Subsampling Hypothesis Tests for Nonstationary Panels with Applications to the PPP Hypothesis*

In Choi† and Timothy Chu‡

Abstract

This paper studies subsampling hypothesis tests for panel data that are possibly nonstationary, and cross-sectionally correlated and cross-sectionally cointegrated. The tests include panel unit root and cointegration tests as special cases. The number of cross-sectional units in the panel data is assumed to be finite, and that of time series observations infinite. Cross-sectional correlation is allowed for both regressors and error terms. Two types of subsampling, non-centered and centered, are considered. It is shown that empirical distributions using subsamples uniformly converge to corresponding limiting distributions. For the non-centered subsampling, the result is shown in the mode of almost sure convergence and discontinuous limiting distributions are allowed. For the centered subsampling, the uniform convergence result is obtained in the mode of convergence in probability and only for continuous limiting distributions. Test consistency using the critical values from the empirical distributions is also established. These results are applied to panel unit root and stationarity tests. The panel unit root tests considered are Levin, Lin and Chu (2002)'s t-test, Im, Pesaran and Shin's (2003) averaged t-test and Choi's (2001) Z test. For the null of stationarity, Hadri's (2000) test is used. Block sizes of subsamples are chosen by stochastic calibration. Simulation results show that the subsampling distributions of the panel unit root tests using the stochastic calibration provide reasonably good approximations to the finite sample distributions of the tests.

1 Introduction

Recently there has been much research interest in nonstationary panels. The initial motivation of using panel data was higher power of unit root tests using them. Levin, Lin and Chu's (2002) tests and other subsequent panel unit root tests like Im, Pesaran

*This version is incomplete and additional research results will be added to the paper in due course. The first author acknowledges that the research reported in this paper was financially supported by the RGC Competitive Earmarked Research Grant 2003-2004 under the Project No. HKUST6223/03H.

†Department of Economics, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong. E-mail: inchoi@ust.hk. Home page: http://ihome.ust.hk/~inchoi.

‡Department of Economics, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong. E-mail: timchue@ust.hk.

and Shin (2003) and Choi (2001) confirm power advantages of using panel data. Cointegration tests using panel data have also been proposed by Kao (1999), Pedroni (1995), McCoskey and Kao (1998), Larsson, Lyhagen and Löthgren (2001) and Groen and Kleibergen (2003). Linear regression methods for nonstationary panel data are developed in, among others, Pedroni (2000), Kao and Chiang (2000), Phillips and Moon (1999), Kauppi (2000) and Choi (2002a). More references can be found in survey articles by Baltagi and Kao (2000), Banerjee (1999) and Phillips and Moon (2000).

However, all the aforementioned work except Groen and Kleibergen (2003) assume cross-sectional independence that may be inappropriate for applications. In response to this, researchers have developed various methods that can be applied to cross-sectionallly correlated panels. Bootstrap methods are used for panel unit root tests in Maddala and Wu (1999) and Chang (2003). Taylor and Sarno (1998) study the multivariate augmented Dickey–Fuller test. O'Connell (1998) considers a GLS-based unit root test for homogeneous panels. Error component model allowing cross-sectional correlation is used in Choi (2002b). Dynamic factor modelling is used in Bai and Ng (2002), Moon and Perron (2003), Phillips and Sul (2003) and Pesaran (2003). Last, Groen and Kleibergen (2003) use the vector autoregressive model.

This paper studies subsampling to devise approximations to the finite-sample distributions of hypothesis tests for possibly nonstationary, cross-sectionally correlated and cross-sectionally cointegrated panels. The subsampling method has mainly been developed by Dimitris Politis, Joseph Romano and Michael Wolf, and their major research results are summarized in Politis, Romano and Wolf (1999). Initial applications of the subsampling methods have been to the construction of confidence intervals. But their usefulness for hypothesis testing is also demonstrated in Delgado, Rodríguez-Poo and Wolf (2001), Gonzalo and Wolf (2002) and Choi (2004), among others. The main idea of subsampling is closely related to the generalized jackknife method of Wu (1990). In addition, Sherman and Carlstein (1996) also use subsamples to estimate sampling distributions. But their study is confined to stationary time series.

In the subsampling approach, the statistic of interest is computed at subsamples of the data (consecutive sample points in the case of the time series), and the subsampled values of the statistic are used to estimate its finite sample distribution. It has profitably been used for time series analysis, although its application to unit root nonstationary time series can be found only in Romano and Wolf (2001) and Choi (2004). The major strength of the subsampling method is that it can work even when the bootstrap method fails, as illustrated in Romano and Wolf (1999, 2001) and Choi (2004). In applying the subsampling method to time series, subsamples should use consecutive sample points in order to retain the structure of serial correlation.

The subsampling approach to hypothesis testing for possibly nonstationary panel data is new and have several advantages. First, the regressors in the panel data model may be stationary, or nonstationary with unit roots, or both types mixed. Specifying the order of integration for the regressors is not required. Second, cross-sectional correlation is allowed and parameters related to the cross-sectional correlation need not be estimated. Third, cross-sectional cointegration is also allowed. Knowledge on the cointegration coefficients and ranks are not required. The cross-sectional cointegration may bring difficulties if conventional asymptotic methods are used. Fourth, as will be shown, the subsampling approach is able to provide tests with reasonably sound finite-sample properties.

The tests we consider for the panel data model are general enough to include existing panel unit root tests like Levin, Lin and Chu (2002), Im, Pesaran and Shin (2003) and combination tests in Maddala and Wu (1999) and Choi (2001). The tests also include panel stationarity tests like Hadri (2000) and tests for the nulls of cointegration and non-cointegration. Tests on structural coefficients for panel data models are also examples of our tests.

When computing the panel tests with subsamples, we use the tests of the original form and the tests with the subsample OLS coefficient estimates centered at the full-sample estimates. Subsampling using the former will be called non-centered subsampling; and that using the latter centered subsampling. It is shown that empirical distributions using subsamples uniformly converge to corresponding limiting distributions. For the non-centered subsampling, the result is shown in the mode of almost sure convergence and discontinuous limiting distributions are allowed. For the centered subsampling, the uniform convergence result is obtained in the mode of convergence in probability and only for continuous limiting distributions. Test consistency using the critical values from the empirical distributions is also established.

The general results of subsampling for panel tests are applied to panel unit root and stationarity tests. The panel unit root tests considered are Levin, Lin and Chu (2002)'s t-test, Im, Pesaran and Shin's (2003) averaged t-test and Choi's (2001) inverse normal test. For the null of stationarity, Hadri's (2000) test is used. Block sizes of subsamples are chosen by stochastic calibration. Simulation results show that the subsampling distributions of the panel unit root tests using the stochastic calibration provide reasonably good approximations to the finite sample distributions of the tests.

This paper is organized as follows. Section 2 introduces the model, hypothesis and tests and establishes the asymptotic validity of the tests using the non-centered and centered subsamplings. Section 3 applies the results of Section 2 to panel unit root and stationarity tests. Section 4 reports simulation results for the empirical size and power of the tests studied in Section 3. All the proofs are contained in the Appendix.

2 Subsampling tests for cross-sectionally correlated panels

2.1 The model, hypotheses and tests

Consider the model for panel data

$$y_{it} = f(x_{it}, u_{it}, \theta_i),\ \ (i = 1, ..., N; t = 1, ..., n), \tag{1}$$

where θ_i is a parameter vector related to both x_{it} and u_{it}, x_{it} a random regressor of dimension k, and u_{it} an error term. In model (1), the index i denotes households, individuals, countries, etc., and the index t time. The number of cross-sectional units N is assumed to be fixed.

Let the j-th component of x_{it} be x_{itj} ($j = 1, ..., k$), and the collection of x_{itj} over the cross-sectional units is denoted as $\mathbf{x}_{tj}$ (i.e., $\mathbf{x}_{tj} = [x_{1tj}, ..., x_{Ntj}]'$). We assume $(I - B)^{\lambda_{xj}}\mathbf{x}_{tj}$ is mixing where $\lambda_{xj} = 0$ or 1 and B denote the lag operator. Detailed mixing conditions will be given later. Here $(I - B)^{\lambda_{xj}}$ denotes a transformation that makes $\mathbf{x}_{tj}$ mixing. If $\mathbf{x}_{tj}$ is an integrated process, $\lambda_{xj} = 1$. On the other hand, if $\mathbf{x}_{tj}$ is a mixing process, $\lambda_{xj} = 0$. Under this set-up, regressors are allowed to have mixed orders of integration. But each regressor has the same order of integration across cross-sectional units. For $I(1)$ regressors, cross-sectional cointegration is allowed as long as Assumption 1 below is satisfied. In other words, there may exist N by 1 vectors $c_1, ..., c_r$ ($r < N$) such that $c_k'\mathbf{x}_{tj} = I(0)$ for all $k = 1, ..., r$ when $\mathbf{x}_{tj} = I(1)$.

Letting $\mathbf{u}_t = [u_{1t}, ..., u_{Nt}]'$, cross-sectional dependence for $\mathbf{v}_t = [(I-B)^{\lambda_{x1}}\mathbf{x}_{t1}', ..., (I-B)^{\lambda_{xk}}\mathbf{x}_{tk}', \mathbf{u}_t']'$ is assumed to take the form

$$Cov(\mathbf{v}_t) = [\sigma_{ij}]_{i,j=1,...(k+1)N} \tag{2}$$

where $\sigma_{ij} < \infty$ for all i and j. Time series dependence between x_{it} and u_{it} may be allowed. If some of x_{it} are mixing and correlated with u_{it}, instrumental variables z_{it} of size l are assumed to be available. The number of the instrumental variables is assumed to be larger than that of the mixing regressors correlated with the error term. For the instruments, we use the notation $\mathbf{z}_{tj} = [z_{1tj}, ..., z_{Ntj}]'$ ($j = 1, ..., l$). For $I(1)$ regressors, instruments are not required.

More detailed conditions on x_{it}, z_{it} and u_{it} will be required for Assumption 1 below to be satisfied. These should be given in each application.

We are interested in testing the null hypothesis

$$H_0 : R\alpha_i = r \text{ for all } i, \tag{3}$$

where α_i is a subvector of θ_i.When α_i are heterogeneous across cross-sectional units, tests for the null hypothesis (3) are assumed to take the form

$$\xi_{Nn} = g\left((R\hat{\alpha}_{1n} - r)'\,\Theta_{1n}(R\hat{\alpha}_{1n} - r), ..., (R\hat{\alpha}_{Nn} - r)'\Theta_{Nn}(R\hat{\alpha}_{Nn} - r)\right), \tag{4}$$

where $\hat{\alpha}_{in}$ and Θ_{in} are a consistent estimator of α_i and the corresponding weight matrix, respectively, that use the whole demeaned[1] time series sample for the i-th individual. Using the demeaned data is appropriate, because x_{it} rarely has zero mean for each i.

When α_i are fixed throughout individuals ($\alpha_i = \alpha$ for all i), tests based on the pooled data are assumed to take the form

$$\tau_{Nn} = h(\left(R\hat{\alpha}^p_{Nn} - r)'\Theta^p_{Nn}(R\hat{\alpha}^p_{Nn} - r)\right), \tag{5}$$

[1] In some cases, detrending is also required. The additional detrending does not bring any changes to the arguments of this section. For simplicity, "demeaning" in this section is understood to be "demeaning and detrending if necessary."

where $\hat{\alpha}^p_{Nn}$ and Θ^p_{Nn} are a consistent estimator of α and the corresponding weight matrix, respectively, using the pooled panel data. The data are assumed to be demeaned before being used for τ_{Nn}.

In some applications (e.g. tests for the null of stationarity), null hypotheses cannot be written like (3). Instead, it is more appropriate to put the null hypotheses for some constants a and b as

$$H_0 : a < \theta_{ik} < b \text{ for all } i, \tag{6}$$

where θ_{ik} is the $k-$th element of θ_i. The test for this null hypothesis is assumed to take the form

$$\varphi_{Nn} = k(\zeta_{1n}, ..., \zeta_{Nn}) \tag{7}$$

where ζ_{in} is a test statistic for the null hypothesis $a < \theta_{ik} < b$ using the demeaned data of the $i-$th individual. Using pooled data for the null hypothesis (6) does not seem to be appropriate because parameter θ_{ik} may take different values for each i.

2.2 Subsampling

The tests considered in the previous subsection are likely to have asymptotic distributions involving nuisance parameters due to the cross-sectional correlation (2). The subsampling method can provide approximations to the limiting distributions of the tests by computing the tests using smaller blocks of consecutively observed time series and formulating empirical distribution functions out of the computed values of the tests. Notably, the method does not require estimating the nuisance parameters.

Let

$$\mathbf{w}_t = [\mathbf{x}'_{t1}, ..., \mathbf{x}'_{tk}, \mathbf{z}'_{t1}, ..., \mathbf{z}'_{tl}, \mathbf{u}'_t]'$$

and

$$\delta_{bs} = [\mathbf{w}_s, \mathbf{w}_{s+1}, ..., \mathbf{w}_{s+b-1}](1 \leq s \leq n - b + 1).$$

Here, δ_{bs} denotes a consecutive subsample that starts from s and has sample size b. Parameter b is also called the block size.

2.2.1 Non-centered subsampling

Let $\hat{\alpha}_{1bs}, ..., \hat{\alpha}_{Nbs}, \hat{\alpha}^p_{Nbs}$ be parameter estimators using the demeaned subsample

$$\bar{\delta}_{bs} = [\mathbf{w}_s - \bar{\mathbf{w}}_{bs}, \mathbf{w}_{s+1} - \bar{\mathbf{w}}_{bs}, ..., \mathbf{w}_{s+b-1} - \bar{\mathbf{w}}_{bs}] \ (1 \leq s \leq n - b + 1),$$

where $\bar{\mathbf{w}}_{bs} = \frac{1}{b}\sum_{k=s}^{s+b-1} \mathbf{w}_k$, and let $\Theta_{1bs}, ..., \Theta_{Nbs}, \Theta^p_{Nbs}$ be corresponding weight matrices. Subsample versions of tests (4) and (5) are, respectively,

$$\xi_{Nbs} = g\left((R\hat{\alpha}_{1bs} - r)'\,\Theta_{1bs}(R\hat{\alpha}_{1bs} - r), ..., (R\hat{\alpha}_{Nbs} - r)'\Theta_{Nbs}(R\hat{\alpha}_{Nbs} - r)\right) \tag{8}$$

and

$$\tau_{Nbs} = h\left(\left(R\hat{\alpha}^p_{Nbs} - r\right)'\,\Theta^p_{Nbs}\left(R\hat{\alpha}^p_{Nbs} - r\right)\right). \tag{9}$$

Subsample version of test (7) is

$$\varphi_{Nbs} = k(\zeta_{1bs}, ..., \zeta_{Nbs}). \tag{10}$$

There are $n-b+1$ subsample tests that will be used to estimate the liming distributions of the tests.

Using the subsample tests (8), (9) and (10), empirical distributions

$$L^{\xi}_{Nnb}(x) = \frac{1}{n-b+1}\sum_{s=1}^{n-b+1} 1\{\xi_{Nbs} \leq x\} \tag{11}$$

$$L^{\tau}_{Nnb}(x) = \frac{1}{n-b+1}\sum_{s=1}^{n-b+1} 1\{\tau_{Nbs} \leq x\} \tag{12}$$

and

$$L^{\varphi}_{Nnb}(x) = \frac{1}{n-b+1}\sum_{s=1}^{n-b+1} 1\{\varphi_{Nbs} \leq x\} \tag{13}$$

are formulated.

It will be shown that the empirical distributions (11), (12) and (13) converge to the limiting distributions of the tests (4), (5) and (7), respectively. To this end, we require a few assumptions. Let $J^{\xi}_{Nn}(x) = P\{\xi_{Nn} \leq x\}$, $J^{\tau}_{Nn}(x) = P\{\tau_{Nn} \leq x\}$ and $J^{\varphi}_{Nn}(x) = P\{\varphi_{Nn} \leq x\}$. These are finite sample distribution functions of the tests (4), (5) and (7), respectively. Regarding the finite sample distributions, assume

Assumption 1 *(i) Under the null hypothesis (3), $J^{\xi}_{Nn}(x) \to J^{\xi}_{N}(x)$ and $J^{\tau}_{Nn}(x) \to J^{\tau}_{N}(x)$ for every continuity point x of $J^{\xi}_{N}(\cdot)$ and $J^{\tau}_{N}(\cdot)$, respectively, as $n \to \infty$. Under the null hypothesis (6), $J^{\varphi}_{Nn}(x) \to J^{\varphi}_{N}(x)$ for every continuity point x of $J^{\varphi}_{N}(\cdot)$ as $n \to \infty$.*

(ii) Under the null hypothesis (3), $J^{\xi}_{Nn}(x\pm) \to J^{\xi}_{N}(x\pm)$ and $J^{\tau}_{Nn}(x\pm) \to J^{\tau}_{N}(x\pm)$ for every discontinuity point x of $J^{\xi}_{N}(\cdot)$ and $J^{\tau}_{N}(\cdot)$, respectively, as $n \to \infty$. Under the null hypothesis (6), $J^{\varphi}_{Nn}(x\pm) \to J^{\varphi}_{N}(x\pm)$ for every discontinuity point x of $J^{\varphi}_{N}(\cdot)$ as $n \to \infty$.

Part (i) of this assumption requires that the tests have limiting distributions, and will be satisfied in most cases. Part (ii) introduces a condition for discontinuous limiting distributions. Under this assumption, it follows from Lemma 3 of Chow and Teicher (1988, p.265) that

$$\sup_x |J^{a}_{Nn}(x) - J^{a}_{N}(x)| \to 0 \ (a = \xi, \tau \text{ or } \varphi) \text{ as } n \to \infty.$$

The following assumption assumes that the subsample tests have the same distributions regardless of the starting point of the subsamples. One way of satisfying this assumption is to assume strict stationarity for $\mathbf{r}_t$ in Assumption 3 below.

Assumption 2 *The sampling distributions of ξ_{Nbs}, τ_{Nbs} and φ_{Nbs} are invariant to s at any fixed b and n.*

Temporal dependence structure of the data is assumed to be:

Assumption 3 $\mathbf{r}_t = [(I-B)^{\lambda_{x1}}\mathbf{x}'_{t1}, ..., (I-B)^{\lambda_{xk}}\mathbf{x}'_{tk}, \mathbf{z}'_{t1}, ..., \mathbf{z}'_{tl}, \mathbf{u}'_t]'$ *is either strong mixing with its mixing coefficients* $\alpha_{\mathbf{r},m}$ *satisfying for* $\delta > 0$

$$\sum_{m=1}^{\infty} \alpha_{\mathbf{r},m}^{\delta/(2+\delta)} < \infty$$

or uniform mixing with its mixing coefficients $\phi_{\mathbf{r},m}$ *satisfying*

$$\sum_{m=1}^{\infty} \phi_{\mathbf{r},m}^{1/2} < \infty.$$

The following theorem reports asymptotic properties of L^a_{Nnb} ($a = \xi$, τ and φ) under the null hypotheses (3) and (6).

Theorem 1 *Suppose that Assumptions 1-3 hold. Also, assume* $b/n \to 0$ *and* $b \to \infty$ *as* $n \to \infty$*. Then, under the null hypotheses (3) and (6),*

(i) $\sup_x \mid L^a_{Nnb}(x) - J^a_N(x) \mid \xrightarrow{a.s.} 0$ *as* $n \to \infty$*, where* $a = \xi$, τ *or* φ*;*

(ii) For $\lambda \in (0,1)$*, let* $c^a_{Nnb}(1-\lambda) = \inf\{x : L^a_{Nnb}(x) \geq 1-\lambda\}$*. If* $J^a_N(\cdot)$ *is continuous at* $c^a_{Nnb}(1-\lambda)$*, for* $a = \xi$, τ *or* φ*,*

$$P\{a_{Nn} \leq c^a_{Nnb}(1-\lambda)\} \to 1-\lambda \text{ as } n \to \infty.$$

Several aspects of this theorem deserve our attention. First, part (i) of Theorem 1 establishes that the subsample distribution $L^a_{Nnb}(x)$ becomes closer to the limiting distribution of a_{Nn} for every point x of $J^a_N(\cdot)$. The limiting distributions are allowed to be discontinuous as long as the discontinuity points satisfy part (ii) of Assumption 1. In the literature of subsampling, uniform convergence results have been available only for continuous limiting distributions and only in the mode of convergence in probability.

Second, part (ii) justifies using the percentiles from the subsample distribution $L^a_{Nnb}(\cdot)$ for hypothesis testing in the sense that the a_{Nn} test using the subsample critical values $c^a_{Nnb}(1-\lambda)$ has correct asymptotic size.

Third, without part (ii) of Assumption 1, we can obtain almost sure convergence of the empirical distributions only at the continuity points.[2]

Fourth, if the limiting distribution is continuous, the uniform convergence result of part (i) can be obtained in the mode of convergence in probability using Lemma A.1. This allows us to relax the mixing condition Assumption 3 to Assumption 6 that follows in next subsubsection.

[2] Uniform convergence of $B_{Nb}(\cdot)$ in the proof of Theorem 1 cannot be obtained without part (ii) of Assumption 1, though that of $A_{Nb}(\cdot)$ can be without it.

2.2.2 Centered subsampling

For the null hypothesis (3), we may use the subsample tests centered at the full-sample parameter estimates. This may bring higher power as observed in Choi (2004). Centered versions of tests (8) and (9) are, respectively,

$$\xi^{\bullet}_{Nbs} = g\left((\hat{\alpha}_{1bs} - \hat{\alpha}_{1n})' R'\Theta_{1bs}R(\hat{\alpha}_{1bs} - \hat{\alpha}_{1n}), ..., (\hat{\alpha}_{Nbs} - \hat{\alpha}_{Nn})'R'\Theta_{Nbs}R(\hat{\alpha}_{Nbs} - \hat{\alpha}_{Nn})\right) \tag{14}$$

and

$$\tau^{\bullet}_{Nbs} = h\left(\left(\hat{\alpha}^{p}_{Nbs} - \hat{\alpha}^{p}_{Nn}\right)' R'\Theta^{p}_{Nbs}R(\hat{\alpha}^{p}_{Nbs} - \hat{\alpha}^{p}_{Nn}\right). \tag{15}$$

Using the subsample tests (14) and (15), empirical distributions

$$L^{\bullet\xi}_{Nnb}(x) = \frac{1}{n-b+1}\sum_{s=1}^{n-b+1} 1\{\xi^{\bullet}_{Nbs} \leq x\}$$

$$L^{\bullet\tau}_{Nnb}(x) = \frac{1}{n-b+1}\sum_{s=1}^{n-b+1} 1\{\tau^{\bullet}_{Nbs} \leq x\}$$

are formulated.

For the centering at the full-sample parameter estimates to be innocuous under the null hypothesis (3), the following assumption is required.

Assumption 4 *(i)* $(R\hat{\alpha}_{in} - r)\Theta_{ibs}(R\hat{\alpha}_{in} - r) \xrightarrow{p} 0$ *and* $(R\hat{\alpha}_{ibs} - r)\Theta_{ibs}(R\hat{\alpha}_{in} - r) \xrightarrow{p.} 0$ *as* $n \to \infty$ *for all* i *and* s.

(ii) $(R\hat{\alpha}^{p}_{Nn} - r)'\Theta^{p}_{Nbs}(R\hat{\alpha}^{p}_{Nn} - R) \xrightarrow{p} 0$ *and* $(R\hat{\alpha}^{p}_{Nbs} - r)'\Theta^{p}_{Nbs}(R\hat{\alpha}^{p}_{Nn} - R) \xrightarrow{p} 0$ *as* $n \to \infty$ *for all* s.

This assumption can be shown to hold in applications using the requirement that $b/n \to 0$ as $n \to \infty$.

In relation to the centering, we also require

Assumption 5 *(i) First-order partial derivatives of the function* g *exist and satisfy* $\frac{\partial g}{\partial x_i}_{|x_i = f_{in}} = O_p(1)$ *for all* i *where* $f_{in} = o_p(1)$.

(ii) First-order partial derivative of the function h *exists and satisfy* $\frac{\partial h}{\partial x}_{|x = f_n} = O_p(1)$ *where* $f_n = o_p(1)$.

For the centered subsampling, our results are in the mode of convergence in probability. Corresponding results in the mode of almost sure convergence can be obtained using Lemma A.2 in the Appendix. But these require almost sure relations in Assumptions 4 and 5, which are cumbersome to verify in applications. Thus, we are content with developing theory for the centered subsampling in the mode of convergence in probability. Using the mode of convergence in probability instead of almost sure convergence does not bring any practical difficulties as long as limiting distributions are continuous.

Using the mode of convergence in probability allows us to relax Assumption 3 to the following:

Assumption 6 $\mathbf{r}_t = [(I-B)^{\lambda_{x1}}\mathbf{x}'_{t1}, ..., (I-B)^{\lambda_{xk}}\mathbf{x}'_{tk}, \mathbf{z}'_{t1}, ..., \mathbf{z}'_{tl}, \mathbf{u}'_t]'$ *is either strong mixing with its mixing coefficients* $\alpha_{\mathbf{r},m}$ *satisfying*

$$\sum_{m=1}^{\infty} m^{-1}\alpha_{\mathbf{r},m}^{\delta/(2+\delta)} < \infty$$

or uniform mixing with its mixing coefficients $\phi_{\mathbf{r},m}$ *satisfying*

$$\sum_{m=1}^{\infty} m^{-1}\phi_{\mathbf{r},m}^{1/2}.$$

The following theorem reports asymptotic properties of $L_{Nnb}^{\bullet a}$ ($a = \xi$ or τ) under the null hypotheses (3).

Theorem 2 *Suppose that Assumptions 1, 2, 4, 5 and 6 hold. Also, assume* $b/n \to 0$ *and* $b \to \infty$ *as* $n \to \infty$*. Then, under the null hypotheses (3),*

(i) $L_{Nnb}^{\bullet a}(x) - J_N^a(x) \xrightarrow{p} 0$ *as* $n \to \infty$*, where* $a = \xi$ *or* τ *and* x *is a continuity point of* $J_N^a(\cdot)$*;*

(ii) If $J_N^a(\cdot)$ *is continuous,* $\sup_x \mid L_{Nnb}^{\bullet a}(x) - J_N^a(x) \mid \xrightarrow{p} 0$ *as* $n \to \infty$*, where* $a = \xi$ *or* τ*;*

(iii) For $\lambda \in (0,1)$*, let* $c_{Nnb}^{\bullet a}(1-\lambda) = \inf\{x : L_{Nnb}^{\bullet a}(x) \geq 1-\lambda\}$*. If* $J_N^a(\cdot)$ *is continuous at* $c_{Nnb}^{\bullet a}(1-\lambda)$*, for* $a = \xi$ *or* τ*,*

$$P\{a_{Nn} \leq c_{Nnb}^{\bullet a}(1-\lambda)\} \to 1-\lambda \text{ as } n \to \infty.$$

This theorem shows that $L_{Nnb}^{\bullet a}(\cdot)$ has essentially the same asymptotic properties as $L_{Nnb}^a(\cdot)$ under the null hypothesis (3) as long as the limiting distribution $J_N^a(\cdot)$ is continuous. But they have different implications for the power of tests as will be discussed later. In addition, note that $L_{Nnb}^{\bullet a}(x)$ will be indeterminate at a discontinuity point x.[3] Thus, the assumption of continuity in part (i) is essential, and the uniform convergence result can be obtained only for continuous limiting distributions.

2.3 Test consistency

Suppose that the alternative hypothesis against the null hypothesis (3) can be written as

$$H_A : R\alpha_i = r^* \ (\neq r) \text{ for at least one } i. \tag{16}$$

The alternative hypothesis against the null hypothesis (6) is assumed to be

$$H_A : \theta_{ik} \leq a \text{ or } \theta_{ik} \geq b \text{ for at least one } i. \tag{17}$$

This subsection studies consistency of tests (4), (5) and (7) under the alternative hypotheses (16) and (17) that use critical values from the subsamplings of the previous subsections.

For the consistency of the tests, assume under the alternative hypotheses:

[3] Equation (A.11) in the Appendix shows that $L_{Nnb}^{\bullet a}(x)$ will be in between $J_N^a(x-)$ and $J_N^a(x+)$ in the limit at the discontinuity point x, which makes $L_{Nnb}^{\bullet a}(x)$ indeterminate in the limit.

Assumption 7 *(i)* $\xi_{Nn}/n^{\epsilon_\xi} \xrightarrow{p} \xi_N$, $\tau_{Nn}/n^{\epsilon_\tau} \xrightarrow{p} \tau_N$, *and* $\varphi_{Nn}/n^{\epsilon_\varphi} \xrightarrow{p} \varphi_N$, *where* $\epsilon_a > 0$ *for* $a = \xi, \tau$ *or* φ.

(ii) If a_{Nn} $(a = \xi, \tau$ *or* $\varphi)$ *rejects the relevant null hypothesis ((3) or (6)) when* a_{Nn} *is less (greater) than a critical value from its limiting distribution,* $P[a_N < 0]$ $(P[a_N > 0]) = 1$.

Part (i) of this theorem should be satisfied by all tests. Otherwise, the tests are inconsistent. Part (ii) is easy to verify in applications as will be shown in next section.

The following theorem reports that the probability of rejection under the alternative hypotheses (16) and (17) converges to one as the sample size n grows when the λ-level critical values $c^a_{Nnb}(1-\lambda)$ and $c^{\bullet a}_{Nnb}(1-\lambda)$ from the subsamplings are used.

Theorem 3 *(i) Suppose that Assumption 7 holds. Under the alternative hypotheses (16) and (17), for* $a = \xi$, τ *or* φ,

$$P\{a_{Nn} \leq c^a_{Nnb}(1-\lambda)\} \ (P\{a_{Nn} \geq c^a_{Nnb}(1-\lambda)\}) \to 1 \ \textit{as } n \to \infty \ \textit{as } n \to \infty.$$

(ii) Suppose that Assumptions 4 with r^* *replacing* r *and 7 hold. Under the alternative hypothesis (16), for* $a = \xi$ *or* τ,

$$P\{a_{Nn} \leq c^{\bullet a}_{Nnb}(1-\lambda)\} \ (P\{a_{Nn} \geq c^{\bullet a}_{Nnb}(1-\lambda)\}) \to 1 \ \textit{as } n \to \infty.$$

The a_{Nn} test of part (i) of this theorem is consistent because it diverges faster than the subsample critical value $c^a_{Nnb}(1-\lambda)$. Since $c^{\bullet a}_{Nnb}(1-\lambda)$ is stochastically bounded under the alternative, the power of the test in part (ii) will diverge faster than that of part (i).

3 Applications

3.1 Unit root tests for panel data

Consider the autoregressive model

$$\Delta y_{it} = \rho_i y_{i,t-1} + \sum_{j=1}^{p_i} \varphi_{ij} \Delta y_{i,t-j} + \mu_i + u_{it}, \ (i = 1, ..., N; t = \mathbf{p}+2, ..., n), \qquad (18)$$

where $[u_{1t}, ..., u_{Nt}]' \sim iid(0, \Sigma)$ $(\Sigma > 0)$, every element of matrix Σ is finite and $\mathbf{p} = \max_{1 \leq i \leq N} p_i$. Off-diagonal elements of the matrix Σ are allowed to be non-zero, which introduces cross-sectional correlation in the panel data. Obviously, this is an example of model (1). We assume p_i are known, though in practice these can be estimated by using information criteria and sequential testing. Since the index t starts from $\mathbf{p}$, every cross-sectional unit has the same sample and subsample sizes. For the autoregressive process (18), the number of time series observations is $n-\mathbf{p}-1$, and the number of subsamples with block size b is $n-b-\mathbf{p}$. Thus, all the subsample empirical distributions for the panel unit root tests should involve $n - b - \mathbf{p}$ in the denominator rather than $n - b + 1$.

It will be convenient to write model (18) in matrix notation as

$$\Delta \mathbf{y}_{in} = \rho_i \mathbf{y}_{i,-1,n} + \mathbf{Q}_{in}\gamma_i + u_{in},\ \ (i = 1, ..., N),$$

where $\Delta \mathbf{y}_{in} = [\Delta y_{1,p+2}, ..., \Delta y_{N,n}]'$, $\mathbf{Q}_{in}$=($\mathbf{1}$,$\Delta \mathbf{y}_{i,-1,n}, \Delta \mathbf{y}_{i,-2,n}, ..., \Delta \mathbf{y}_{i,-p_i,n}$) with $\mathbf{1} = [1, ..., 1]'$ and $\gamma_i = (\varphi_{i1}, ..., \varphi_{ip_i}, \mu_i)$.

The null and alternative hypotheses for panel unit root tests are:

$$H_0 : \rho_i = 0 \text{ for all } i \tag{19}$$

$$H_1 : \rho_i < 0 \text{ for at least one } i. \tag{20}$$

Under the null hypothesis, it is assumed that $\mu_i = 0$ for all i.

In addition to the iid assumption on $[u_{1t}, ..., u_{Nt}]'$, we assume

Assumption 8 $[\Delta y_{1t}, ..., \Delta y_{Nt}]'$ *satisfies Assumption 3.*

Under the null hypothesis, $[\Delta y_{1t}, ..., \Delta y_{Nt}]'$ is a strictly stationary linear process. Still, it requires extra conditions in order to be mixing. We do not state them here for brevity, but the reader may consult Davidson (1994, pp.219-228) for sufficient conditions that make a linear process mixing. These involve assumptions on the probability density functions of the underlying process as well as those on the coefficients of the linear process.

This section considers Levin, Lin and Chu's (2002; hereafter LLC), Im, Pesaran and Shin's (2003; hereafter IPS) and combination tests for the null hypothesis (19). The combination tests have independently been developed by Maddala and Wu (1999) and Choi (2001). Tests using model (18) with an additional linear time trend are easy to understand once those for model (18) are fully explored, and will be discussed later.

3.1.1 The LLC test

LLC pool the data and propose the t-test

$$t_{Nn} = \frac{\hat{\rho}^p_{Nn}}{\sqrt{\hat{\sigma}^2_{Nn}\left(\sum_{i=1}^N \mathbf{y}'_{i,-1,n}\mathbf{M}_{Q_{in}}\mathbf{y}_{i,-1,n}\right)^{-1}}} \tag{21}$$

for the null hypothesis (19), where $\hat{\rho}^p_{Nn}$ is a pooled OLS estimator of ρ_1 defined by $\hat{\rho}^p_{Nn} = \frac{\sum_{i=1}^N \mathbf{y}_{i,-1,n}\mathbf{M}_{Q_{in}}\Delta\mathbf{y}_{in}}{\sum_{i=1}^N \mathbf{y}_{i,-1,n}\mathbf{M}_{Q_{in}}\mathbf{y}_{i,-1,n}}$ with $\mathbf{M}_{Q_{in}} = I - \mathbf{Q}_{in}\left(\mathbf{Q}'_{in}\mathbf{Q}_{in}\right)^{-1}\mathbf{Q}'_{in}$ and $\hat{\sigma}^2_{Nn} = \frac{1}{N}\sum_{i=1}^N \frac{1}{n-p_i-2}\left(\mathbf{M}_{Q_{in}}\Delta\mathbf{y}_{in} - \hat{\rho}^p_{Nn}\mathbf{M}_{Q_{in}}\mathbf{y}_{i,-1,n}\right)'\left(\mathbf{M}_{Q_{in}}\Delta\mathbf{y}_{in} - \hat{\rho}^p_{Nn}\mathbf{M}_{Q_{in}}\mathbf{y}_{i,-1,n}\right)$ estimates $\frac{1}{N}\sum_{i=1}^N \sigma_i^2$. They use a modification of test (21) when N is infinite such that it has a standard normal distribution in the limit, but this is not required here because N is fixed.

Under the null hypothesis (19), as $n \to \infty$,

$$t_{Nn} \Rightarrow \frac{\sum_{i=1}^{N} \sigma_{li}\sigma_i \int_0^1 \bar{W}_i(r) dW_i(r)}{\sqrt{\sum_{i=1}^{N} \sigma_{li}^2 \int_0^1 \bar{W}_i^2(r) dr \frac{1}{N} \sum_{i=1}^{N} \sigma_i^2}}, \tag{22}$$

where $\bar{W}(r) = W(r) - \int_0^1 W(s)ds$ is the demeaned standard Brownian motion and σ_{li}^2 is the long-run variance of Δy_{it}. Even when σ_i^2 and σ_{li}^2 do not change over individuals, this distribution depends on the nuisance parameters due to the cross-sectional correlation and cannot be simulated unless the matrix Σ is known.

If y_{it} is cross-sectionally cointegrated, the distribution (22) should change and it depends on cointegrating vectors that are unknown. Simulating such distributions in each application will be quite cumbersome. For example, if $N = 2$ and $y_{1t} - cy_{2t} = I(0)$, we have $W_1(r) = cW_2(r)$ and this relation should be in (22). But still, note that Assumption 1 continue to hold in either case. Thus, we may ignore the case of cross-sectional cointegration for the LLC test.

The LLC test (21) is an example of the τ_{Nn} test given in (5) and Theorems 1, 2 and 3 can be applied to it as reported in the following corollary. In this corollary, X_{bs} denotes a submatrix of X_n that is made up of consecutive b rows of X_n beginning from the s-th row. Meanings of $t_{Nbs}, \hat{\rho}_{Nbs}^p$ and $\hat{\sigma}_{Nbs}^2$ should be obvious from the previous discussions. In addition, we set $p_i = \mathbf{p}$ in calculating the subsample tests in the following corollary, and this also applies to the tests in Corollary 5 and 6 below.

Corollary 4 *Let* $\tau_{Nbs} = t_{Nbs}$ *and* $\tau_{Nbs}^{\bullet} = \frac{\hat{\rho}_{Nbs}^p - \hat{\rho}_{Nn}^p}{\sqrt{\hat{\sigma}_{Nbs}^2 \left(\sum_{i=1}^{N} \mathbf{y}_{i,-1,bs}' \mathbf{M}_{Q_{ibs}} \mathbf{y}_{i,-1,bs}\right)^{-1}}}$.

(i) Under the null hypothesis (19), conclusions in Theorems 1 and 2 hold with $a = \tau$ *if Assumption 8 holds.*

(ii) Under the alternative hypothesis (20), conclusions in Theorem 3 hold with $a = \tau$.

Since the limiting distribution of the t_{Nn} test is continuous, we may obtain the uniform convergence result in part (i) of Theorem 1 in the mode of convergence in probability using Assumption (6) in Assumption (8) instead of Assumption (3) that is stronger (see the fourth comment after Theorem 1). The same is true of Corollaries 5 and 6 that follow.

3.1.2 The IPS test

Let t_{in} be the augmented Dickey-Fuller test for the null hypothesis $H_0 : \rho_i = 0$, i.e.,

$$t_{in} = \frac{\hat{\rho}_{in}}{\sqrt{\hat{\sigma}_{in}^2 \left(\mathbf{y}_{i,-1,n}' \mathbf{M}_{Q_{in}} \mathbf{y}_{i,-1,n}\right)^{-1}}},$$

where $\hat{\rho}_{in}$ is the OLS estimator of ρ_i using the time series data for the i-th individual and $\hat{\sigma}_{in}^2 = \frac{1}{n-p_i-2} \left(\mathbf{M}_{Q_{in}} \Delta \mathbf{y}_{in} - \hat{\rho}_{in} \mathbf{M}_{Q_{in}} \mathbf{y}_{i,-1,n}\right)' \left(\mathbf{M}_{Q_{in}} \Delta \mathbf{y}_{in} - \hat{\rho}_{in} \mathbf{M}_{Q_{in}} \mathbf{y}_{i,-1,n}\right)$.

The IPS t–bar statistic is defined as the average of the individual augmented Dickey-Fuller test as

$$\bar{t}_{Nn} = \frac{1}{N}\sum_{i=1}^{N} t_{in}. \tag{23}$$

For a fixed N, as $T \to \infty$,

$$\bar{t}_{Nn} \Rightarrow \frac{1}{N}\sum_{i=1}^{N} \frac{\int_0^1 \bar{W}_i(r)dW_i(r)}{\sqrt{\int_0^1 \bar{W}_i^2(r)dr}} \tag{24}$$

under the null hypothesis (19). Though the limiting distribution appears to be free of nuisance parameters, indeed it involves them due to the cross-sectional correlation. The limiting distribution (24) prevails even if y_{it} is cross-sectionally cointegrated, because the test is the average of the individual Dickey-Fuller test that is not affected by cross-sectional cointegration.

Test (23) is an example of the ξ_{Nn} test given in (4) and Theorems 1, 2 and 3 can be applied to it.

Corollary 5 *Let* $\xi_{Nbs} = \bar{t}_{Nbs}$ *and* $\xi^{\bullet}_{Nbs} = \frac{1}{N}\sum_{i=1}^{N} t^{\bullet}_{ibs}(p_i, \varphi_i)$ *with* $t_{ibs}(p_i, \varphi_i) = \frac{\hat{\rho}_{ibs} - \hat{\rho}_{in}}{\sqrt{\hat{\sigma}^2_{ibs}\left(\mathbf{y}'_{i,-1,bs}\mathbf{M}_{Q_{i,bs}}\mathbf{y}_{i,-1,bs}\right)^{-1}}}$.

(i) Under the null hypothesis (19), conclusions in Theorems 1 and 2 hold with $a = \xi$ *if Assumption 8 holds.*

(ii) Under the alternative hypothesis (20), conclusions in Theorem 3 hold with $a = \xi$.

3.1.3 Combination tests

A unit root test is applied to each cross-sectional unit and the resulting p-values are combined to formulate combination tests. Choi (2001) uses Elliott, Rothenberg and Stock's (1996)'s Dickey-Fuller GLS test as an underlying unit root test.[4] To calculate the Dickey-Fuller-GLS tests, let $a_i = 1 - c_i/n$, $\Delta_c \mathbf{y}_{in} = (y_{i1}, y_{i2} - a_i y_{i1}, ..., y_{in} - a_i y_{i,n-1})'$ and $\mathbf{w}_{a_i n} = (1, 1 - a_i, ..., 1 - a_i)'$. Then, using $c_i = -7$ as suggested in Elliott, Rothenberg and Stock, run the OLS regression

$$\Delta_c \mathbf{y}_{in} = \mathbf{w}_{a_i n}\hat{\beta} + \hat{\mathbf{f}}_{a_i n}. \tag{25}$$

Next, formulate the GLS-detrended series $y^g_{it} = y_{it} - \hat{\beta}$ by using $\hat{\beta}$ from equation (25) and run the augmented Dickey-Fuller regression

$$\triangle \mathbf{y}^g_{in} = \hat{\rho}^g_{in}\mathbf{y}^g_{i,-1,n} + \hat{\varphi}^g_{i1n}\triangle \mathbf{y}^g_{i,-1,n} + ... + \hat{\varphi}^g_{ip_{in}}\triangle \mathbf{y}^g_{i,-p_i,n} + \hat{\mathbf{g}}_{in},$$

[4] The Dickey-Fuller-GLS test provides better empirical size and power than the Dickey-Fuller test according to some unreported simulation results.

where $\mathbf{y}_{in}^g = [y_{i,p_i+2} - \hat{\beta}, ..., y_{i,n} - \hat{\beta}]'$ and $\mathbf{y}_{i,-k,n}^g$ are similarly defined. The Dickey-Fuller-GLS test is defined by

$$dfg_{in}^{\mu} = \frac{\hat{\rho}_{in}^g}{\sqrt{\hat{\sigma}_{in}^{g2}\left(\mathbf{y}_{i,-1,n}^{g\prime}\mathbf{M}_{Q_{in}^g}\mathbf{y}_{i,-1,n}^g\right)^{-1}}},$$

where $\hat{\sigma}_{in}^{g2} = \frac{1}{n-p_i-2}\left(\mathbf{M}_{Q_{in}^g}\Delta\mathbf{y}_{in}^g - \hat{\rho}_{in}^g\mathbf{M}_{Q_{in}^g}\mathbf{y}_{i,-1,n}^g\right)'\left(\mathbf{M}_{Q_{in}^g}\Delta\mathbf{y}_{in}^g - \hat{\rho}_{in}^g\mathbf{M}_{Q_{in}^g}\mathbf{y}_{i,-1,n}^g\right)$ with $\mathbf{M}_{Q_{in}^g} = I - \mathbf{Q}_{in}^g(\mathbf{Q}_{in}^{g\prime}\mathbf{Q}_{in}^g)^{-1}\mathbf{Q}_{in}^{g\prime}$ and $\mathbf{Q}_{in}^g$=$(\mathbf{1}$,$\Delta\mathbf{y}_{i,-1,n}^g, \Delta\mathbf{y}_{i,-2,n}^g, ..., \Delta\mathbf{y}_{i,-p_i,n}^g)$.

The asymptotic p-value for the Dickey-Fuller-GLS test is defined as

$$p_{in} = F(dfg_{in}^{\mu})$$

where $F(\cdot)$ denotes the Dickey-Fuller-GLS test's limiting cumulative distribution function.

There are various combination tests (cf. Hedges and Olkin, 1985). The most desirable one, according to the results in Choi (2000), appears to be the inverse normal test defined by

$$Z_{Nn} = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Phi^{-1}(p_{in}) \tag{26}$$

where $\Phi(.)$ is the standard normal cumulative distribution function. Under the null hypothesis (19),

$$Z_{Nn} \Rightarrow \frac{1}{\sqrt{N}}\sum_{i=1}^{N} z_i$$

as $n \to \infty$, where $z_i \sim N(0,1)$ and z_i are dependent due to the cross-sectional correlation. This distribution is also valid even if y_{it} is cross-sectionally cointegrated for the same reason as for the IPS test.

The Z_{Nn} test is an example of the ξ_{Nn} test, and the following results are deduced from Theorems 1, 2 and 3.

Corollary 6 *Let $\xi_{Nbs} = Z_{Nbs}$ and $\xi_{Nbs}^{\bullet} = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Phi^{-1}(p_{ibs}^{\bullet})$ with $p_{ibs}^{\bullet} = F(dfg_{in}^{\bullet\mu})$ and $dfg_{in}^{\bullet\mu} = \frac{\hat{\rho}_{ibs}^g - \hat{\rho}_{in}^g}{\sqrt{\hat{\sigma}_{ibs}^{g2}\left(\mathbf{y}_{i,-1,bs}^{g\prime}\mathbf{M}_{Q_{ibs}^g}\mathbf{y}_{i,-1,bs}^g\right)^{-1}}}$.*

(i) Under the null hypothesis (19), conclusions in Theorems 1 and 2 hold with $a = \xi$ if Assumption 8 holds.

(ii) Under the alternative hypothesis (20), conclusions in Theorem 3 hold with $a = \xi$.

3.2 Stationarity test for panel data

Consider the linear panel regression model

$$y_{it} = \mu_i + u_{it},\ u_{it} = \rho_i u_{i,t-1} + v_{it}, (i = 1, ..., N; t = 2, ..., n), \tag{27}$$

which is a special case of model (1). Assuming that v_{it} are $I(0)$ for all i and that v_{it} are cross-sectionally correlated, we are interested in testing the null hypothesis

$$H_0 : |\rho_i| < 1 \text{ for all } i \tag{28}$$

against the alternative hypothesis

$$H_A : \rho_i = 1 \text{ for at least one } i. \tag{29}$$

We will use Hadri's (2000) Lagrange Multiplier test for the null hypothesis (28). Letting $s_{it} = \sum_{k=1}^{t}(y_{ik} - \bar{y}_i)$ with $\bar{y}_i = \frac{1}{n}\sum_{t=1}^{n} y_{it}$, it is defined as

$$\eta_{Nn} = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{n^2\hat{\sigma}_i^2}\sum_{t=1}^{n} s_{it}^2, \tag{30}$$

where $\hat{\sigma}_i^2$ is an estimator of the long-run variance of y_{it} defined by $\hat{\sigma}_i^2 = \sum_{j=-l}^{l} C_i(j)k(\frac{j}{l})$ with $C_i(j) = \frac{1}{n}\sum_{t=1}^{n-j}(y_{it} - \bar{y}_i)(y_{i,t+j} - \bar{y}_i)$ and $k(\cdot)$ being a lag window. This test may be considered as the average of Kwiatkowski, Phillips, Schmidt and Shin's (1992) test of level-stationarity for each individual.

Assume under the null hypothesis (28):

Assumption 9 *(i) u_{it} is strictly stationary, satisfies Assumption 3 for all i and*

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]} u_{it} \Rightarrow \sigma_i W_i(r) \text{ as } n \to \infty.$$

(ii) $\hat{\sigma}_i^2 \xrightarrow{p} \sigma_i^2$ for all i as $n \to \infty$.

More primitive conditions for this assumption to hold are well documented in the literature. See, for example, Kwiatkowski, Phillips, Schmidt and Shin (1992).

Under Assumptions 9, it follows that

$$\eta_{Nn} \Rightarrow \frac{1}{N}\sum_{i=1}^{N}\int_0^1 (W_i(r) - rW_i(1))^2\, dr \text{ as } n \to \infty. \tag{31}$$

Since v_{it} are cross-sectionally correlated, $W_i(r)$ are not independent.

Under the alternative hypothesis (29), assume

Assumption 10

$$\frac{1}{n^2\hat{\sigma}_i^2}\sum_{t=1}^{n} s_{it}^2 = O_p(l/n).$$

Note that this assumption is proven in Kwiatkowski, Phillips, Schmidt and Shin (1992) based on more primitive conditions. Obviously, this assumption holds even when y_{it} is cross-sectionally cointegrated under the alternative.

Test (30) is an example of φ_{Nn} test given in (7). Thus, Theorems 1, 2 and 3 provide the following corollary.

Corollary 7 *Let* $\varphi_{Nbs} = \eta_{Nbs}$.

(i) Under the null hypothesis (28), conclusions in Theorems 1 hold with $a = \varphi$ *if Assumption 9 holds.*

(ii) Under the alternative hypothesis (29), conclusions in Theorem 3 hold with $a = \varphi$ *if Assumption 10 holds.*

3.3 Extensions

This subsection considers extensions of the unit root and stationarity tests in previous subsections to more complex models. The models considered are the linear cointegration model and extensions of models (18) and (27) with a linear time trend. The extensions are rather straightforward given the results in previous subsections, and hence discussions on them will be brief.

3.3.1 Panel regression and cointegration tests

We may employ the unit root and stationarity tests in previous subsections to test for cointegration. Consider the linear panel regression model with $I(1)$ regressors x_{it}

$$y_{it} = \mu_i + \alpha_i' x_{it} + u_{it},\ u_{it} = \rho_i u_{i,t-1} + v_{it}, \tag{32}$$

which is a special case of model (1). In order to test for cointegration, we run OLS on this model using x_{it} as regressors and get residuals $\hat{u}_{it}$. For the null of noncointegration (i.e., $\rho_i = 1$ for all i), we use tests (21), (23) and (26) assuming that v_{it} follows autoregressive processes. The limiting distributions of these tests change with the presence of the $I(1)$ regressor and depend on nuisance parameters, but the subsampling method still work as long as $[(\Delta x_{1t}', v_{it}), ..., (\Delta x_{Nt}', v_{Nt})]'$ satisfies Assumption 3. Thus, critical values of these tests can be estimated using the subsampling method. Similarly, critical values of test (30) for the null of cointegration can be estimated by using the OLS residuals and the subsampling method. It is well known in time series regression that efficient estimation method (e.g., lead-and-lags regression) should be used to make the limiting distributions of stationarity tests free of nuisance parameters (cf. Choi and Ahn, 1995). However, the OLS regression may be used for the subsampling, because the nuisance parameters resulting from OLS can be handled by the subsampling method. In finite samples, though, using efficient estimation methods may induce the tests to perform better.

3.3.2 Models with a linear time trend

A linear time trend variable t may be added to models (18), (27) and (32). Limiting distributions of the unit root and stationarity tests will change, but no doubt the subsampling method still works since the trend variable is nonstochastic and brings no meaningful changes.

4 Simulation for panel unit root tests

4.1 Choice of block sizes

No doubt, choice of block sizes will affect the performance of unit root tests using subsamplings. Here we devise simulation-based calibration rule for the choice of block sizes. Choi (2004) shows that this method works better for the VAR causality test than other existing methods like minimum volatility and bootstrap-based calibration rules. We assume that an adequate approximation for an optimal block size at each nominal size λ has the following relation to the sample sizes and the common AR order $p+1$[5]:

$$b^{opt,\lambda} = \exp(\theta_0 + \theta_1 \times (p+1) + \theta_2 \times N) n^{\beta}. \tag{33}$$

Other factors (e.g., characteristic roots of the AR model, degree of cross-sectional correlation, etc.) should also affect optimal block sizes. But relation (33) is simple to use and provides reasonably good approximations to optimal block sizes according to simulation results that will be reported. Relation (33) with estimated parameters provides a calibration rule for the choice of the block size b. In practice, the numbers of the lagged difference terms will differ across cross-sectional units. In this case, one may use mean or median of $\{p_i\}$. This will not affect the optimal choice of the block size in any significant way because the effect of parameter p on the choice is minimal as will be shown in Tables I and II.

In order to estimate the parameters of relation (33), we ran simulations for various sample sizes and data generating processes, and relate optimal block sizes to these. The data generating process used is (18) with $\mu_i = 0$ for all i; $N = 5, 10, 20$; $n = 80, 130, 180$ and $p_i = 0, 2, 4$ for all i[6]. We let all of the off-diagonal elements of Σ be 0.3 and all of the diagonal elements 1.[7] In this experimental format, there are 27 different data generating processes. The calibration rule was devised separately for the 5% and 10% significance levels and for the centered and non-centered subsamplings. The algorithm we used for the calibration rule at each significance level is:

> *Step 1:* Generate each data 1,000 times and calculate the subsample critical values for every block size from 5 to $0.8 \times n$. In addition, record the finite sample critical values of the full-sample unit root tests out of the 1,000 iterations.
>
> *Step 2:* Record the median of the 1,000 subsample critical values of Step 2 for each block size.
>
> *Step 3:* Record the median block size of Step 2 that best approximates the finite sample critical values of the full-sample unit root test of Step 1 in terms of absolute discrepancy.

[5] Here p is the number of lagged difference terms as in model (18). So the AR order is $p+1$.

[6] For the AR(3) model, the values of the characteristic roots are 1, 0.5 and 0.3. For the AR(5) model, they are 1, 0.5, 0.4, 0.3 and 0.2.

[7] We also performed some experiments with the value 0.5 used for the off-diagonal elements. These did not bring any noticeable changes in the optimal block sizes.

Step 4: Regress the natural logarithm of the median block size from Step 3 on 1, AR order, N and $\ln(n)$ to estimate parameters $\theta_0, \theta_1, \theta_2$ and β.

Steps 1-3 provide median block sizes (out of 1,000 iterations) that produce subsample critical values closest to the corresponding finite sample critical values of the unit root tests. The calibration rule obtained from the above algorithm is reported in Table 1. Table 1 shows that the optimal block sizes are increasing functions of n. Effects of M and p differ across the unit root tests. But p does not affect the optimal block sizes much as the magnitudes of the numbers for θ_1 in Table 1 show. Table 1 indicates, for example, that the optimal block size of the non-centered subsampling for the LLC test at the 5% level is the integer nearest to

$$exp(0.5462 + 0.1441 \times (p + 1) - 0.01814 \times M) \times n^{0.5469}. \tag{34}$$

Table I: Simulation-Based Calibration Rules for Subsampling Panel Unit Root Tests With an Intercept

(1) The LLC test

	Significance level	θ_0	θ_1	θ_2	β
Non-centered	5%	0.5462	0.1441	-0.01814	0.5469
	10%	0.08549	0.1370	-0.01632	0.6490
Centered	5%	-0.4187	0.2211	-0.02256	0.5261
	10%	0.01289	0.2260	-0.02267	0.4053

(2) The IPS test

	Significance level	θ_0	θ_1	θ_2	β
Non-centered	5%	0.2656	-0.01091	-0.02577	0.6391
	10%	1.302	-0.06591	-0.03439	0.4456
Centered	5%	0.6402	0.1253	-0.01850	0.3173
	10%	0.8806	0.1232	-0.01573	0.2285

(3) The Z_{Nn} test

	Significance level	θ_0	θ_1	θ_2	β
Non-centered	5%	0.03522	-0.02201	0.01485	0.8185
	10%	0.1640	-0.01685	0.01098	0.8232
Centered	5%	-0.5293	-0.02970	0.003535	0.8607
	10%	-0.6189	-0.02995	0.002260	0.8898

We follow the same steps as for Table 1 using model (18) with a linear time trend term added. The coefficient values for the intercept and linear time trend terms were set at zero for all i. The results for the optimal choice of block sizes are reported in the following table. Again, the optimal block sizes are increasing functions of n; and the effects of M and p differ across the unit root tests.

Table II: Simulation-Based Calibration Rules for Subsampling Panel Unit Root Tests With an Intercept and a Linear Time Trend

(1) The LLC test

	Significance level	θ_0	θ_1	θ_2	β
Non-centered	5%	0.6321	0.07859	-0.004126	0.6059
	10%	1.121	0.06738	-0.007120	0.5238
Centered	5%	-0.1063	0.1679	-0.01377	0.5304
	10%	-0.3636	0.1660	-0.01507	0.5651

(2) The IPS test

	Significance level	θ_0	θ_1	θ_2	β
Non-centered	5%	1.521	-0.08973	-0.03099	0.4681
	10%	2.838	-0.1451	-0.03660	0.1919
Centered	5%	0.8209	0.1012	-0.01641	0.3002
	10%	0.9355	0.1084	-0.01460	0.2388

(3) The Z_{Nn} test

	Significance level	θ_0	θ_1	θ_2	β
Non-centered	5%	0.006859	-0.04031	0.009892	0.8350
	10%	-0.1006	-0.03730	0.008551	0.8854
Centered	5%	-0.4720	-0.03896	-0.003550	0.7690
	10%	-0.5714	-0.04727	-0.003045	0.7944

4.2 Empirical size and power

This subsubsection studies empirical size and power of the panel unit root tests using subsampling critical values. Data were generated using model (18). A linear time trend term was added to the model for the unit root tests for the null of a unit root with drift. The coefficient values for the intercept and linear time trend terms were set at zero for all i. Other aspects of the data generating scheme are $p_i = 0, 2$; $M = 5, 30, 50$; $n = 100, 200$ and $\Sigma = \begin{bmatrix} 1 & \omega & \cdots & \omega \\ \omega & 1 & & \vdots \\ \vdots & & \ddots & \omega \\ \omega & \cdots & \omega & 1 \end{bmatrix}$. For the AR(1) model, we used the data generating scheme $(1-\alpha L)y_{it} = u_{it}$ where L denotes the usual lag operator. For the AR(3) model, we used $(1-\alpha L)(1-0.5L)(1-0.3L)y_{it} = u_{it}$. The null of a unit root corresponds to $\alpha = 1$. For the alternative of stationarity, $\alpha = 0.97$ and $\alpha = 0.95$ were considered. Random numbers were generated by $[u_{1t}, ..., u_{Nt}]' \sim iid$ $N(0, \Sigma)$ for $t = 1, ..., n + 30$ and the last n vectors were used for data generation. Empirical size was calculated using $\omega = 0.3$ and $\omega = 0.6$. Empirical power used $\omega = 0.3$ only. Note that the data generating scheme for the empirical size and power is chosen to be different from that for the calibration rule except that $M = 5$ and $\omega = 0.3$ are used. The purpose is to check the validity of the calibration rule for different data generating processes. The numbers of iterations are 5,000 for empirical size and 2,000 for empirical power. The nominal size used was 5%.

Empirical size of the panel unit root tests using subsample critical values under the calibration rule is reported in Table III, and empirical power in Table IV. Results in the Table III can be summarized as follows.

- The LLC test using critical values from the non-centered subsampling keep empirical size reasonably well both in the cases $\omega = 0.3$ and $\omega = 0.6$. Differences between the two cases are minimal.

- The IPS and Z_{Nn} tests using critical values from the non-centered subsampling keep empirical size reasonably well both in the cases $\omega = 0.3$ and $\omega = 0.6$ unless $M = 50$. Again,the two cases $\omega = 0.3$ and $\omega = 0.6$ do not show any noticeable differences.

- Excluding the case $M = 50$, the three tests show similar size performance when critical values from the non-centered subsamplings are used. In fact, the mean empirical sizes of the LLC, IPS and Z_{Nn} tests excluding the case $M = 50$ are 0.039, 0.040 and 0.041, respectively, which are reasonably close to the nominal size 0.05.

- The centered subsamplings tend to bring size distortions for all the tests, though these are mild for the LLC test. In particular, the Z_{Nn} test using the centered subsampling tend to overreject.

- In empirical practice, when M is not large relative to n, all the tests with the non-centered subsamplings may be used. When M is large relative to n, the LLC test with the non-centered subsampling is recommended. Using the centered subsampling for the Z_{Nn} test is not recommended in any case.

Results in Table IV are summarized as follows.

- All the tests show significantly improved empirical power when the centered subsamplings are used. But note that the centered subsamplings bring size distortions especially for the Z_{Nn} test as we have seen in Table III.

- Comparing the power of the tests using non-centered subsamplings, the LLC and Z_{Nn} tests tend to be more powerful than the IPS test.

- Empirical power improves as the sample sizes increase, though there are a few exceptions.

- In practice, if keeping the nominal size is utmost important, it is appropriate to use the LLC and Z_{Nn} tests with the non-centered subsamplings because these tests are more powerful than the IPS test and keep the nominal size well. But if risking a slight degree of size distortion is allowed, using the LLC test with the centered subsampling appears to be a proper choice because this test keeps the nominal size relatively well and is reasonably powerful.

5 Applications to the PPP hypothesis

To be written.

Table III: Empirical Size

Notes: 1. For the AR(1) model, the data generating scheme $(1-\alpha L)y_{it} = u_{it}$ was used. For the AR(3) model, we used $(1-\alpha L)(1-0.5L)(1-0.3L)y_{it} = u_{it}$.

2. Random numbers were generated by $[u_{1t},...,u_{Nt}]' \sim iid\ N(0,\Sigma)$ for $t = 1,...,n+30$ and the last n vectors were used for data generation.

3. The off-diagonal elements of Σ are all ω, and the diagonal elements are ones.

4. The number of iterations is 5,000.

(1) $\omega = 0.3$

	AR order	n	N	Non-centered (5%) LLC	IPS	Z	Centered (5%) LLC	IPS	Z
Demeaned	$p=1$	100	5	0.033	0.038	0.019	0.069	0.084	0.105
			30	0.040	0.027	0.018	0.060	0.034	0.116
			50	0.035	0.009	0.093	0.028	0.000	0.171
		200	5	0.044	0.042	0.029	0.063	0.060	0.110
			30	0.042	0.034	0.028	0.055	0.030	0.123
			50	0.043	0.027	0.056	0.042	0.003	0.165
	$p=3$	100	5	0.037	0.033	0.017	0.058	0.068	0.099
			30	0.022	0.044	0.011	0.061	0.019	0.112
			50	0.026	0.007	0.078	0.053	0.000	0.147
		200	5	0.038	0.051	0.025	0.053	0.066	0.104
			30	0.027	0.055	0.030	0.034	0.058	0.120
			50	0.022	0.043	0.068	0.036	0.000	0.151
Demeaned and Detrended	$p=1$	100	5	0.049	0.054	0.040	0.102	0.112	0.120
			30	0.033	0.033	0.030	0.066	0.060	0.106
			50	0.032	0.005	0.066	0.022	0.000	0.077
		200	5	0.057	0.053	0.046	0.091	0.074	0.105
			30	0.046	0.038	0.052	0.054	0.038	0.080
			50	0.037	0.015	0.075	0.017	0.001	0.048
	$p=3$	100	5	0.040	0.037	0.035	0.066	0.073	0.101
			30	0.031	0.053	0.026	0.043	0.026	0.103
			50	0.025	0.000	0.035	0.019	0.000	0.078
		200	5	0.052	0.051	0.048	0.078	0.060	0.098
			30	0.038	0.057	0.042	0.042	0.077	0.072
			50	0.033	0.027	0.054	0.019	0.000	0.040

(2) $\omega = 0.6$

	AR order	n	N	Non-centered (5%)			Centered (5%)		
				LLC	IPS	Z	LLC	IPS	Z
Demeaned	$p = 1$	100	5	0.027	0.032	0.023	0.055	0.069	0.116
			30	0.040	0.023	0.053	0.033	0.024	0.152
			50	0.023	0.008	0.141	0.007	0.000	0.204
		200	5	0.051	0.049	0.033	0.060	0.054	0.117
			30	0.039	0.037	0.058	0.036	0.022	0.149
			50	0.031	0.022	0.010	0.017	0.003	0.193
	$p = 3$	100	5	0.029	0.031	0.018	0.045	0.054	0.103
			30	0.017	0.020	0.022	0.018	0.011	0.129
			50	0.017	0.004	0.092	0.007	0.000	0.172
		200	5	0.034	0.041	0.034	0.045	0.050	0.115
			30	0.024	0.030	0.059	0.016	0.021	0.148
			50	0.014	0.015	0.092	0.006	0.000	0.168
Demeaned and Detrended	$p = 1$	100	5	0.056	0.053	0.054	0.104	0.090	0.113
			30	0.036	0.028	0.077	0.044	0.032	0.101
			50	0.028	0.006	0.125	0.011	0.000	0.083
		200	5	0.053	0.053	0.062	0.071	0.065	0.106
			30	0.044	0.031	0.094	0.036	0.019	0.090
			50	0.040	0.012	0.141	0.012	0.002	0.063
	$p = 3$	100	5	0.041	0.036	0.040	0.058	0.059	0.091
			30	0.027	0.025	0.059	0.024	0.012	0.102
			50	0.021	0.000	0.062	0.005	0.000	0.087
		200	5	0.051	0.051	0.059	0.068	0.049	0.105
			30	0.037	0.027	0.071	0.020	0.023	0.082
			50	0.033	0.006	0.097	0.007	0.000	0.062

Table IV: Empirical Power

Notes for Table III apply here too except that the number of iterations is 2,000 and that only $\omega = 0.3$ was used.

(1) $\alpha = 0.97$

	AR order	n	N	Non-centered (5%)			Centered (5%)		
				LLC	IPS	Z	LLC	IPS	Z
Demeaned	$p = 1$	100	5	0.137	0.098	0.157	0.287	0.254	0.512
			30	0.439	0.309	0.287	0.595	0.433	0.863
			50	0.502	0.218	0.275	0.522	0.006	0.904
		200	5	0.473	0.441	0.417	0.668	0.587	0.810
			30	0.920	0.864	0.657	0.970	0.892	0.982
			50	0.959	0.893	0.493	0.971	0.712	0.989
	$p = 3$	100	5	0.101	0.081	0.129	0.182	0.186	0.465
			30	0.188	0.296	0.239	0.372	0.183	0.839
			50	0.251	0.083	0.258	0.407	0.000	0.892
		200	5	0.354	0.352	0.376	0.537	0.508	0.799
			30	0.810	0.847	0.678	0.879	0.872	0.979
			50	0.852	0.871	0.558	0.908	0.043	0.983
Demeaned and Detrended	$p = 1$	100	5	0.086	0.087	0.080	0.175	0.187	0.235
			30	0.125	0.103	0.127	0.248	0.185	0.374
			50	0.115	0.026	0.199	0.096	0.000	0.343
		200	5	0.232	0.209	0.279	0.371	0.303	0.519
			30	0.526	0.446	0.522	0.656	0.464	0.791
			50	0.609	0.413	0.569	0.554	0.155	0.793
	$p = 3$	100	5	0.060	0.060	0.064	0.109	0.121	0.175
			30	0.092	0.125	0.112	0.138	0.067	0.324
			50	0.084	0.000	0.111	0.074	0.000	0.301
		200	5	0.199	0.180	0.247	0.307	0.217	0.465
			30	0.416	0.470	0.470	0.498	0.503	0.722
			50	0.467	0.341	0.522	0.414	0.000	0.746

(2) $\alpha = 0.95$

	AR order	n	N	Non-centered (5%)			Centered (5%)		
				LLC	IPS	Z	LLC	IPS	Z
Demeaned	$p = 1$	100	5	0.284	0.252	0.261	0.536	0.495	0.738
			30	0.790	0.662	0.422	0.889	0.770	0.957
			50	0.856	0.610	0.325	0.873	0.043	0.978
		200	5	0.882	0.825	0.604	0.958	0.931	0.933
			30	0.999	0.996	0.760	1.00	0.996	0.998
			50	1.00	0.997	0.511	1.00	0.985	0.997
	$p = 3$	100	5	0.184	0.165	0.230	0.357	0.344	0.652
			30	0.461	0.556	0.364	0.694	0.419	0.944
			50	0.567	0.267	0.308	0.710	0.000	0.963
		200	5	0.705	0.719	0.549	0.884	0.851	0.926
			30	0.986	0.990	0.763	0.995	0.993	0.993
			50	0.994	0.991	0.996	0.993	0.411	0.594
Demeaned and Detrended	$p = 1$	100	5	0.166	0.152	0.174	0.312	0.287	0.395
			30	0.319	0.276	0.288	0.541	0.406	0.698
			50	0.357	0.125	0.383	0.362	0.001	0.699
		200	5	0.547	0.526	0.535	0.744	0.672	0.844
			30	0.915	0.912	0.805	0.962	0.923	0.986
			50	0.942	0.907	0.784	0.954	0.725	0.983
	$p = 3$	100	5	0.101	0.100	0.137	0.194	0.196	0.319
			30	0.196	0.223	0.246	0.292	0.140	0.604
			50	0.214	0.001	0.255	0.192	0.000	0.620
		200	5	0.431	0.419	0.466	0.642	0.497	0.763
			30	0.816	0.864	0.770	0.895	0.882	0.955
			50	0.836	0.790	0.767	0.848	0.017	0.966

Appendix: Proofs

Functions of a mixing process with a fixed number of arguments are also mixing. However, functions of a mixing process with a growing number of arguments are not necessarily mixing and the laws of large numbers for these have not been available. The following two lemmas prove the weak and strong laws of large number for the functions. These will be used to prove Theorems 1 and 2. In the lemmas, we use the notation $\|a\|_p = (E|a|^p)^{1/p}$.

Lemma A.1 *Let* $Y_t = f(X_t, X_{t+1}, ..., X_{t+b})$ *where* b *is an integer satisfying* $b = O(n^\eta)$ *with* $0 < \eta < 1$. *Suppose for some constants* c *and* d *that either*

(1a) $\sup_{t\geq 1} \|Y_t\|_{2+\delta} < c < \infty$ *;*

(1b) $\{X_t\}$ *is strong mixing with its mixing coefficients* $\alpha_{X,m}$ *satisfying for* $\delta > 0$

$$\sum_{m=1}^{\infty} m^{-1} \alpha_{X,m}^{\delta/(2+\delta)} < \infty$$

or

(2a) $\sup_{t\geq 1}\|Y_t\|_2 < d < \infty$*;*

(2b) $\{X_t\}$ *is uniform mixing with its mixing coefficients* $\phi_{X,m}$ *satisfying*

$$\sum_{m=1}^{\infty} m^{-1}\phi_{X,m}^{1/2} < \infty$$

hold. Let $S_{n-b} = \sum_{t=1}^{n-b}(Y_t - E(Y_t))$. *Then, as* $n \to \infty$,

$$\frac{1}{n-b}S_{n-b} \xrightarrow{p} 0.$$

Proof. Suppose that $\{X_t\}$ is a strong mixing sequence, and let $\mathcal{G}_{-\infty}^{t} = \sigma(..., Y_{t-1}, Y_t)$ and $\mathcal{G}_{t+m}^{\infty} = \sigma(Y_{t+m}, Y_{t+m-1}, ...)$ where $\sigma(\cdot)$ denotes the smallest σ-field generated by the random variables in the parenthesis. Since Y_t is measurable on any $\sigma-$field on which $X_t, X_{t+1}, ..., X_{t+b}$ are measurable, $\mathcal{G}_{-\infty}^{t} \subseteq \mathfrak{F}_{-\infty}^{t+b} = \sigma(..., X_{t+b-1}, X_{t+b})$ and $\mathcal{G}_{t+m}^{\infty} \subseteq \mathfrak{F}_{t+m}^{\infty} = \sigma(X_{t+m}, X_{t+m+1}, ...)$. For $m > b$, these relations give

$$\alpha_{Y,m} \leq \alpha_{X,m-b}, \tag{A.1}$$

where $\alpha_{Y,m}$ is the mixing coefficients for $\{Y_t\}$ defined by $\alpha_{Y,m} = \sup_t \alpha\left(\mathcal{G}_{-\infty}^{t}, \mathcal{G}_{t+m}^{\infty}\right)$. Letting $Cov(Y_t, Y_{t+m}) = \sigma_{t,t+m}$, we have due to the strong mixing inequality for $\{Y_t\}$ (cf. Davidson, 1994, p.212),

$$|\sigma_{t,t+m}| \leq 2(2^{1-1/(2+\delta)}+1)\alpha_{Y,m}^{\delta/(2+\delta)}\|Y_t\|_{2+\delta}\|Y_{t+m}\|_{2+\delta}. \tag{A.2}$$

Choosing $m = 0$ in (A.2) gives $Var(Y_t) = \sigma_t^2 < B < \infty$ for all t. Now, consider the following elementary inequality relations

$$\begin{aligned}
E\left(\frac{1}{n-b}S_{n-b}\right)^2 &\leq \frac{1}{(n-b)^2}\sum_{t=1}^{n-b}\sigma_t^2 + \frac{2}{(n-b)^2}\sum_{m=1}^{n-b-1}\sum_{t=1}^{n-b-m}|\sigma_{t,t+m}| \\
&\leq \frac{B}{n-b} + \frac{2}{(n-b)^2}\sum_{m=1}^{n-b-1}(n-b-m)B_m \\
&= \frac{B}{n-b} + \frac{2}{(n-b)^2}\sum_{m=1}^{b}(n-b-m)B_m + \frac{2}{(n-b)^2}\sum_{m=b+1}^{n-b-1}(n-b-m)B_m \\
&\leq \frac{B}{n-b} + \frac{2B}{(n-b)^2}\sum_{m=1}^{b}(n-b-m) + \frac{2}{(n-b)}\sum_{m=b+1}^{n-b-1}B_m \qquad (A.3)
\end{aligned}$$

where $B_m = \sup_t|\sigma_{t,t+m}| \leq B$. The second term in the last inequality of (A.3) is $O(\frac{1}{n^{1-\eta}})$ because $b = O(n^\eta)$. The third term in the last inequality is equivalent to $\frac{n-2b-1}{n-b}\frac{2}{(n-2b-1)}\sum_{m=1}^{n-2b-1}B_{m+b}$. Thus, if $\lim_{n\to\infty}\sum_{m=1}^{n-2b-1}m^{-1}B_{m+b} < \infty$, Kronecker's lemma implies that the third term is $o(1)$. But inequalities (A.1), (A.2) and the given conditions provide

$$\lim_{n\to\infty}\sum_{m=1}^{n-2b-1} m^{-1}B_{m+b} \;\leq\; 2(2^{1-1/(2+\delta)}+1)c^2 \lim_{n\to\infty}\sum_{m=1}^{n-2b-1} m^{-1}\alpha_{X,m}^{\delta/(2+\delta)} < \infty. \tag{A.4}$$

Thus, $E\left(\frac{1}{n-b}S_{n-b}\right)^2 = o(1)$. The conclusion follows from this using the Chebyshev inequality.

If $\{X_t\}$ is a uniform mixing sequence, we obtain the conclusion using the uniform mixing inequality (cf. Davidson, 1994, p.214)

$$|Cov(Y_t, Y_{t+m})| \leq 2\phi_{Y,m}^{1/2}\left\|Y_t\right\|_2\left\|Y_{t+m}\right\|_2,$$

and proceeding as previously. ■

The following lemma proves the strong law of large number for functions of a mixing process with a growing number of arguments. This lemma requires more stringent moment conditions than for Lemma A.1.

Lemma A.2 *Suppose that the same assumptions as in Lemma A.1 hold except that*

$$\sum_{m=1}^{\infty}\alpha_{X,m}^{\delta/(2+\delta)} < \infty \text{ and } \sum_{m=1}^{\infty}\phi_{X,m}^{1/2} < \infty \tag{A.5}$$

replace the corresponding conditions in Lemma A.1. Then, as $n\to\infty$,

$$\frac{1}{n-b}S_{n-b} \xrightarrow{a.s.} 0.$$

Proof. This lemma will be proven by adopting the Borel-Cantelli lemma using subsequences (see, for example, Chung, 1974). For simplicity, let $n-b=l$. Using condition (A.5) for inequality (A.3), we obtain $E\left(\frac{1}{l}S_l\right)^2 = O(\frac{1}{l^{1-\eta}})$ as in the proof of Lemma A.1. Thus, using the notation $\sum_{l^2} a_{l^2} = a_1 + a_4 + a_9 + ...$ and the Chebyshev inequality, we have for $\varepsilon > 0$

$$\sum_{l^2} P\left(|S_{l^2}| > l^2\varepsilon\right) \leq \sum_{l^2} E(\frac{S_{l^2}}{l^2})^2\frac{1}{\varepsilon^2} < \infty,$$

which implies $P\left(|S_{l^2}| > l^2\varepsilon \ i.o.\right) = 0$ by the Borel-Cantelli lemma. Thus,

$$\frac{S_{l^2}}{l^2} \xrightarrow{a.s.} 0 \text{ as } n\to\infty. \tag{A.6}$$

Now, we show that S_k is essentially no different from S_{l^2} when $l^2 \le k < (l+1)^2$. Put $D_l = \max_{l^2 \le k < (l+1)^2} |S_k - S_{l^2}|$. Because

$$\begin{aligned} E(S_k - S_{l^2})^2 &= E\left(\sum_{t=l^2+1}^{k} (Y_t - E(Y_t))\right)^2 \\ &= \sum_{t=l^2+1}^{k} \sigma_t^2 + 2\sum_{m=l^2+1}^{k-1}\sum_{j=m+1}^{k} |\sigma_{m,j}| \\ &\le (k-l^2)B + 2(k-l^2-1)\sum_{i=1}^{\infty} B_i \end{aligned}$$

and because $\sum_{i=1}^{\infty} B_i < \infty$ due to inequalities (A.1), (A.2) and (A.4) without m^{-1}, it follows that

$$E(D_l^2) \le \sum_{k=l^2+1}^{(l+1)^2-1} E(S_k - S_{l^2})^2 \le \sum_{k=l^2+1}^{(l+1)^2-1} \left((k-l^2)B + 2(k-l^2-1)\sum_{i=1}^{\infty} B_i\right) = O(l^2).$$

The Chebyshev inequality gives for $\varepsilon > 0$

$$P\left(D_l > l^2\varepsilon\right) \le \frac{E(D_l^2)}{l^4\varepsilon^2} = O(\frac{1}{l^2}).$$

Using the Borel-Cantelli lemma as above, we obtain

$$\frac{D_l}{l^2} \xrightarrow{a.s.} 0 \text{ as } n \to \infty. \tag{A.7}$$

Since $\frac{|S_k|}{k} \le \frac{|S_{l^2}| + D_l}{l^2}$ for $l^2 \le k < (l+1)^2$, the conclusion follows from (A.6) and (A.7). ■

Proof of Theorem 1: (i) Using Assumption 2, we obtain an inequality

$$\begin{aligned} \left|L_{Nnb}^{\xi}(x) - J_N^{\xi}(x)\right| &\le \left|\frac{1}{n-b+1}\sum_{s=1}^{n-b+1}(1\{\xi_{Nbs} \le x\} - E1\{\xi_{Nb1} \le x\})\right| \\ &\quad + \left|E1\{\xi_{Nb1} \le x\} - J_N^{\xi}(x)\right| \\ &= A_{Nnb}(x) + B_{Nb}(x), \text{ say.} \end{aligned} \tag{A.8}$$

We may write $1\{\xi_{Nbs} \le x\} = f(\mathbf{w}_s - \bar{\mathbf{w}}_{bs}, \mathbf{w}_{s+1} - \bar{\mathbf{w}}_{bs}, ..., \mathbf{w}_{s+b-1} - \bar{\mathbf{w}}_{bs})$. Because of the time series demeaning, $1\{\xi_{Nbs} \le x\}$ is not affected by the initial values of the $I(1)$ regressors and, therefore, is a function of only $\mathbf{r}_s, ..., \mathbf{r}_{s+b-1}$. The indicator function and $\mathbf{r}_t$ satisfy the conditions of Lemma A.2, and hence $A_{Nnb}(x)$ in this inequality converges to zero almost surely. This, owing to the Glivenko-Cantelli lemma (see Davidson, 1994, p.332)[8], leads to $\sup_x A_{Nnb}(x) \to 0$. Because of the relation $E1\{\xi_{Nb1} \leq x\} = J^{\xi}_{Nb}(x)$ and Assumption 1, $B_{Nb}(x)$ in inequality (A.8) also converges to zero uniformly in x (cf. Lemma 3 of Chow and Teicher, 1988, p.265). Thus, the proof for L^{ξ}_{Nnb} is complete. The conclusions for L^{τ}_{Nnb} and L^{φ}_{Nnb} can be proven in the same manner. ■

(ii) Since $J^{a}_{N}(\cdot)$ is continuous at $c^{a}_{Nnb}(1-\lambda)$, this can be shown using the same method as in Beran (1987, p.14). ■

Proof of Theorem 2: (i) Write for all i

$$\begin{aligned} & (\hat{\alpha}_{ibs} - \hat{\alpha}_{in})' R' \Theta_{ibs} R (\hat{\alpha}_{ibs} - \hat{\alpha}_{in}) \\ = \; & (R\hat{\alpha}_{ibs} - r)'\Theta_{ibs}(R\hat{\alpha}_{ibs} - r) - 2(R\hat{\alpha}_{in} - r)'\Theta_{ibs}(R\hat{\alpha}_{ibs} - r) + (R\hat{\alpha}_{in} - r)'\Theta_{ibs}(R\hat{\alpha}_{in} - r) \\ = \; & (R\hat{\alpha}_{ibs} - r)'\Theta_{ibs}(R\hat{\alpha}_{ibs} - r) + \kappa_{i,nbs}, \text{ say.} \end{aligned} \tag{A.9}$$

The mean-value expansion of $\xi^{\bullet}_{Nbs}$ gives

$$\xi^{\bullet}_{Nbs} = \xi_{Nbs} + \sum_{\iota=1}^{N} \frac{\partial g}{\partial x_i}_{|x_i=\nu_{i,nbs}} \kappa_{i,nbs}, \tag{A.10}$$

where ξ_{Nbs} is defined in (8) and $\nu_{i,nbs}$ lies on the line joining $\kappa_{i,nbs}$ and 0 and is $o_p(1)$. Using (A.10), write

$$\begin{aligned} L^{\bullet\xi}_{Nnb}(x) &= \frac{1}{n-b+1} \sum_{s=1}^{n-b+1} 1\{\xi^{\bullet}_{Nbs} \leq x\} \\ &= \frac{1}{n-b+1} \sum_{s=1}^{n-b+1} 1\{\xi_{Nbs} \leq x - \varrho_{nbs}\}, \end{aligned}$$

where $\varrho_{nbs} = \sum_{\iota=1}^{N} \frac{\partial g}{\partial x_i}_{|x_i=\nu_{i,nbs}} \kappa_{i,nbs}$. As in the proof of Theorem 1 of Romano and Wolf (2001), we have an inequality for $\varepsilon > 0$

$$L^{\xi}_{Nnb}(x-\varepsilon)1\{E_n\} \leq L^{\bullet\xi}_{Nnb}(x)1\{E_n\} \leq L^{\xi}_{Nnb}(x+\varepsilon), \tag{A.11}$$

where $1\{E_n\}$ is the indicator of the event $\{|\, \varrho_{nbs} \,| \leq \varepsilon\}$. Because $\varrho_{nbs} \xrightarrow{p} 0$ for all s due to Assumptions 4 and 5, the event E_n holds with probability approaching one as n goes to infinity. Thus, (A.11) implies that the relation

$$L^{\xi}_{Nnb}(x-\varepsilon) \leq L^{\bullet\xi}_{Nnb}(x) \leq L^{\xi}_{Nnb}(x+\varepsilon) \tag{A.12}$$

holds with probability tending to one. Using Lemma A.1 and the same method as in the proof of Theorem 1, we obtain $L^{\xi}_{Nnb}(x) \xrightarrow{p} J^{\xi}_{N}(x)$. Because x is a continuity

[8] Davidson states the Glivenko-Cantelli lemma for iid random variables. But it is easily verified that it can be used for dependent random variables as long as the strong law of large numbers for them holds.

point, it also follows that $L^{\xi}_{Nnb}(x \pm \varepsilon) \xrightarrow{p} J^{\xi}_{N}(x)$. This and (A.12) imply the result for $L^{\bullet\xi}_{Nnb}$. The conclusion for $L^{\bullet\tau}_{Nnb}$ follows in the same way. ■

(ii) If $J^{\xi}_{N}(x)$ is continuous, the uniform convergence follows from the pointwise convergence of part (i). See Lemma 3 of Chow and Teicher (1988, p.265). The conclusion for $L^{\bullet\tau}_{Nnb}$ follows in the same way. ■

(iii) The proof of this result is identical with that of part (ii) of Theorem 1. ■

Proof of Theorem 3: (i) Suppose without loss of generality that the null hypothesis (3) is rejected when ξ_{Nn} is less than a critical value from its limiting distribution. Since $\xi_{Nn}/n^{\epsilon_\xi} \xrightarrow{p} \xi_N$, $c^{\xi}_{Nnb}(1-\lambda)/b^{\epsilon_\xi} = O_p(1)$ and $\frac{b}{n} \to 0$ as $n \to \infty$,

$$\begin{aligned} P[\xi_{Nn} &< c^{\xi}_{Nnb}(1-\lambda)] \\ &= P[\xi_{Nn}/n^{\epsilon_\xi} < c^{\xi}_{Nnb}(1-\lambda)/b^{\epsilon_\xi} \times \left(\frac{b}{n}\right)^{\epsilon_\xi}] \\ &= E1\left\{\xi_{Nn}/n^{\epsilon_\xi} < c^{\xi}_{Nnb}(1-\lambda)/b^{\epsilon_\xi} \times \left(\frac{b}{n}\right)^{\epsilon_\xi}\right\} \\ &\to E1\left\{\xi_N < 0\right\} = P[\xi_N < 0] = 1 \end{aligned}$$

by the dominated convergence theorem. The same relation holds for τ_{Nn} and φ_{Nn}. These prove the consistency of the tests using critical values from the non-centered subsamplings. ■

(ii) Using relation (A.9) with r replaced by r^*, we find that $\xi^{\bullet}_{Nbs}$ and $\tau^{\bullet}_{Nbs}$ are $O_p(1)$ under the alternative hypothesis. Thus, $c^{\bullet a}_{Nnb}(1-\lambda) = O_p(1)$, and it follows that

$$\begin{aligned} P[a_{Nn} &< c^{\bullet a}_{Nnb}(1-\lambda)] \\ &= P[a_{Nn}/n^{\epsilon_a} < c^{\xi}_{Nnb}(1-\lambda) \times \left(\frac{1}{n}\right)^{\epsilon_a}] \\ &= E1\left\{\xi_{Nn}/n^{\epsilon_a} < c^{\xi}_{Nnb}(1-\lambda) \times \left(\frac{1}{n}\right)^{\epsilon_a}\right\} \\ &\to E1\left\{\xi_N < 0\right\} = 1 \end{aligned}$$

as desired. ■

Proof of Corollary 4: (i) Assumption 1 holds due to relation (22). Let $[\Delta y_{1t}, ..., \Delta y_{Nt}]' = \mathfrak{y}_t$ Then, t_{Nn} is a function of $\mathfrak{y}_2, ..., \mathfrak{y}_n$, and t_{Nnbs} is a function of $\mathfrak{y}_{s-p}, ..., \mathfrak{y}_{s+b-1}$ ($s \geq p+2$). Since $\mathfrak{y}_t$ is strictly stationary under the null hypothesis, t_{Nbs} has the same distribution for each fixed N and b. This implies Assumption 2. Under Assumption 8, Assumptions 3 and 6 are obviously satisfied. Since $\hat{\rho}^{p}_{Nn} = O_p(n^{-1})$, $\hat{\rho}^{p}_{Nbs} = O_p(b^{-1})$ and $\frac{1}{\sqrt{\hat{\sigma}^2_{Nbs}\left(\sum_{i=1}^{N}\mathbf{y}'_{i,-1,bs}\mathbf{M}_{Q_{ibs}}\mathbf{y}_{i,-1,bs}\right)^{-1}}} = O_p(b)$, Assumption 4 holds. Assumption 5 is obviously satisfied. Thus, the results for τ_{Nbs} and $\tau^{\bullet}_{Nbs}$ follow. ■

(ii) Assumption 7 holds because

$$t_{Nn}/\sqrt{n} = \frac{\hat{\rho}^p_{Nn}-\rho^*}{\sqrt{\hat{\sigma}^2_{Nn}\left(\sum_{i=1}^N \mathbf{y}'_{i,-1,n}\mathbf{M}_{Q_{in}}\mathbf{y}_{i,-1,n}\right)^{-1}}}/\sqrt{n}+\frac{\rho^*}{\sqrt{\hat{\sigma}^2_{Nn}\left(\sum_{i=1}^N \mathbf{y}'_{i,-1,n}\mathbf{M}_{Q_{in}}\mathbf{y}_{i,-1,n}/n\right)^{-1}}}$$
$$\xrightarrow{p} p\lim_{n\to\infty}\frac{\rho^*}{\sqrt{\hat{\sigma}^2_{Nn}\left(\sum_{i=1}^N \mathbf{y}'_{i,-1,n}\mathbf{M}_{Q_{in}}\mathbf{y}_{i,-1,n}/n\right)^{-1}}} \quad \text{(A.13)}$$

as $n\to\infty$ with $\rho^*<0$ and because the probability limit in relation (A.13) is negative almost surely. Since $\hat{\rho}^p_{Nn}-\rho^*=O_p(1/\sqrt{n})$, $\hat{\rho}^p_{Nbs}-\rho^*=O_p(1/\sqrt{b})$ and $\frac{1}{\sqrt{\hat{\sigma}^2_{Nbs}\left(\sum_{i=1}^N \mathbf{y}'_{i,-1,bs}\mathbf{M}_{Q_{ibs}}\mathbf{y}_{i,-1,bs}\right)^{-1}}}=O_p(\sqrt{b})$, part (ii) of Assumption 4 holds with $r^*=\rho^*$and other relevant changes. ■

Proof of Corollary 5: (i) These can be shown using the same arguments as in the proof of part (i) of Corollary 4. The only notable differences are $\hat{\rho}_{in}=O_p(n^{-1})$, $\hat{\rho}_{ibs}=O_p(b^{-1})$ and $\frac{1}{\sqrt{\hat{\sigma}^2_{ibs}\left(\mathbf{y}'_{i,-1,bs}\mathbf{M}_{Q_{i,bs}}\mathbf{y}_{i,-1,bs}\right)^{-1}}}=O_p(b)$, which imply Assumption 4. ■

(ii) Suppose without loss of generality that the null is violated for $i=1,2,...,N_1$. Then,

$$\bar{t}_{Nn}/\sqrt{n}=\frac{1}{N}\sum_{i=1}^{N_1}t_{in}/\sqrt{n}+o_p(1),$$

and the probability limit of $\frac{1}{N}\sum_{i=1}^{N_1}t_{in}(p_i,\varphi_i)/\sqrt{n}$ is negative almost surely as in the proof of Corollary 4. Thus, Assumption 7 holds. Part (i) of Assumption 4 also holds as in the proof of Corollary 4. ■

Proof of Corollary 6: (i) Assumptions 1, 2, 3, 4 and 6 hold as in the proof of part (i) of Corollary 4. To show that Assumption 5 is satisfied, write

$$\frac{\partial g}{\partial x_i}_{|x_i=f_{in}}=\frac{1}{\sqrt{N}}\left(\frac{\partial \Phi}{\partial y_i}_{|y_i=F(f_{in})}\right)^{-1}\frac{\partial F}{\partial x_i}_{|x_i=f_{in}},$$

where $f_{in}=o_p(1)$. Because $0<\frac{\partial\Phi}{\partial y_i}_{|y_i=F(f_{in})}<\infty$ almost surely and $\frac{\partial F}{\partial x_i}_{|x_i=f_{in}}<\infty$ almost surely, we have $\frac{\partial g}{\partial dx_i}_{|x_i=f_{in}}=O_p(1)$ as required. The exact formula of the function $F(\cdot)$ is unknown, but it is continuous. Thus, it can be expressed as the sum of absolutely continuous and singular parts. The derivative of the former part exists and is finite, and that of the latter also exists and equals zero. Using the notation $\frac{\partial F}{\partial x_i}$ is, therefore, justifiable. ■

(ii) A mean value expansion of $\Phi^{-1}((F(dfg^{\mu}_{in}))$ gives

$$\begin{aligned}
& \Phi^{-1}((F(dfg^{\mu}_{in}))\,/\sqrt{n} \\
= \; & \Phi^{-1}((F(0))\,/\sqrt{n} + \frac{\partial \Phi^{-1}}{\partial y_i}_{|y_i=F(v_{in})} \frac{\partial F}{\partial x_i}_{|x_i=v_{in}} dfg^{\mu}_{in}/\sqrt{n} \\
= \; & \left(\frac{\partial \Phi}{\partial y_i}_{|y_i=F(v_{in})} \right)^{-1} \frac{\partial F}{\partial x_i}_{|x_i=v_{in}} dfg^{\mu}_{in}/\sqrt{n} + o(1), \qquad \text{(A.14)}
\end{aligned}$$

where v_{in} lies on the line connecting 0 and dfg^{μ}_{in}. Now suppose without loss of generality that the null is violated for $i = 1, 2, ..., N_1$. Then, using relation (A.14), we obtain

$$\begin{aligned}
Z_{Nn}/\sqrt{n} &= \frac{1}{\sqrt{N}} \sum_{i=1}^{N_1} \Phi^{-1}((F(dfg^{\mu}_{in}))\,/\sqrt{n} + o_p(1) \\
&= \frac{1}{\sqrt{N}} \sum_{i=1}^{N_1} \left(\frac{\partial \Phi}{\partial y_i}_{|y_i=F(v_{in})} \right)^{-1} \frac{\partial F}{\partial x_i}_{|x_i=v_{in}} dfg^{\mu}_{in}/\sqrt{n} + o_p(1).
\end{aligned}$$

Because $\left(\frac{\partial \Phi}{\partial y_i}_{|y_i=F(v_{in})} \right)^{-1} > 0$ and $\frac{\partial F}{\partial x_i}_{|x_i=v_{in}} > 0$ with probability one and because the probability limit of $dfg^{\mu}_{in}/\sqrt{n}$ is a negative constant, Assumption 7 holds. Part (i) of Assumption 4 also holds as in the proof of Corollary 4 because dfg^{μ}_{in} is consistent against the alternative of stationarity in the form of $\rho_i < 0$. ■

Proof of Corollary 7: (i) Assumption 1 holds due to relation (31). Assumption 9 implies Assumptions 2 and 3. ■

(ii) Assumption 7 holds due to Assumption 10. ■

References

Bai, J. and S. Ng (2002): "A Panic Attack on Unit Roots and Cointegration," Mimeo, Boston College.

Baltagi, B. and C. Kao (2000): "Nonstationary Panels, Cointegration in Panels and Dynamic Panels: A Survey," *Advances in Econometrics,* 15, 7–51.

Banerjee, A. (1999): "Panel Data Unit Roots and Cointegration: An Overview," *Oxford Bulletin of Economics and Statistics,* 61, 607–629.

Beran, R. (1987): "Bootstrap Methods in Statistics," *Jaresberichte des Deutschen Mathematischen Vereins,* 86, 14-30.

Chang, Y. (2003): "Bootstrap Unit Root Tests in Panels with Cross-Sectional Dependency," *Journal of Econometrics*, forthcoming..

Choi, I. (2001): "Unit Root Tests for Panel Data," *Journal of International Money and Finance,* 20, 249–272.

Choi, I. (2002a): "Instrumental Variables Estimation of a Nearly Nonstationary, Heterogeneous Error Component Model," *Journal of Econometrics,* 109, 1–32.

Choi, I. (2002b): "Combination Unit Root Tests for Cross-Sectionally Correlated Panels," http://ihome.ust.hk/~inchoi.

Choi, I. (2004): "Subsampling Vector Autoregressive Tests of Linear Constraints," *Journal of Econometrics,* forthcoming.

Choi, I. and B.C. Ahn (1995): "Testing for Cointegration in a System of Equations," *Econometric Theory,* 11, 952–983.

Chow, Y.S. and H. Teicher (1988): *Probability Theory: Independence, Interchangeability and Martingales*, Springer–Verlag, New York.

Chung, K. L. (2001): *A Course in Probability Theory*, 3rd edition, Academic Press, New York.

Davidson, J. (1994): *Stochastic Limit Theory,* Oxford University Press, Oxford.

Delgado, M., J.M. Rodríguez-Poo and M. Wolf (2001): "Subsampling inference in cube root asymptotics with an Application to Manski's maximum score estimator. *Economics Letters,* 73, 241-250.

Elliott, G., T.J. Rothenberg J.H. Stock. (1996): "Efficient Tests for an Autoregressive Unit Root," *Econometrica,* 64, 813–836.

Gonzalo, J. and M. Wolf (2002): "Subsampling inference in threshold autoregressive models," Mimeo, Universitat Pompeu Fabra.

Groen, J.J.J. and F.R. Kleibergen (2003): "Likelihood–based Cointegration Analysis in Panels of Vector Error Correction Models," *Journal of Business and Economic Statistics*, 21, 295-318.

Hadri, K. (2000): "Testing for Stationarity in Heterogeneous Panel Data," *Econometrics Journal,* 3, 148–161.

Hedges, L.V. and Olkin, I. (1985): *Statistical Methods for Meta-Analysis*, San Diego, Academic Press.

Im, K. S., M.H. Pesaran and S. Shin (2003): "Testing for Unit Roots in Heterogeneous Panels," *Journal of Econometrics*, 115, 53-74.

Kao, C. (1999): "Spurious Regression and Residual-Based Tests for Cointegration in Panel Data," *Journal of Econometrics,* 90, 1–44.

Kao, C. and M.-H. Chiang (2000): "On the Estimation and Inference of a Cointegrated Regression in Panel Data," in B. Baltagi (eds.) *Advances in Econometrics,* 15, 179-222.

Kauppi, H. (2000): "Panel Data Limit Theory and Asymptotic Analysis of a Panel Regression with Near Integrated Regressors," in B. Baltagi (eds.) *Advances in Econometrics,* 15, 239–274.

Kwiatkowski, D., P.C.B. Phillips, P. Schmidt and Y. Shin (1992): "Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root: How Sure Are We that Economic Time Series Have a Unit Root?" *Journal of Econometrics,* 54, 159–178.

Larsson, R., J. Lyhagen and M. Löthgren (2001): "Likelihood-based Cointegration Tests in Heterogeneous Panels," *Econometrics Journal,* 4, 109–147.

Levin, A., C.-F. Lin and C.-S.J. Chu (2002): "Unit Root Tests in Panel Data: Asymptotic and Finite-Sample Properties," *Journal of Econometrics,* 108, 1-25.

Maddala, G.S. and S. Wu (1999): "A Comparative Study of Unit Root Tests with Panel Data and a New Simple Test," *Oxford Bulletin of Economics and Statistics,* 61, 631–652.

McCosky, S. and C. Kao (1998): "A Residual-Based Test of the Null of Cointegration in Panel Data," *Econometric Review,* 17, 57–84.

Moon, H.R. and B. Perron (2001): "Testing for a Unit Root in Panels with Dynamic Factors," Mimeo USC.

Pedroni, P. (1995): "Panel Cointegration: Asymptotic and Finite Sample Properties of Pooled Time Series With an Application to the PPP Hypothesis," Mimeo, Department of Economics, Indiana University.

Pedroni, P. (2000): "Fully modified OLS for heterogeneous cointegrated panels and the case of purchasing power parity," in B. Baltagi (eds.) *Advances in Econometrics*, 15, 93-130.

Politis, D.N., J. P. Romano and M. Wolf (1999): *Subsampling,* Springer-Verlag, New York.

O'Connell, P.G.J. (1998): "The Overvaluation of Purchasing Power Parity," *Journal of International Economics*, 44, 1–19.

Phillips, P.C.B. and H. Moon (1999): "Linear Regression Limit Theory for Nonstationary Panel Data," *Econometrica,* 67, 1057–1112.

Phillips, P.C.B. and H. Moon (2000): "Nonstationary Panel Data Analysis: An Overview of Some Recent Developments," *Econometric Review,* 19, 263–286.

Phillips, P.C.B. and D. Sul. (2003), “Dynamic Panel Estimation and Homogeneity Testing Under Cross Section Dependence,” *Econometrics Journal*, 6, 217-259.

Romano, J.P. and M. Wolf (1999):. Inference for the Mean in the Heavy-Tailed Case. *Metrika,* 50, 55-69.

Romano, J.P. and M. Wolf (2001): Subsampling Intervals in Autoregressive Models with Linear Trend. *Econometrica,* 69, 1283-1314.

Sherman, M. and E. Carlstein (1996): “Replicate Histograms,” *Journal of the American Statistical Association,* 91, 566-576.

Taylor, M. P. and Sarno, L. (1998): “The Behavior of Real Exchange Rates during the Post-Bretton Woods Period,” *Journal of International Economics*, 46, 281–312.

Wu, C.F.J. (1990): “On the Asymptotic Properties of the Jackknife Histogram,” *Annals of Statistics,* 18, 1438-1452.

Yin, Y. and S. Wu (2000): “Stationarity Tests in Heterogeneous Panels,” in B. Baltagi (eds.) *Advances in Econometrics,* 15, 275–296.