b) *condition* (b)–(d) *in Theorem* 3 *hold, under* $\delta = 0$, $N_T^{-1/2} \sum_{i=1}^{N_T} \gamma_{1iT}(\xi_{1iT} - s_i\mu_{m1}) \Rightarrow \mathrm{N}(0, \sigma_{m1}^2 V)$, *as* $\sqrt{N_T}/T \to 0$. *In addition, suppose*
(c) *the truncation lag parameter* $\bar{K}$ *increases at some rate* T^β, *where* $\beta \in (0,1)$, *then* $N_T^{-1/2} \sum_{i=1}^{N_T} \gamma_{1iT}(\xi_{1iT} - \hat{\mu}_{1iT}) \Rightarrow \mathrm{N}(0, \sigma_{m1}^2 V)$, *as* $N_T/T \to 0$.

It is useful compare Theorem 4 with Theorem 3. From the first part of Theorem 4, $N_T^{-1/2} \sum_{i=1}^{N_T} \gamma_{1iT}(\xi_{1iT} - s_i\mu_{m1}) \Rightarrow \mathrm{N}(0, \sigma_{m1}^2 V)$, if there is no individual-specific effects as in Theorem 3, then $\mu_{11} = 0$, and $N_T^{-1/2} \sum_{i=1}^{N_T} \gamma_{1iT} \xi_{1iT} \Rightarrow \mathrm{N}(0, \sigma_{m1}^2 V)$, as $\sqrt{N_T}/T \to 0$. On the other hand, since $\mu_{m1} \neq 0$ under Models 2 and 3, mean adjustment is required and we want to make sure that such a demeaning method does not influence the asymptotic distribution. The sufficient condition is that $N_T^{-1/2} \sum_{i=1}^{N_T} \gamma_{1iT}(s_i\mu_{m1} - \hat{\mu}_{1iT}) \Rightarrow 0$. But $\hat{\mu}_{1iT} = s_i\mu_{m1} + \mathrm{O_p}(T^{-1/2})$, thus the second term of (A.4.3) in Appendix A will not converge to zero unless $N_T/T \to 0$. In contrast to the case of no individual specific effect (as in Theorem 3), we now need faster growth rate in time series dimension in order to establish the asymptotic normality for the adjusted t-statistic.

Some special cases are of interest. Assume that individual regression error ε_{it} is i.i.d. for each i over T, but heteroskedasticity is present across individuals, i.e. $\sigma_{\varepsilon i}^2 \neq \sigma_{\varepsilon j}^2$. Then we still require $N_T/T \to 0$ to establish the asymptotic normality. In this case, even $s_i = 1$, for all i, the convergence speed of $\hat{\mu}_{1iT} = \gamma_{1iT}^{-1}\mu_{m\tilde{T}}^*$ continues to be dominated by $\gamma_{1iT} = 1 + \mathrm{O_p}(T^{-1/2})$, and hence $\hat{\mu}_{1iT} = \mu_{m1} + \mathrm{O_p}(T^{-1/2})$. The reason is clearly due to the heteroskedastic error across individuals, and that $\hat{\sigma}_{\varepsilon i} = \sigma_{\varepsilon i} + \mathrm{O_p}(T^{-1/2})$. If we further assume away the heteroskedasticity such that $\sigma_{\varepsilon i}^2 = \sigma_{\varepsilon j}^2$, then we have the advantage to pool all observations to estimate the common variance so that $\hat{\sigma}_\varepsilon = \sigma_\varepsilon + \mathrm{O_p}(T^{-1/2}N_T^{-1/2})$. Thus, the mean adjustment term will vanish under the condition that $\sqrt{N_T}/T \to 0$.

The foregoing discussions suggest that the relative divergence speed of N_T and T is largely determined by whether there are individual-specific effects in the panel or not. When there is no individual-specific effect, we can establish asymptotic normality under quite general error process with slower growth rate in T. On the other hand, the inclusion of individual-specific effects requires faster growth rate in T even the individual regression error is a very favorable i.i.d. process.

Theorem 5. In addition to the conditions in Theorem 4, *if* $\lim_{N_T\to\infty} N_T^{-1} \sum_{i=1}^{N_T} s_i^4$ *exist, then* $N_T^{-1} \sum_{i=1}^{N_T} \gamma_{2iT}\xi_{2iT} \to^p \mu_{m2}V$, *and* $t_\delta^* \to \mathrm{N}(0,1)$.

The influence of extending the time dimension T compared with extending the cross-section N can be further clarified by examining the asymptotic distribution of $\hat{\delta}$.

$$\hat{\delta} \approx \mathrm{N}\left(\frac{1}{\tilde{T}}\frac{\mu_{m1}S}{\mu_{m2}V}, \frac{1}{N\tilde{T}^2}\frac{\sigma_{m1}^2}{\mu_{m2}^2 V}\right). \tag{23}$$

First, the Eq. (23) indicates that the variance of $\hat{\delta}$ delta falls at rate $O(N^{-1}T^{-2})$. The higher convergence rate with respect to T is referred to as “super-consistency” in the time series literature. Intuitively, the presence of a unit root causes the variation of the “signal” y_{it-1} to grow arbitrarily large while the variation of the noise ζ_{it} remains constant. Thus, an increment in the time series dimension T adds more variation to the sample than an increment in the cross-section dimension N.

Second, the inclusion of individual-specific intercepts and time trends (Models 2 and 3) causes a downward bias in $\hat{\delta}$ (i.e., $\mu_{m1} < 0$ for $m = 2, 3$). The downward bias vanishes as the time series dimension T grows large (so that $\hat{\delta}$ converges to 0 asymptotically), but is not influenced by the cross-section dimension N. Thus, in contrast to the case of stationary panel data, the presence of a unit root causes the regression specification to influence the asymptotic distribution of the panel estimate.

Third, since an increment in the cross-section dimension N reduces the variance of $\hat{\delta}$ but not its downward bias, the regression t-statistic diverges to negative infinity for Models 2 and 3. In other words, adding an individual to the panel reduces the variance of $\hat{\delta}$ around a non-zero mean, so that the regression t-statistic increasingly rejects the null hypothesis at conventional significance levels. Consequently, it is necessary to use the estimated long run versus short run standard deviation ratios $\hat{s}_i$ and the adjustment factor μ_{mT}^* to demean the individual sample statistics. As the panel grows large, $N/T \to 0$ ensures that the sum of individual differences between the estimated mean $\hat{\mu}_{1iT}$ and the true mean μ_{1iT} does not influence the limiting distribution.

The suggested test procedure requires that the data is generated independently across individuals. As in stationary panel data models, this assumption can be somewhat relaxed to allow for a limited degree of dependence via time-specific aggregate effects by subtracting the cross sectional averages $\bar{y}_t = \frac{1}{N}\sum_{i=1}^{n} y_{it}$ from the observed date. (cf. Hsiao, 1986). The removal of cross-section averages from the data is equivalent to including time-specific intercepts in the regression models above, which does not affect the limiting distributions of the panel unit root test (see Levin and Lin, 1992 for further details). It is appropriate under the assumption of a single aggregate common factor which has an identical impact on all individuals in the panel, however, Quah and Sargent (1992) have recently proposed a method of identifying multiple common factors in panel data; their method may be

Table 2
Mean and standard deviation adjustments[a]

$\tilde{T}$	$\bar{K}$	$\mu^*_{1\tilde{T}}$	$\sigma^*_{1\tilde{T}}$	$\mu^*_{2\tilde{T}}$	$\sigma^*_{2\tilde{T}}$	$\mu^*_{3\tilde{T}}$	$\sigma^*_{3\tilde{T}}$
25	9	0.004	1.049	−0.554	0.919	−0.703	1.003
30	10	0.003	1.035	−0.546	0.889	−0.674	0.949
35	11	0.002	1.027	−0.541	0.867	−0.653	0.906
40	11	0.002	1.021	−0.537	0.850	−0.637	0.871
45	11	0.001	1.017	−0.533	0.837	−0.624	0.842
50	12	0.001	1.014	−0.531	0.826	−0.614	0.818
60	13	0.001	1.011	−0.527	0.810	−0.598	0.780
70	13	0.000	1.008	−0.524	0.798	−0.587	0.751
80	14	0.000	1.007	−0.521	0.789	−0.578	0.728
90	14	0.000	1.006	−0.520	0.782	−0.571	0.710
100	15	0.000	1.005	−0.518	0.776	−0.566	0.695
250	20	0.000	1.001	−0.509	0.742	−0.533	0.603
∞	—	0.000	1.000	−0.500	0.707	−0.500	0.500

[a] *Notes*: The adjustment factors $\mu^*_{1\tilde{T}}$ and $\sigma^*_{1\tilde{T}}$ are used to adjust the mean and standard deviation of the panel unit root test statistic defined in Eq. (12) for a given regression model ($m=1,2,3$), average time series dimension $\tilde{T}$, and lag truncation parameter $\bar{K}$. In all cases, the standard error of the estimated mean adjustment is less than 0.007, and the standard error of the estimated standard deviation adjustment is less than 0.011.

useful for removing more complex contemporaneous correlation from the data.

4. Monte Carlo simulations

In this section, we briefly discuss the results of three Monte Carlo experiments. The first set of simulations were used to determine appropriate values of the mean and standard deviation adjustments, μ^*_{mT} and σ^*_{mT}, used in the adjusted t-statistic given in Eq. (12) for a particular deterministic specification ($m=1,2,3$) and time series dimension $\tilde{T}$. The lag truncation parameter $\bar{K}$ was selected according to the formula $\bar{K}=3.21T^{1/3}$. Due to computational limitations, the ADF lag length p_i was set to the true value of zero rather than being individually estimated. At each replication, Gaussian random numbers with unit variance were used to generate 250 independent random walks of length $\tilde{T}+1$ (i.e. a panel of dimensions $N=250$ and $T=\tilde{T}+1$), and this data was used to construct the sample statistics given in Eqs. (7)–(11) above. Based on 25,000 replications, the adjustment factor $\mu^*_{m\tilde{T}}$ was estimated by the mean value of $t_\delta/N\tilde{T}\hat{S}_N\hat{\sigma}_{\varepsilon}^{-2}STD(\hat{\delta})$ and the adjustment factor $\sigma^*_{m\tilde{T}}$ was estimated by the standard deviation of $t_\delta - N\tilde{T}\hat{S}_N\hat{\sigma}_{\varepsilon}^{-2}STD(\hat{\delta})$. The resulting estimates of $\mu^*_{m\tilde{T}}$ and $\sigma^*_{m\tilde{T}}$ are given in Table 2. In all cases, the standard error of the estimated mean adjustment is less than 0.007, and the standard error of the estimated standard deviation adjustment is less than 0.011. As $\tilde{T}$ grows large, Table 2 indicates that the adjustment factors converge to the asymptotic values predicted by Table 1 and Theorems 3–5.

The second set of simulations examines the empirical size properties of the panel unit root test procedure under alternative deterministic specifications and panel dimensions. For a given replication, Gaussian random numbers with unit variance were used to construct N independent random walks with T time periods each; then the data was used to calculate the adjusted t-statistic using the simulated values of $\mu^*_{m\tilde{T}}$, and $\sigma^*_{m\tilde{T}}$, and the test was performed using the critical values of the standard normal distribution at nominal sizes of 1, 5 and 10 percent. Based on 10,000 Monte Carlo replications, the empirical size for each model specification and set of panel dimensions are given in Table 3. The standard error of the estimated empirical size is less than 0.003 in all cases. Table 3 indicates that in panels with a moderate number of individuals, the nominal size slightly underestimates the empirical size. For larger panels, the difference between nominal and empirical size falls within the range of the Monte Carlo sampling error, consistent with the prediction of asymptotic normality given in Theorems 3–5.

Finally, the empirical power properties of the panel unit root test procedure were analyzed. For a given replication, the data was generated by the process $y_{it} = 0.9 y_{it-1} + \varepsilon_{it}$, (i.e. $\delta = -0.10$) where the disturbances ε_{it} are i.i.d. Gaussian random numbers with unit variance. This data was used to calculate the adjusted t-statistic as in the previous Monte Carlo experiment, and the test was performed using the critical values of the standard normal distribution at nominal sizes of 1, 5 and 10 percent. Based on 10,000 Monte Carlo replications, the empirical power for each model specification and set of panel dimensions are given in Table 4. For purposes of comparison, Table 4 also includes the empirical power of the Dickey–Fuller test for a single time series ($N = 1$). The standard error of the estimated empirical power is less than 0.01 in all cases.

The middle panel of Table 4 illustrates the power advantages of performing unit root tests with panel data. In the absence of individual-specific effects ($m = 1$), the power of the standard Dickey–Fuller test is quite low for short time series ($T \leqslant 50$), whereas very high power can be achieved by performing a joint test for a small number of independent time series ($N \geqslant 10$). If the model allows for individual-specific intercepts and trends, the standard Dickey–Fuller test has very low power even for relatively long time series ($T < 100$), whereas substantial power can be achieved by using the panel unit root test procedure with a panel of moderate dimensions. (i.e. $N = 10$ and $T = 50$, or $N = 25$ and $T = 25$).

Maddla and Wu (1999) have done various simulations to compare the performance of competing tests, including IPS test, LL test and the Fisher's test. Care must be taken to interpret their results. Strictly speaking,

Table 3
Empirical size properties[a]

Model	$\tilde{T}$	$N=10$	$N=25$	$N=50$	$N=100$	$N=250$
$m=1$	25	0.128	0.112	0.107	0.106	0.101
	50	0.131	0.112	0.110	0.105	0.099
	100	0.130	0.114	0.110	0.100	0.099
	250	0.127	0.117	0.109	0.110	0.100
$m=2$	25	0.104	0.101	0.099	0.098	0.098
	50	0.096	0.098	0.098	0.104	0.099
	100	0.100	0.103	0.097	0.106	0.105
	250	0.097	0.100	0.095	0.092	0.097
$m=3$	25	0.102	0.103	0.100	0.100	0.099
	50	0.103	0.104	0.102	0.098	0.102
	100	0.108	0.107	0.100	0.101	0.097
	250	0.105	0.102	0.099	0.105	0.106
$m=1$	25	0.067	0.061	0.054	0.054	0.049
	50	0.071	0.061	0.055	0.053	0.050
	100	0.068	0.059	0.057	0.050	0.050
	250	0.067	0.058	0.056	0.053	0.050
$m=2$	25	0.049	0.050	0.048	0.048	0.047
	50	0.046	0.045	0.049	0.054	0.049
	100	0.050	0.047	0.049	0.052	0.048
	250	0.048	0.048	0.045	0.048	0.050
$m=3$	25	0.051	0.051	0.049	0.051	0.049
	50	0.052	0.049	0.051	0.049	0.054
	100	0.052	0.054	0.053	0.053	0.051
	250	0.049	0.049	0.051	0.052	0.052
$m=1$	25	0.014	0.014	0.010	0.011	0.010
	50	0.015	0.013	0.010	0.010	0.011
	100	0.014	0.012	0.013	0.011	0.011
	250	0.014	0.011	0.012	0.012	0.010
$m=2$	25	0.008	0.009	0.008	0.009	0.008
	50	0.008	0.008	0.009	0.0‘0	0.009
	100	0.009	0.008	0.009	0.009	0.010
	250	0.010	0.010	0.009	0.009	0.009
$m=3$	25	0.011	0.009	0.010	0.011	0.008
	50	0.011	0.010	0.010	0.010	0.010
	100	0.011	0.009	0.009	0.011	0.009
	250	0.009	0.010	0.010	0.011	0.011

[a]*Note*: Top, middle and lower panel correspond to the ten percent, five percent and one percent level test, respectively.

comparisons between the IPS test and LL test are not valid. Though both tests have the same null hypothesis, but the alternatives are quite different. The alternative hypothesis in this article is that all individual series are stationary with identical first order autoregressive coefficient, while the individual first

Table 4
Empirical power properties[a]

Model	$\tilde{T}$	$N=1$	$N=10$	$N=25$	$N=50$	$N=100$	$N=250$
$m=1$	25	0.46	1.00	1.00	1.00	1.00	1.00
	50	0.75	1.00	1.00	1.00	1.00	1.00
	100	0.99	1.00	1.00	1.00	1.00	1.00
	250	1.00	1.00	1.00	1.00	1.00	1.00
$m=2$	25	0.15	0.35	0.59	0.82	0.98	1.00
	50	0.21	0.65	0.93	1.00	1.00	1.00
	100	0.31	0.98	1.00	1.00	1.00	1.00
	250	0.76	1.00	1.00	1.00	1.00	1.00
$m=3$	25	0.14	0.25	0.38	0.53	0.73	0.96
	50	0.23	0.57	0.83	0.96	1.00	1.00
	100	0.37	0.96	1.00	1.00	1.00	1.00
	250	0.82	1.00	1.00	1.00	1.00	1.00
$m=1$	25	0.27	0.99	1.00	1.00	1.00	1.00
	50	0.51	1.00	1.00	1.00	1.00	1.00
	100	0.92	1.00	1.00	1.00	1.00	1.00
	250	1.00	1.00	1.00	1.00	1.00	1.00
$m=2$	25	0.08	0.22	0.43	0.68	0.93	1.00
	50	0.12	0.49	0.86	0.99	1.00	1.00
	100	0.18	0.96	1.00	1.00	1.00	1.00
	250	0.60	1.00	1.00	1.00	1.00	1.00
$m=3$	25	0.08	0.16	0.24	0.37	0.59	0.90
	50	0.13	0.42	0.72	0.93	1.00	1.00
	100	0.24	0.92	1.00	1.00	1.00	1.00
	250	0.71	1.00	1.00	1.00	1.00	1.00
$m=1$	25	0.05	0.91	1.00	1.00	1.00	1.00
	50	0.17	1.00	1.00	1.00	1.00	1.00
	100	0.49	1.00	1.00	1.00	1.00	1.00
	250	1.00	1.00	1.00	1.00	1.00	1.00
$m=2$	25	0.02	0.06	0.16	0.37	0.73	1.00
	50	0.03	0.17	0.58	0.94	1.00	1.00
	100	0.05	0.74	1.00	1.00	1.00	1.00
	250	0.27	1.00	1.00	1.00	1.00	1.00
$m=3$	25	0.02	0.04	0.09	0.16	0.31	0.72
	50	0.04	0.17	0.46	0.80	0.98	1.00
	100	0.08	0.74	0.99	1.00	1.00	1.00
	250	0.43	1.00	1.00	1.00	1.00	1.00

[a]*Note*: Top, middle and lower panel correspond to the ten percent, five percent and one percent level test, respectively.

order autoregressive coefficients in IPS test are allowed to vary under the alternative. If the stationary alternative with identical AR coefficients across individuals is appropriate, pooling would be more advantageous than Im et al. average t-statistics without pooling. Also note that the power simulations reported in Maddla and Wu (1999) are not size-corrected.

5. Conclusion

In this paper, we have developed a procedure utilizing pooled cross-section time series data to test the null hypothesis that each individual time series contains a unit root against the alternative hypothesis that each time series is stationary. As both the cross-section and time series dimensions of the panel grow large, the panel unit root test statistic has a limiting normal distribution. The Monte Carlo simulations indicate that the normal distribution provides a good approximation to the empirical distribution of the test statistic in relatively small samples, and that the panel framework can provide dramatic improvements in power compared to performing a separate unit root test for each individual time series. Thus, the use of panel unit root tests may prove to be particularly useful in analyzing industry-level and cross-country data.

The proposed panel based unit root test does have its limitations. First, there are some cases in which contemporaneous correlation cannot be removed by simply subtracting the cross sectional averages. The research reported in this paper depends crucially upon the independence assumption across individuals, and hence not applicable if cross sectional correlation is present. Secondly, the assumption that all individuals are identical with respect to the presence or absence of a unit root is somewhat restrictive. Readers are referred to Im et al. (1995) for a panel unit root test without the assumption of identical first order correlation under the alternative.

Acknowledgements

This is a revised version of the Levin and Lin's (1992) earlier work. We would like to thank Clive Granger, Rob Engle, Neil Ericsson, Soren Johansen, Cheng Hsiao, Thomas MaCurdy, James McKinnon, Max Stinchcombe, C.-R. Wei, Hal White, and an anonymous referee for helpful comments and suggestions. We appreciate the assistance of Paul Hoffman and the staff of the UCLA supercomputer center.

Appendix A.

Proof of Theorem 3. For each $T, \{\tilde{\xi}_{1iT},\ i=1,2,\ldots,N_T\}$ a zero-mean, independent sequence, where $\tilde{\xi}_{1iT} \equiv \xi_{1iT} - \mu_{1iT}$. Part one of the proof is accomplished in three steps. The following results are used in the subsequent proof.

$$\sup_i \mu_{1iT} = \mathrm{O_p}(T^{-1}) \tag{A.3.1}$$

and

$$\sup_i(\sigma_{iT}^2 - 0.5s_i^2) = \mathrm{O_p}(T^{-1}), \sigma_{iT}^2 \equiv \mathrm{E}(\tilde{\tilde{\xi}}_{1iT}^2) \equiv \sigma_{iT}^2. \tag{A.3.2}$$

In step 1, we prove that $N_T^{-1/2}\tilde{\tilde{\xi}}_{1iT}$ obeys the central limit theorem. For this, we need to ensure the following Lindeberg condition holds (Billingsley, 1986, pp. 368–369):

$$\sum_{i=1}^{N_T}\left(\sum_{i=1}^{N_T}\sigma_{iT}^2\right)^{-1}\int \tilde{\tilde{\xi}}_{1iT}^2 I_{\left\{|\tilde{\tilde{\xi}}_{1iT}|\geqslant\varepsilon\left(\sum_{i=1}^{N_T}\sigma_{iT}^2\right)^{1/2}\right\}}\,\mathrm{d}\mathbb{P}\to 0, \quad as\ T\to\infty.$$

It is well-known that the sufficient condition for the Lindeberg condition is

$$\mathrm{E}|\tilde{\tilde{\xi}}_{1iT}|^{2+\delta} < \infty, \text{ for some } \delta > 0 \text{ and all } i = 1,\ldots,N_T. \tag{A.3.3}$$

Straightforward but tedious algebra leads to result similar to (A.3.1) and (A.3.2), $\sup_i[\mathrm{E}|\tilde{\tilde{\xi}}_{1iT}|^4 - (15/4)s_i^4] = \mathrm{O_p}(T^{-1})$. Hence there exists some T such that (A.3.3) holds for all i, and the Lindeberg condition is fulfilled. We have from the CLT that

$$N_T^{-1/2}\sum_{i=1}^{N_T}\tilde{\tilde{\xi}}_{1iT} = N_T^{-1/2}\sum_{i=1}^{N_T}(\xi_{1iT}-\mu_{1iT}) \sim A\mathrm{N}(0, 0.5V),$$

$$V = \lim_{N_T\to\infty} N_T^{-1}\sum_{i=1}^{N_T}s_i^2. \tag{A.3.4}$$

In step 2, we show that $N_T^{-1/2}\sum_{i=1}^{N_T}\xi_{1iT}$ also obeys the central limit theorem. Using (A.3.1),

$$N_T^{-1/2}\sum_{i=1}^{N_T}(\xi_{1iT}-\mu_{1iT}) = N_T^{-1/2}\sum_{i=1}^{N_T}\xi_{1iT} + N_T^{-1/2}\sum_{i=1}^{N_T}\mu_{1iT}. \tag{A.3.5}$$

From (A.3.1), The second term in (A.3.5) $\leqslant (\sqrt{N_T}/T)(1/N_T)\sum_{i=1}^{N_T} T\mathrm{O_p}(T^{-1}) \leqslant (\sqrt{N_T}/T)M$, since $T\mathrm{O_p}(T^{-1})$ is a bounded sequence for all i. It follows that $N_T^{-1/2}\sum_{i=1}^{N_T}(\xi_{1iT}-\mu_{1iT})$ and $N_T^{-1/2}\sum_{i=1}^{N_T}\xi_{1iT}$ are asymptotic equivalent as $\sqrt{N_T}/T\to 0$.

In the final step, we prove the asymptotic equivalence between $N_T^{-1/2}\sum_{i=1}^{N_T}\xi_{1iT}$ and $N_T^{-1/2}\sum_{i=1}^{N_T}\gamma_{1iT}\xi_{1iT}$. The strategy is to show that the difference between them converges to zero in probability. To do so, we consider the sequence of triangular array, $\{(\gamma_{1iT}-1)\xi_{1iT}\}$, and verify the condition in Theorem 6.2 of Billingsley (1986, p. 81). Define $c_{iT} = \mathrm{E}[(\gamma_{1iT}-1)\xi_{1iT}]$ and $\eta_{iT}^2 = \mathrm{var}[(\gamma_{1iT}-1)\xi_{1iT}]$. The required condition is that there exists a sequence v_T such that $v_T^{-1}\sqrt{\sum_{i=1}^{N_T}\eta_{iT}^2}\to 0$. In fact, putting $v_T = N_T^{1/2}$ works.

Since $\eta_{iT}^2 \leqslant \mathrm{E}[(\gamma_{1iT}-1)^2\xi_{1iT}^2]$,

$$v_T^{-1}\sqrt{\sum_{i=1}^{N_T}\eta_{iT}^2} \leqslant \left\{N_T^{-1}\sum_{i=1}^{N_T}\mathrm{E}[(\gamma_{1iT}-1)^2\xi_{1iT}^2]\right\}^{1/2}$$

$$=\left\{N_T^{-1}T^{-1}\sum_{i=1}^{N_T}\mathrm{E}[T(\gamma_{1iT}-1)^2\xi_{1iT}^2]\right\}^{1/2}$$

$$\leqslant \left\{MN_T^{-1}T^{-1}\sum_{i=1}^{N_T}\mathrm{E}(\xi_{1iT}^2)\right\}^{1/2},$$

because $T(\gamma_{1iT}-1)^2$ is bounded in probability. It follows that

$$\left\{MN_T^{-1}T^{-1}\sum_{i=1}^{N_T}\mathrm{E}(\xi_{1iT}^2)\right\}^{1/2}$$

$$\leqslant \left\{MN_T^{-1}T^{-1}\sum_{i=1}^{N_T}[0.5s_i^2+\mathrm{O_p}(T^{-1})]\right\}^{1/2}$$

$$\leqslant \left\{0.5MN_T^{-1}T^{-1}\sum_{i=1}^{N_T}s_i^2+MN_T^{-1}T^{-2}\sum_{i=1}^{N_T}T\mathrm{O_p}(T^{-1})\right\}^{1/2}.$$

It is seen that the last term converges to zero, hence

$$N_T^{-1/2}\sum_{i=1}^{N_T}[(\gamma_{1iT}-1)\xi_{1iT}]-N_T^{-1/2}\sum_{i=1}^{N_T}c_{iT}\to^{\mathrm{p}} 0.$$

However,

$$N_T^{-1/2}\sum_{i=1}^{N_T}c_{iT}=N_T^{-1/2}\sum_{i=1}^{N_T}\mathrm{E}[(\gamma_{1iT}-1)\xi_{1iT}]$$

$$=\sqrt{\frac{N_T}{T}}\frac{1}{N_T}\sum_{i=1}^{N_T}\mathrm{E}[\sqrt{T}(\gamma_{1iT}-1)\xi_{1iT}]$$

$$\leqslant M\sqrt{\frac{N_T}{T}}\frac{1}{N_T}\sum_{i=1}^{N_T}\mathrm{E}(\xi_{1iT})$$

$$\leqslant M\sqrt{\frac{N_T}{T}}\frac{1}{N_T}\sum_{i=1}^{N_T}\mathrm{O_p}(1/T)$$

$$\leqslant MM'\sqrt{\frac{N_T}{T}}\frac{1}{N_T}\frac{1}{T}\sum_{i=1}^{N_T} T\mathrm{O_p}(1/T)$$

$$\leqslant MM'\sqrt{N_T}/T^{3/2} \to 0.$$

This completes proving the first part of theorem 3.

For the second part of the proof, we verify that the condition of LLN for triangular array holds (Billingsley, 1986, p. 81). Since $\{\xi_{2iT}, i = 1, 2, \ldots\}$ is an independent sequence with mean $\sup_i[\mathrm{E}(\xi_{2iT}) - 0.5s_i^2] = \mathrm{O_p}(T^{-1})$, and $\sup_i[\mathrm{Var}(\xi_{2iT}) - \frac{1}{3}s_i^4] = \mathrm{O_p}(T^{-1})$, if the required condition $[\sum_{i=1}^{N_T} \mathrm{var}(\xi_{2iT})]^{1/2}/v_T \to 0$ holds for some positive sequence $\{v_T\}$, then

$$v_T^{-1}\sum_{i=1}^{N_T}[\xi_{2iT} - \mathrm{E}(\xi_{2iT})] \to^{\mathrm{p}} 0. \tag{A.3.6}$$

Put $v_T = N_T$. Assumption 3(e) ensures that $[\sum_{i=1}^{N_T} \mathrm{var}(\xi_{2iT})]^{1/2}/v_T \to 0$, hence (A.3.6) follows.

But

$$N_T^{-1}\sum_{i=1}^{N_T}\mathrm{E}(\xi_{2iT}) \leqslant N_T^{-1}\sum_{i=1}^{N_T}[0.5s_i^2 + \mathrm{O_p}(T^{-1})]$$

$$= 0.5N_T^{-1}\sum_{i=1}^{N_T} s_i^2 + T^{-1}N_T^{-1}\sum_{i=1}^{N_T} T\mathrm{O_p}(T^{-1}) \to 0.5V.$$

Desired result follows from combining this result with $\gamma_{2iT} \to 1$ in Theorem 1. Finally, since $\hat{\sigma}_{\tilde{\varepsilon}} \to 1, t_\delta \Rightarrow \mathrm{N}(0,1)$ follows immediately. □

Proof of Theorem 4. Since $\{\tilde{\xi}_{1iT}, i = 1, 2, \ldots\}$, $\tilde{\xi}_{1iT} \equiv \xi_{1iT} - \mu_{1iT}$, is a zero mean, independent sequence across individuals such that $\sup_i(\mu_{1iT} - s_i\mu_{m1}) = \mathrm{O_p}(T^{-1})$ and $\sup_i(\sigma_{iT}^2 - s_i^2\sigma_{m1}^2) = \mathrm{O_p}(T^{-1}), \sigma_{iT}^2 \equiv \mathrm{Var}(\tilde{\xi}_{1iT})$, where $m = 2$ or 3, is the model index. Analogous to the proof in Theorem 3, the Lindeberg condition holds, and the CLT implies that

$$\left(\sum_{i=1}^{N_T}\sigma_{iT}^2\right)^{-1/2}\sum_{i=1}^{N_T}\tilde{\xi}_{1iT} = \left(N_T^{-1}\sum_{i=1}^{N_T}\sigma_{iT}^2\right)^{-1/2} N_T^{-1/2}\sum_{i=1}^{N_T}\tilde{\xi}_{1iT} \Rightarrow \mathrm{N}(0,1), \text{ or}$$

$$N_T^{-1/2}\sum_{i=1}^{N_T}\tilde{\xi}_{1iT} \Rightarrow \mathrm{N}(0, \sigma_{m1}^2 V).$$

Now, $N_T^{-1/2}\sum_{i=1}^{N_T}(\xi_{1iT} - \mu_{1iT}) \leqslant N_T^{-1/2}\sum_{i=1}^{N_T}(\xi_{1iT} - s_i\mu_{m1}) + N_T^{-1/2}\sum_{i=1}^{N_T}\mathrm{O_p}(T^{-1})$. Since $N_T^{-1/2}\sum_{i=1}^{N_T}\mathrm{O_p}(T^{-1}) \to 0$, as $\sqrt{N_T}/T \to 0$, we conclude that $N_T^{-1/2}\sum_{i=1}^{N_T}(\xi_{1iT} - s_i\mu_{m1}) \Rightarrow \mathrm{N}(0, \sigma_{m1}^2 V)$. Similar to the proof in Theorem 3, we can also prove that

$$N_T^{-1/2}\sum_{i=1}^{N_T}\gamma_{1iT}(\xi_{1iT}-s_i\mu_{m1})\Rightarrow \mathrm{N}(0,\sigma_{m1}^2 V). \tag{A.4.1}$$

For the second part of Theorem 4, consider replacing μ_{m1} in (A.4.1) with its consistent estimate $\hat{\mu}_{1iT}$. Write

$$N_T^{-1/2}\sum_{i=1}^{N_T}\gamma_{1iT}(\xi_{1iT}-\hat{\mu}_{1iT})=N_T^{-1/2}\sum_{i=1}^{N_T}\gamma_{1iT}(\xi_{1iT}-s_i\mu_{m1}) +N_T^{-1/2}\sum_{i=1}^{N_T}\gamma_{1iT}(s_i\mu_{m1}-\hat{\mu}_{1iT}). \tag{A.4.2}$$

We know that the first term in (A.4.2) converges to a normal distribution. The asymptotic behavior of the second term is of interest. Note that $\hat{s}_i=\hat{\sigma}_{yi}/\hat{\sigma}_{\varepsilon i}$ is a bounded sequence for all i, due to $\hat{\sigma}_{\varepsilon i}=\sigma_{\varepsilon i}+\mathrm{O_p}(T^{-1/2})$ and $\hat{\sigma}_{yi}=\sigma_{yi}+\mathrm{O_p}(T^{-1/2})$, c.f. Andrews (1991). Since $\mu^*_{m\tilde{T}}=\mu_{m1}+\mathrm{O_p}(T^{-1})$ by construction and $\gamma_{1iT}=1+\mathrm{O_p}(T^{-1/2})$, $\hat{\mu}_{1iT}=\hat{s}_i/\gamma_{1iT}\mu^*_{m\tilde{T}}=s_i\mu_{m1}+\mathrm{O_p}(T^{-1/2})$.

Now rewrite the second term in (A.4.2) as

$$\begin{aligned}
&N_T^{-1/2}\sum_{i=1}^{N_T}\gamma_{1iT}(s_i\mu_{m1}-\hat{\mu}_{1iT})\\
&=N_T^{-1/2}\sum_{i=1}^{N_T}(\gamma_{1iT}-1)(s_i\mu_{m1}-\hat{\mu}_{1iT})+N_T^{-1/2}\sum_{i=1}^{N_T}(s_i\mu_{m1}-\hat{\mu}_{1iT})\\
&\leqslant N_T^{-1/2}\sum_{i=1}^{N_T}\mathrm{O_p}(T^{-1/2})\mathrm{O_p}(T^{-1/2})+N_T^{-1/2}\sum_{i=1}^{N_T}\mathrm{O_p}(T^{-1/2}).
\end{aligned}\tag{A.4.3}$$

The first term in (A.4.3) $\to 0$, as $\sqrt{N_T/T}\to 0$. However, the second term requires that $N_T/T\to 0$. Complete the second part of Theorem 4. □

Proof of Theorem 5. $\{\xi_{2iT}, i=1,2,\ldots\}$ is an independent sequence with $\sup_i[\mathrm{E}(\xi_{2iT})-s_i^2\mu_{m2}]=\mathrm{O_p}(T^{-1})$, and $\sup_i[\mathrm{Var}(\xi_{2iT})-s_i^4\sigma_{m2}^2]=\mathrm{O_p}(T^{-1})$. Applying the LLN as in the proof of Theorem 3, $N_T^{-1}\sum_{i=1}^{N_T}\xi_{2iT}-N_T^{-1}\sum_{i=1}^{N_T}\mathrm{E}(\xi_{2iT})\to^{\mathrm{p}} 0$. But $N_T^{-1}\sum_{i=1}^{N_T}\mathrm{E}(\xi_{2iT})\leqslant N_T^{-1}\sum_{i=1}^{N_T}[s_i^2\mu_{m2}+\mathrm{O_p}(T^{-1})]=\mu_{m2}N_T^{-1}\sum_{i=1}^{N_T}s_i^2+T^{-1}N_T^{-1}\sum_{i=1}^{N_T}T\mathrm{O_p}(T^{-1})\to\mu_{m2}V$.

The standard deviation adjustment $\sigma^*_{m\tilde{T}} \to \sigma_{m1}/\sqrt{\mu_{m2}}$ by construction, and $\hat{\sigma}_{\tilde{\varepsilon}} \to 1$, we have $t^*_\delta \Rightarrow \mathrm{N}(0,1)$. □

References

Anderson, T.W., Hsiao, C., 1982. Formulation and estimation of dynamic models using panel data. Journal of Econometrics 18, 47–82.

Andrews, D.W.K., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 59, 817–858.

Billingsley, P., 1986. Probability and Measure, 2nd Edition. Wiley Press, New York.

Box, G.E.P., Jenkins, G.M., 1970. Time Series Analysis: Forecasting and Control. Holden Day Press, San Francisco, CA.

Breitung, J., Meyer, W., 1991. Testing for unit roots in panel data: are wages on different bargaining levels cointegrated? Institut fur Quantitative Wirtschaftsforschung, Working paper.

Campbell, J., Perron, P., 1991. Pitfalls and opportunities: what macroeconomists should know about unit roots. In: Blanchard, O., Fischer, S. (Eds.), NBER Macroeconomics Annual. MIT Press, Cambridge, MA.

Dickey, D.A., Fuller, W.A., 1979. Distribution of the estimators for autoregressive time series with a unit root. Journal of American Statistical Association 74, 427–431.

Dickey, D.A., Fuller, W.A., 1981. Likelihood ratio statistics for autoregressive time series with a unit root. Econometrica 49, 1057–1072.

Diebold, F.X., Nerlove, M., 1990. Unit roots in economic time series: a selective survey. Advances in Econometrics 8, 3–69.

Fuller, W., 1976. Introduction to Statistical Time Series. John Wiley.

Hall, A., 1990. Testing for a Unit Root in Time Series with Pretest Data-based Model Selection. North Carolina State University, Working Paper.

Harris, R., Tzavalis, E., 1996. Inference for unit root in dynamic panels. Unpublished manuscript.

Holtz-Eakin, D., Newey, W.K., Rosen, H., 1988. Estimating vector autoregressions with panel data. Econometrica 56, 1371–1395.

Hsiao, C., 1986. Analysis of panel data, Econometric Society Monograph #11.

Im, K.S., Pesaran, H., Shin, Y., 1995, Testing for unit roots in heterogenous panels. Unpublished manuscript.

Johansen, S., 1991. Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregression models. Econometrica 59, 1551–1580.

Johansen, S., 1992. Determination of cointegration rank in the presence of a linear trend. Oxford Bulletin of Economics and Statistics 54, 383–397.

Levin, A., Lin, C.-F., May 1992. Unit root tests in panel data: asymptotic and finite-sample properties. UC San Diego, Working Paper 92-23.

MaCurdy, T.E., 1982. The use of time series processes to model the error structure of earnings in a longitudinal data analysis. Journal of Econometrics 18, 83–114.

Maddla, G.S., Wu, S., 1999. Comparative study of unit root tests with panel data and a new simple test. Oxford Bulletin of Economics and Statistics 61, 631–652.

Park, J.Y., 1990. Testing for unit roots and cointegration by variable addition. Advances in Econometrics 8, 107–133.

Phillips, P.C.B., 1987. Time series regression with a unit root. Econometrica 55, 277–301.

Phillips, P.C.B., Moon, H.R., 1999. Linear regression limit theory for nonstationary panel data. Econometrica 67, 1057–1111.

Phillips, P.C.B., Perron, P., 1988. Testing for a unit root in time series regression. Biometrika 75, 335–346.

Quah, D., 1990. International patterns of growth I: persistence in cross-country disparities. MIT working paper.

Quah, D., Sargent, T.J., 1992. A dynamic index model for large cross-country disparities. Institute for Empirical Macroeconomics DP#77.

Quah, D., 1994. Exploiting cross section variation for unit root inference in dynamic data. Economic Letters 44, 9–19.

Said, S.E., Dickey, D.A., 1984. Testing for unit roots in autoregressive-moving average processes of unknown order. Biometrika 71, 599–607.

Schwert, G.W., 1989. Tests for unit roots: a Monte Carlo investigation. Journal of Business and Economic Statistics 7, 147–160.