On the Power and Interpretation of Panel Unit Root Tests

Sune Karlsson and Mickael Löthgren

Department of Economic Statistics
Stockholm School of Economics
P.O. Box 6501, SE-113 83 Stockholm
Sweden
e-mail: stsk@hhs.se, stml@hhs.se
fax:+46-8-348161

Stockholm School of Economics
SSE/EFI Working Paper Series in Economics and Finance
No. 299

February 1999

Abstract

We demonstrate that panel unit root tests can have high power when a small fraction of the series are stationary and may lack power when a large fraction is stationary. The acceptance or rejection of the null is thus not sufficient evidence to conclude that all series have a unit root or that all are stationary.
Keywords: Dynamic panels; Monte Carlo; Heterogeneity.
JEL Classification: C12; C15; C22; C23.

1 Introduction

The univariate Dickey-Fuller (DF) and Augmented Dickey-Fuller (ADF) tests for stationarity have been extended to panel tests for stationarity under models with various degrees of heterogeneity by, e.g., Levin and Lin (1992) and Im, Pesaran and Shin (1997). One main motivation for the development and application of panel unit root tests is that the power of the tests increase with an increase in the number of panel series as compared with the well-known low power of the standard DF and ADF unit root tests against near unit root alternatives for small samples. The proposed panel tests have found use in various studies, mainly focusing on tests of the existence of purchasing power parity. See, e.g., MacDonald (1996), Oh (1996), Wu (1996), Coakley and Fuertes (1997) and Papell (1997). Other studies include Culver and Papell (1997) who apply panel unit root tests to the inflation rate, Song and Wu (1998) investigate hysteresis in unemployment and McCoskey and Selden (1998) test for units roots in health care expenditure and GDP.

The panel tests differ in the degree of heterogeneity between individuals that is allowed for under the various approaches. The model considered by Quah (1994) is the most restrictive and it is assumed that all individuals in the panel have a common AR parameter and a common intercept. The model by Levin and Lin (1992, LL) allows for individual specific intercepts, but retain the assumption of the common AR parameter. The tests proposed by Quah and LL are based on pooled estimates using least squares and thus implicitly also assumes a common residual variance for the individuals.

The panel model by Im, Pesaran and Shin (1997, IPS) allows complete heterogeneity between individuals. In the IPS setup ADF equation is estimated separately for each individual and thus allows for differing parameter values, residual variance and even different lag lengths. The test statistics are then the average across individuals of the LM unit root test (LM-bar) and the usual Dickey-Fuller τ statistic (t-bar). IPS also advocates studentizing the statistics to obtain statistics ($\Gamma_{\overline{LM}}$ and $\Gamma_{\overline{t}}$) that are asymptotically distributed as $N(0,1)$.

One primary motivation for the use of the panel versions of the unit root tests is that the additional cross-sectional dimension in the panel lead to better power properties of the panel tests as compared with the low power of the standard individual-specific ADF test against near unit root alternatives for small samples.

The null hypothesis in all panel unit root tests is that each series in the panel contains a unit root, and thus is difference stationary. The alternative hypothesis is somewhat more ambiguously specified. In the earlier papers (Quah and LL) the alternative is that *all* individual series in the panel are

stationary. In IPS, the alternative is that at least one of the individual series in the panel is stationary. IPS show that the proposed LM-bar test is consistent if the limiting fraction of stationary series is non-zero as $N \longrightarrow \infty$. They do, however, not investigate the small sample power of the tests when only a fraction of the series are stationary. It is also reasonable to expect the LL tests to have some power against alternatives where a fraction of the series are stationary.

The presence or absence of power against alternatives where a subset of the series are stationary has serious implications for empirical work. If the tests have high power, a rejection of the unit root null can be driven by a few stationary series and the whole panel may erroneously be modelled as stationary. If, on the other hand, the tests have low power it may erroneously be concluded that the panel contain a common unit root even if a majority of the series are stationary.

To shed some light on these issues we investigate the small-sample power properties of the LL and IPS panel unit root LM-bar and t-bar tests by Monte Carlo simulations. Specifically, the power of the tests are investigated under data generating processes where the fraction of stationary series in the panel are varied in the $0 - 100\%$ interval (0% gives the size of the tests whereas a fraction strictly greater than 0% gives the power of the test).

The paper unfolds as follows: Section 2 gives a brief presentation of the dynamic panel data models and the panel unit root tests considered in the paper. Section 3 presents the design and the results of the conducted Monte Carlo simulations study.

2 The panel tests

2.1 The Levin and Lin tests

The LL model is given by

$$\Delta y_{it} = \mu_i + \beta y_{i,t-1} + \sum_{k=1}^{p} \phi_k \Delta y_{i,t-k} + \gamma t + \varepsilon_{it}, i = 1, ..., N; t = 1, ..., T. \quad (1)$$

Corresponding to the maintained hypothesis of common dynamics the null hypothesis and alternative hypothesis are given by

$$H_0 : \beta = 0, H_1 : \beta < 0.$$

The fixed-effect model (1) is easily estimated using the within estimator or the LSDV estimator and the LL test statistic is based on the usual t-statistic

$$t_\beta = \frac{\widehat{\beta}}{\widehat{\sigma}_{\widehat{\beta}}}. \tag{2}$$

LL propose two transformations of the t-statistic t_β that are asymptotically normally distributed as N and $T \to \infty$:

1. $LL_1 = \sqrt{1.25} t_\beta + \sqrt{1.875N} \Rightarrow N(0,1)$ (as $\sqrt{N}/T \to 0$); and
2. $LL_2 = \sqrt{1.25}\left[t_\beta - \frac{\sqrt{N}\mu_{1T}}{\sqrt{\mu_{2T}}}\right] \Rightarrow N(0,1)$, with $\mu_{1T} = -\frac{1}{2} - \frac{1}{2}T^{-1}$ and $\mu_{2T} = \frac{1}{6} + \frac{5}{6}T^{-2}$.

Note that since $\mu_{1T} = -\frac{1}{2} + o(T)$ and $\mu_{2T} = \frac{1}{6} + o(T)$ the two tests LL_1 and LL_2 coincide asymptotically, but in finite samples they will differ.

2.2 The Im, Pesaran and Shin model

The heterogeneous panel data model of IPS is given by

$$\Delta y_{it} = \mu_i + \beta_i y_{i,t-1} + \sum_{k=1}^{p_i} \phi_{i,k} \Delta y_{i,t-k} + \gamma_i t + \varepsilon_{it}, i = 1, ..., N; t = 1, ..., T. \tag{3}$$

The maintained hypothesis of common dynamics is relaxed and the relevant hypotheses are

$$H_0 : \beta_i = 0, \ \forall \ i, \ H_1 : \exists \ i \text{ s.t. } \beta_i < 0.$$

Due to the heterogeneity each equation is estimated separately by OLS and the test statistics are obtained as (studentized) averages of the test statistics for each equation.

2.2.1 The *t*-bar statistic

The IPS t-bar statistic for is simply defined as the average of the individual Dickey-Fuller τ statistics as

$$\overline{t} = \frac{1}{N}\sum_{i=1}^{N} \tau_i, \ \tau_i = \frac{\widehat{\beta}_i}{\widehat{\sigma}_{\widehat{\beta}_i}} \tag{4}$$

Assuming that the cross-sections are independent, IPS propose to use the following standardized t-bar statistic

$$\Gamma_{\overline{t}} = \frac{\sqrt{N}\left(\overline{t} - E\left(\tau_i \left| \beta_i = 0\right.\right)\right)}{\sqrt{Var\left(\tau_i \left| \beta_i = 0\right.\right)}}. \tag{5}$$

The means $E\left(\tau_i \left| \beta_i = 0\right.\right)$ and variances $Var\left(\tau_i \left| \beta_i = 0\right.\right)$ are obtained from Monte Carlo simulations. IPS conjecture that the standardized t-bar statistic $\Gamma_{\overline{t}}$ converge weakly to a standard normal distribution as N and $T \to \infty$. Hence the panel data unit root inference can be conducted by comparing the obtained $\Gamma_{\overline{t}}$ statistic to critical values from the lower tail of the $N\left(0,1\right)$ distribution.

2.2.2 The LM-bar statistic

The individual unit root LM-statistic LM_i for testing the hypothesis that $\beta_i = 0$ for series i is defined as

$$LM_i = \frac{T\Delta\mathbf{y}_i'\mathbf{P}_i\Delta\mathbf{y}_i}{\Delta\mathbf{y}_i'\mathbf{M}_i\Delta\mathbf{y}_i}, \tag{6}$$

where the projection matrices $\mathbf{M}_i = \mathbf{I} - \mathbf{Q}_i\left(\mathbf{Q}_i'\mathbf{Q}_i\right)^{-1}\mathbf{Q}_i'$ and $\mathbf{P}_i = \mathbf{M}_i\mathbf{y}_{i,-1}$ $\left(\mathbf{y}_{i,-1}'\mathbf{M}_i\mathbf{y}_{i,-1}\right)^{-1}\mathbf{y}_{i,-1}'\mathbf{M}_i$ are $T \times T$ idempotent matrices of rank $T - p_i - 2$ and 1, respectively. $\mathbf{Q}_i = \left(\boldsymbol{\iota}, \mathbf{t}, \Delta\mathbf{y}_{i,-1}, \ldots, \Delta\mathbf{y}_{i,-p_i}\right)$, $\boldsymbol{\iota}$ is a vector of ones, $\mathbf{t}$ is the time trend, $\Delta\mathbf{y}_i = \left(\Delta y_{i1}, \Delta y_{i2}, ..., \Delta y_{iT}\right)'$ and $\mathbf{y}_{i,-1} = \left(y_{i0}, y_{i1}, ..., y_{i,T-1}\right)'$.

The IPS LM-bar panel unit root statistic is defined as the average of the individual LM statistics as

$$\overline{LM} = \frac{1}{N}\sum_{i=1}^{N} LM. \tag{7}$$

Assuming that the cross-sections are independent, IPS propose to use the following standardized LM-bar statistic

$$\Gamma_{\overline{LM}} = \frac{\sqrt{N}\left(\overline{LM} - E\left(LM_i \left| \beta_i = 0\right.\right)\right)}{\sqrt{Var\left(LM_i \left| \beta_i = 0\right.\right)}}. \tag{8}$$

$E\left(LM_i \left| \beta_i = 0\right.\right)$ and $Var\left(LM_i \left| \beta_i = 0\right.\right)$ are obtained from Monte Carlo simulations and IPS conjecture that the standardized LM-bar statistic $\Gamma_{\overline{LM}}$ converge weakly to a standard normal distribution as N and $T \to \infty$. Hence the panel data unit root inference can be conducted by comparing the obtained $\Gamma_{\overline{LM}}$ statistic to critical values from the upper tail of the $N\left(0,1\right)$ distribution.

For computational reasons the τ_i and LM_i statistics are computed using partitioned regression. Let $\widehat{\mathbf{u}}_i$ be the residuals from a regression of $\Delta\mathbf{y}_i$ on $\mathbf{Q}_i$ and $\widehat{\mathbf{e}}_i$ the residuals from a regression of $\mathbf{y}_{i,-1}$ on $\mathbf{Q}_i$. The LM_i statistic is then obtained as $LM_i = TR_i^2$, where R_i^2 is the R^2 from a regression of $\widehat{\mathbf{u}}_i$ on $\widehat{\mathbf{e}}_i$, and τ_i is obtained as the t-statistic for the coefficient on $\widehat{\mathbf{e}}_i$.

3 The Simulation Study

The underlying data generating process in the simulations is given by:[1]

$$y_{it} = \mu_i (1 - \rho_i) + \rho_i y_{i,t-1} + \varepsilon_{it},\ t = -50, -49, ..., T,\ i = 1, ..., N, \tag{9}$$

with $\mu_i \sim N(0,1)$, $\varepsilon_{it} \sim N(0, \sigma_i^2)$ where $\sigma_i^2 \sim U[0.5, 1.5]$. A subset of the equations are stationary and the remainder have unit roots. More specifically, $\rho_i = \rho^s < 1,\ i = 1, \ldots, N_s = \delta N$ and $\rho_i = 1,\ i = N_s + 1, \ldots, N$. μ_i and σ_i^2 are generated once and then fixed in all replication for a given proportion δ. The 50 initial observations are discarded to remove dependence on initial observations. The $\Gamma_{\overline{LM}}$ and $\Gamma_{\overline{t}}$ statistics are based on the simple Dickey-Fuller regressions

$$\Delta y_{it} = a_i + b_i y_{i,t-1} + e_{it},\ t = 1, ..., T; i = 1, ..., N. \tag{10}$$

We consider all combinations of $N \in \{5,\ 10,\ 25,\ 50,\ 100,\ 200\}$, $T \in \{5,\ 10,\ 25,\ 50,\ 100,\ 200\}$, and $\rho^s \in \{0.9,\ 0.7,\ 0.5\}$ for the proportion of stationary series $\delta \in \{0.0,\ 0.1,\ ...,\ 1.0\}$. For each combination of N, T, δ and ρ^s the number of replication are set to 10000.[2] The moments needed for the calculations of the standardized t-bar and LM-bar tests, i.e., $E(\tau_i | \beta_i = 0)$, $Var(\tau_i | \beta_i = 0)$, $E(LM_i | \beta_i = 0)$ and $Var(LM_i | \beta_i = 0)$, are obtained using stochastic simulations based on 250000 replicates.

3.1 Results

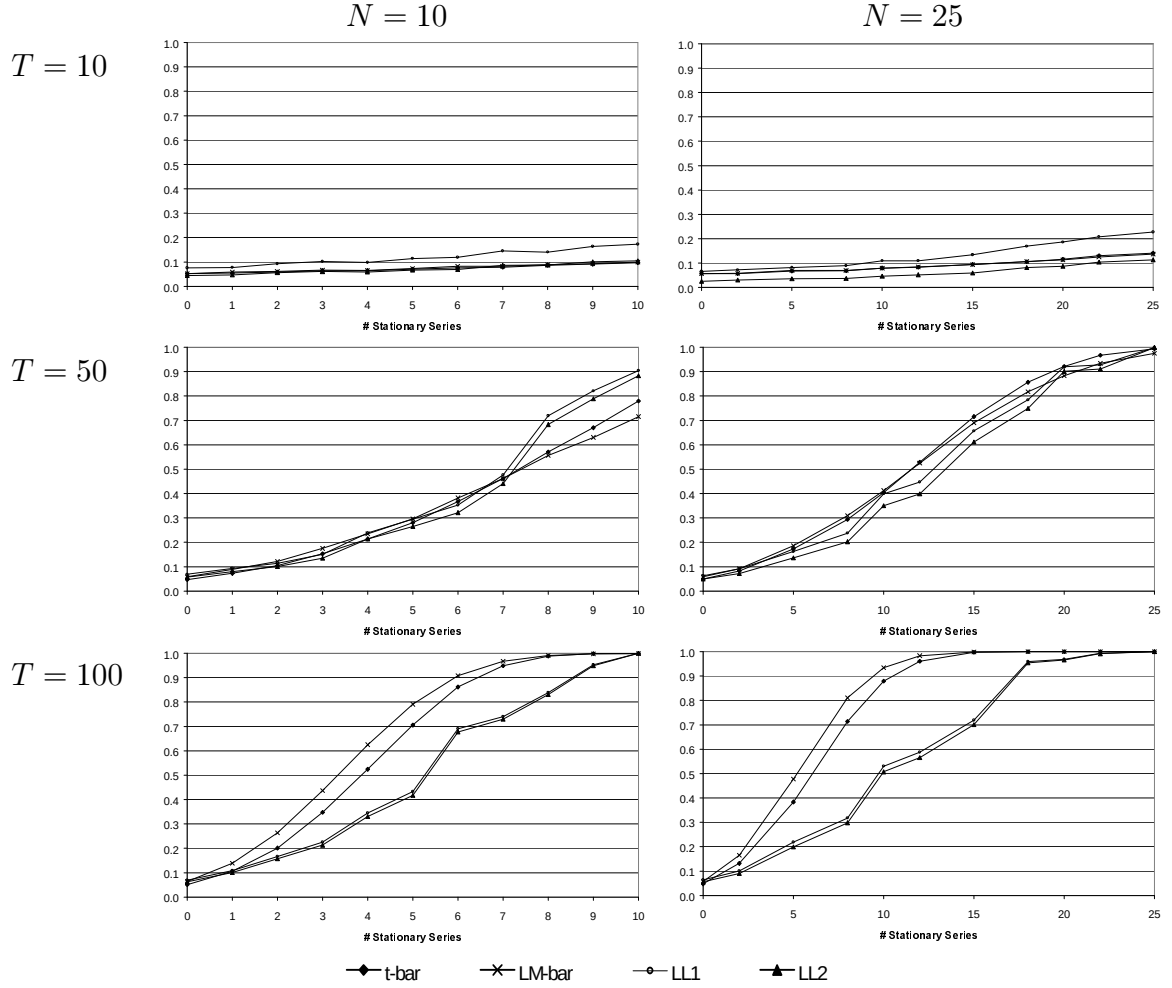
For the sake of brevity, we only present results for $\rho^s = 0.9$ and a subset of sample sizes.[3] As seen from the results in Figure 1 and 2 the power of the panel tests for panels with small time-series dimension ($T = 10$) is fairly low, even as the number of series N in the panel increases. For panels

[1] This is, with the exception that we allow ρ_i to take different values, the same experimental design used for the no serial correlation case in Im, Pesaran and Shin (1997), Table 5.

[2] Based on this number of replicates 95% confidence intervals of the simulated size and power of the tests have a maximum length of $\pm 1.96\sqrt{0.5^2/10000} = \pm 0.0098$.

[3] The complete set of results can be obtained from the authors upon request.

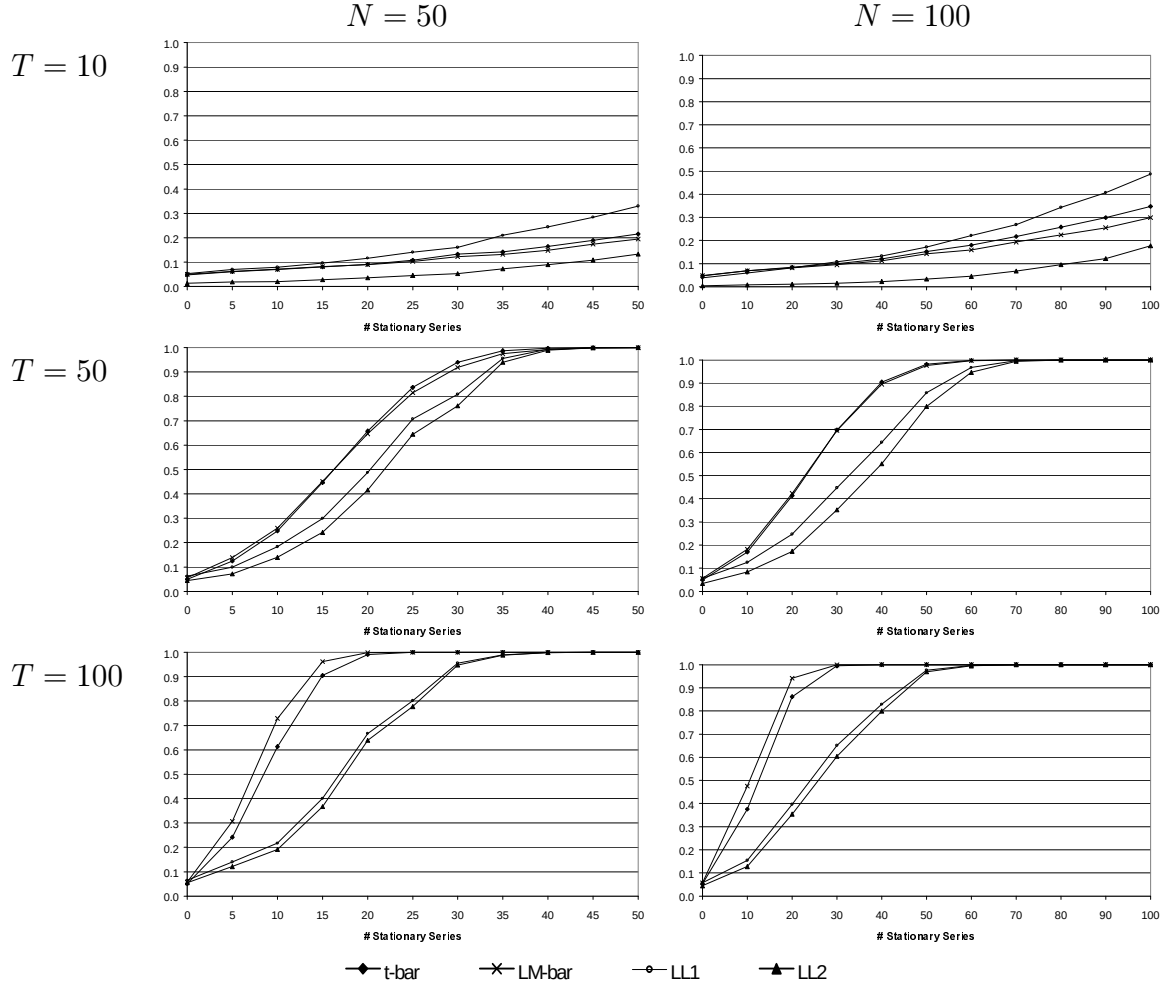
Figure 1: Power of $\Gamma_{\overline{LM}}$, $\Gamma_{\overline{t}}$, LL_1 and LL_2 tests for $\rho^s = 0.9$



with all series stationary, the power is about 0.10, for $N = 10$ and about $0.20 - 0.48$ for $N = 100$ and the power is even lower when only a fraction of the series are stationary. On the other hand, given the cross-sectional dimension N, the power increase dramatically for all tests as the time-series dimension increases. For instance, for $N = 10$, the power increases from about $0.10 - 0.18$ to unity as T increase from 10 to 100. The power increase resulting from increased time-series dimension is even more pronounced for larger N and already for a fairly small proportion of stationary series in the panel the power approaches unity as T increases.

For commonly encountered panel sizes in empirical applications, such as panels of OECD countries ($N = 25$, $T = 50$), the power of the tests are fairly

Figure 2: Power of $\Gamma_{\overline{LM}}$, $\Gamma_{\overline{t}}$, LL_1 and LL_2 tests for $\rho^s = 0.9$



high. Specifically, the power is about 0.40 with 10 stationary series in the panel and about 0.90 with 20 stationary series.[4]

From the results it appears that the IPS-tests outperform the LL-tests for large-T panels (with $T \geq 100$). It also appears that the larger the cross-section, the larger is the power advantage of the IPS-tests. For small-T panels, the LL-tests exhibit size-distortions. For $T = 10$, $N = 10$, the LL_1-test is over-sized and LL_2 is under-sized for $T = 10$ and $N \geq 25$. With $T = 10$ and $N \geq 25$, LL_1 has correct size and is the most powerful test.

Taken together, our main findings are that the power increases monotonically with: (1) an increased number N of series in the panel; (2) an increased

[4]The results for $\rho^s = 0.7$ show a power of near unity for this range of stationary series.

time-series dimension T in each individual series; and (3) an increased proportion δ of stationary series in the panel. Furthermore, as expected, the results indicate that, for a given proportion δ of stationary series in the panel, the power increase due to an increase in the time-series dimension is larger than the corresponding increase in N. The not displayed results for AR parameters ρ^s smaller than 0.9 for the stationary series show that a larger deviation from a unit root for the stationary series increase the power of the test.

To conclude, the results imply the following two major implications for applied work: For large-T panels, there is potential risk that the whole panel may erroneously be modelled as stationary. This is due to the high power of the panel tests for small proportions of stationary series in the panel. For small-T panels, on the other hand, there is a potential risk that the whole panel may be erroneously modelled as non-stationary, due to the relatively low power of the tests even for large proportions of stationary series in the panel. In other words, researchers should be cautious and not impose stationary or non-stationary homogeneity properties of the panel cross-sections solely based on panel unit root test results. Instead, a careful joint analysis of both the individual and the panel unit root test results are called for to fully assess the stationarity properties of the panel data.

Acknowledgments

Financial support to Sune Karlsson from the Swedish Research Council for Humanities and Social Sciences (HSFR) and to Mickael Löthgren from the Jan Wallanders and Tom Hedelius' Foundation for Research in the Social Sciences is gratefully acknowledged.

References

Coakley, J. and Fuertes A.M. 1997. New panel unit root tests of PPP, Economics Letters 57,17 – 22.

Culver, S.E. and Papell D.H. 1997. Is there a unit root in the inflation rate? Evidence from sequential break and panel data models, Journal of Applied Econometrics 12, 435 – 444.

Im, K.S., Pesaran, M.H. and Shin Y. 1997, Testing for unit roots in heterogeneous panels, Mimeo, Department of Applied Economics, University of Cambridge.

Levin, A. and Lin C.F. 1992. Unit root tests in panel data: Asymptotic and finite-sample properties, Mimeo, University of California, San Diego.

MacDonald, R. 1996. Panel unit root tests and real exchange rates, Economics Letters 50, 7 – 11.

McCoskey, S. and Selden S. 1998. Health care expenditures and GDP: Panel data unit root test results, Journal of Health Economics 17, 369-376.

Oh, K.-Y. 1996. Purchasing power parity and unit root tests using panel data, Journal of International Money and Finance 15, 405–418.

Papell, D.H. 1997. Searching for stationarity: Purchasing power parity under the current float, Journal of International Economics 43, 313 – 332.

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

Song, F.M. and Wu Y. 1998. Hysteresis in unemployment: Evidence from OECD countries, The Quarterly Review of Economics and Finance 38, 181–192.

Wu, Y. 1996. Are real exchange rates nonstationary? Evidence from a panel-data test, Journal of Money, Credit, and Banking 28, 54–63.