Empirical Economics (2005) 30:77–91
DOI 10.1007/s00181-004-0222-8

EMPIRICAL ECONOMICS

Testing for PPP: Should we use panel methods?

Anindya Banerjee[1], Massimiliano Marcellino[2], Chiara Osbat[3]

[1] Department of Economics, European University Institute, Via della Piazzuola 43, 50133 Firenze, Italy (e-mail: anindya.banerjee@iue.it)
[2] Istituto di Economia Politica - Universita' Bocconi, IGIER and CEPR Via Salasco 3, 20136 Milano, Italy
[3] European Central Bank, Kaiserstrasse 29, 60311 Frankfurt am Main, Germany

First version received: November 2001/Final version received : October 2003

Abstract. A common finding in the empirical literature on the validity of purchasing power parity (PPP) is that it holds when tested for in panel data, but not in univariate (i.e. country-specific) analysis. The usual explanation for this mismatch is that panel tests for unit roots are more powerful than their univariate counterparts. In this paper we suggest an alternative explanation. Existing panel methods assume that cross-unit cointegrating relationships, that would tie the units of the panel together, are not present. Using simulations, we show that if this important underlying assumption of panel unit root tests is violated, the empirical size of the tests is substantially higher than the nominal level, and the null hypothesis of a unit root is rejected too often even when it is true. More generally, this finding warns against the "automatic" use of panel methods for testing for unit roots in macroeconomic time series.

Key words: PPP, unit root, panel cointegration, cross-unit dependence

JEL classifications: C12, C13, C22, C23

1. Introduction

The last few years have seen a veritable explosion of papers on testing for purchasing power parity (PPP) using panels of macroeconomic data. A

We wish to thank an anonymous referee and Ronnie MacDonald for helpful comments. The financial support of the European University Institute in funding this research is also gratefully acknowledged.
The views expressed in this study are those of the authors and do not reflect those of the ECB.
All correspondence to: Anindya Banerjee.

selection would include *inter alia* Frankel and Rose (1995), Jorion and Sweeney (1996), Oh (1996), Rogoff (1996), O'Connell (1998), and more recently, Bayoumi and MacDonald (1999), Pedroni (1999), Papell and Theodoridis (2001), Papell (2002), and Chang (2002).

The real exchange rate (*rer*) for country *i*, if defined with respect to the United States (US) dollar as the numeraire currency, is constructed as:

$$Q_{it} = E_{it}\frac{P_t^*}{P_{it}}$$

where E_{it} is the nominal exchange rate, P_t^* is the US consumer price index (CPI) and P_{it} is the CPI for country *i*. Denoting logarithms in lower case letters, we therefore have

$$q_{it} = e_{it} + p_t^* - p_{it},$$

The strong form of the test for PPP consists of testing the null hypothesis of a unit root in the q_{it} series, either individually (i.e. country by country) or using panel methods, discussed briefly in the next section.

More generally, a weak form of the hypothesis may be tested by constructing the series $\widetilde{q}_{it}$, where

$$\widetilde{q}_{it} = e_{it} + \alpha p_t^* - \beta p_{it}.$$

If $\widetilde{q}_{it}$ contains a unit root, the weak form of the PPP hypothesis is also rejected. The parameters α and β capture differences in the composition of the CPI baskets across countries, transaction and transportation costs, and other wedges, and may be estimated from single-equation- or system-cointegration methods, given that all three component series of $\widetilde{q}_{it}$ are assumed to be at least integrated of order one (denoted I(1)). Testing the weak form of PPP in panels may thus be seen as testing for cointegration.

The starting point for the use of panel unit root and cointegration tests is the presumption that univariate (i.e. country-specific) tests for PPP, which are typically based on augmented Dickey-Fuller (ADF) regressions on real exchange rates, have low power. Among the large number of papers available in the literature, special attention should be drawn to the work of Papell and his co-authors, Murray and Papell (2002), Murray and Papell (2003), Papell and Theodoridis (2001), Papell (2002), Papell (2003) and Papell and Prodan (2003), who have investigated a large number of issues that are relevant to an appropriate assessment of the evidence derived from panel tests.

The origin of our research in this paper is the evidence presented in Banerjee et al. (2000) (henceforth BMO). In that paper we investigated the properties of panel cointegration tests, particularly those proposed by Larsson and Lyhagen (2000) and Pedroni (1999), and showed that disregarding the cointegrating relationships across the countries in the panel led to serious difficulties in making inference about cointegration within each country in a panel. We showed that when the restriction that there are no cointegrating relationships among the variables across the countries in the panel is valid, the tests have the correct size and high power to detect cointegration. If the restriction is invalid, however, the tests for cointegration tend to be grossly over-sized especially as T increases, so that the null of no cointegration is rejected too often in relation to the nominal confidence level (or size) of the test.

Within the context of testing for the strong form of PPP, the panel approach to testing for unit roots rules out the existence of cointegrating relationships between q_{it} and q_{jt} for all $i \neq j$. In this paper we demonstrate the consequences of the violation of this restriction by looking at the properties of some panel unit root tests commonly used to test for stationarity of the real exchange rate, and show that in common with the critique presented in BMO, size distortions lead to misleading inference. While the analysis in BMO is therefore more directly relevant for the consideration of the weak form of PPP, our main aim here is to explore further the arguments within the context of the strong form of the PPP hypothesis.[1]

In the next section, we provide a brief overview of three of the most commonly used tests for unit roots in panels as proposed in the papers by Levin and Lin (1992) and Levin and Lin (1993) (LL)[2], Im, Pesaran and Shin (2003) (IPS) and Maddala and Wu (1999) (MW). More comprehensive summaries are provided *inter alia* by Banerjee (1999) and Baltagi and Kao (2000). Section 3 presents and discusses an extensive set of Monte Carlo experiments that analyze the size and power properties of these tests under various data generating processes (DGPs). Our results strongly indicate that the rejections of the unit root null hypothesis, commonly attributed in the empirical literature to the higher power of panel unit root tests, may be due simply to the over-sizing that is present when cointegrating relationships link the countries of the panel together. Section 4 provides an empirical example demonstrating that while single-country ADF tests typically fail to reject the null hypothesis of a unit root for the real exchange rate, panel unit root tests reject this null when using standard critical values. However, when using modified critical values for the panel tests that take into account the presence of cross-country cointegration, the real exchange rate is found to be non-stationary, in line with the country-by-country analysis. Section 5 offers conclusions and closing remarks.

2. Testing for unit roots in dynamic panels

In this section we investigate the properties of the three kinds of panel unit root tests proposed respectively by LL, IPS and MW. Generally speaking, all the tests we investigate are based on the following regression:

$$\Delta y_{it} = \gamma_i + \delta_i t + \theta_t + \rho_i y_{it-1} + \zeta_{it}, \; i = 1, 2, .N, \; t = 1, 2, ..., T. \tag{1}$$

which allows for fixed effects and unit-specific time trends (which may be set to zero if necessary depending upon the model to be considered).[3] In the simplest specifications the error term $\zeta_{it} \sim IID(0, \sigma^2)$[4], but all models

[1] This over-sizing property of panel unit root tests has also been discussed by Engel (2000) and O'Connell (1998) in slightly different contexts. In particular, O'Connell shows the effect of *short-run* linkages among the units (for example through the non-zero covariances of the errors across the units) on unit root tests. Our study deals with the effects of long-run or cointegrating relationships.

[2] The theoretical results of Levin and Lin (1992) and Levin and Lin (1993) with updated tables were published in Levin et al. (2002). We have chosen to retain the reference to the original papers for convenience.

[3] As below, we may augment this regression with lagged values of Δy_{it} to allow for serial correlation in the errors.

[4] This is an assumption which can be generalized to allow, for example, heteroscedasticity and some dependence in the error terms in each unit.

are based on the important assumption that $E\left[\zeta_{it}\zeta_{js}\right] = 0 \ \forall \ t, s$ and $i \neq j$. This assumption is frequently made after allowing for the common time effects θ_t, which in practice can be concentrated out of the equation by taking deviations from cross-sectional means. The null hypothesis of interest for all three tests is H_0: $\rho_i = 0 \ \forall \ i$, although as we see below, the tests allow for different degrees of heterogeneity of ρ_i under the alternative hypothesis.

Levin and Lin (1992) consider for the alternative hypothesis the case where the autoregressive coefficient is homogeneous across countries, *i.e.* H_A: $\rho_i = \rho < 0 \ \forall \ i$. This imposes rather restrictive assumptions on the dynamics under the alternative hypothesis, which, at least in the case of testing for PPP, seem to lack an economic rationale. They derive the asymptotic distributions of the panel estimator of ρ under different assumptions on the presence of fixed effects or heterogeneous time trends. The asymptotic distributions of the ordinary least squares (OLS) pooled panel estimator and associated t-statistic (denoted LL) have Gaussian limiting distributions, after allowing for mean and variance adjustments, which are computed by Monte Carlo simulation and tabulated in their paper.

Following the critique of Pesaran and Smith (1995) on pooled panel estimators, such as those used by Levin and Lin (1992) and Levin and Lin (1993), Im et al. (2003) extend the Levin and Lin framework to allow for heterogeneity in the value of ρ_i under the alternative hypothesis. The alternative hypothesis is then specified as $H_A : \rho_i < 0, \ \ i = 1, 2, ..., N_1, \ \ \rho_i = 0,$ $i = N_1 + 1, N_1 + 2, ...N$ and a group-mean t-bar statistic for $\rho_i = 0$ is based on the N ADF regressions given by

$$\Delta y_{it} = \gamma_i + \rho_i y_{it-1} + \sum_{j=1}^{p_i} \theta_{ij} \Delta y_{it-j} + \varepsilon_{it}, \ \ t = 1, 2, ...T, i = 1, 2, ...N. \tag{2}$$

The group-mean t-bar statistic is constructed as

$$\Psi_{\bar{t}} = N^{-1} \sum_{i=1}^{N} t_{iT}$$

where t_{iT} is the $t-$statistic for ρ_i computed for each unit i.

IPS show that under H_0: $\rho_i = 0 \ \forall i$, $\Psi_{\bar{t}}$, after adjustment for mean and variance (with adjustment factors obtainable by stochastic simulation, as tabulated in Im et al. (2003) using 50,000 replications for different values of T and p_i), has a Gaussian limiting distribution. That is,

$$Z_{\bar{t}} = \frac{\sqrt{N}\{\Psi_{\bar{t}} - E(\Psi_{\bar{t}})\}}{\sqrt{Var(\Psi_{\bar{t}})}} \Longrightarrow N(0, 1) \tag{3}$$

as $T, N \longrightarrow \infty$ and $\frac{N}{T} \longrightarrow k$ where k is a finite positive constant.[5]

[5] IPS also discuss the use of a statistic called $Z_{\widetilde{tbar}}$ which is asymptotically equivalent to the $Z_{\bar{t}}$ statistic (if $T \to \infty$ followed by $N \to \infty$) but differs only as far as the choice of the estimator of the variance of ε_{it} is concerned. Note that in IPS (Assumption 3.1, p. 56), the ε_{it} terms are assumed to be independently and normally distributed variables for all i and t with zero means and finite heterogeneous variances σ_i^2.

In an earlier version of their paper, IPS also propose the use of a group-mean Lagrange multiplier (LM) statistic, which is based on averaging the single-country $LM-$ statistics for $\rho_i = 0$ derived from (2). Allowing for mean and variance corrections provided by them, the convergence result stated for $\Psi_{\bar{t}}$ holds also for $\Psi_{\overline{LM}}$, and consistency is guaranteed under the controlled rate of divergence of N and T to infinity. That is, defining $\Psi_{\overline{LM}} = N^{-1} \sum_{i=1}^{N} LM_i$,

$$Z_{\overline{LM}} = \frac{\sqrt{N}\{\Psi_{\overline{LM}} - E(\Psi_{\overline{LM}})\}}{\sqrt{Var(\Psi_{\overline{LM}})}} \Longrightarrow N(0,1) \tag{4}$$

In a Monte Carlo study given in the earlier version, IPS demonstrate better finite-sample performances of the $Z_{\bar{t}}$ and $Z_{\overline{LM}}$ tests in comparison with LL.

Finally, Maddala and Wu, relying on Fisher (1932), suggest combining the p-values of any given test-statistic for a unit root in each cross-sectional unit. The statistic is given by

$$MW : -2\sum_{i=1}^{N} ln(\pi_i),$$

where π_i is the p-value of the test statistic in unit i, and is distributed as a $\chi^2(2N)$ under the usual assumption of cross-sectional independence. The Fisher test is an exact and non-parametric test, and may be computed for any *arbitrary* choice of a test for the unit root in a cross-sectional unit. In this paper, however, we concentrate on using the ADF t-test for each unit in order to construct the MW test statistic. The obvious simplicity of this test and its robustness to the choice of lag length and sample size make its use attractive.

3. Simulation results for panel unit root tests

In this section we describe the Monte Carlo design, and evaluate the performance of the panel unit root tests, both under the standard assumption of no cross-unit cointegrating relationships, and in the presence of such relationships. Our results allow us to evaluate the reliability of the panel methods when applied to test for PPP and, more generally, to evaluate the stationarity of macroeconomic time series for several countries.

3.1. The Monte Carlo design

Although our focus is on univariate panel unit root tests, for our Monte Carlo study we generate data by using a bivariate system for each i, in order to allow for a richer variety of possible data generating processes. Let us consider the variables y_{it}, x_{it}, where $i = 1, ...N$ indexes the countries and $t = 1, ..., T$ is the temporal index. We group the y and x variables for each country into the $N \times 1$ vectors $Y_t = (y_{1t}, ..., y_{Nt})'$ and $X_t = (x_{1t}, ..., x_{Nt})'$. We then consider the following error correction model (ECM) as the DGP:

$$\begin{pmatrix} \Delta Y_t \\ \Delta X_t \end{pmatrix} = \alpha\beta' \begin{pmatrix} Y_{t-1} \\ X_{t-1} \end{pmatrix} + \varepsilon_t, \tag{5}$$

where ε_t is $i.i.d.$ $N(0, \Sigma)$, with

$$\Sigma = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{2N}^2 \end{pmatrix},$$

and σ_j^2 is extracted from a uniform distribution on $[0.5, 1.5]$, for $j = 1, ..., 2N$.

We have also considered DGPs with lagged differences as regressors and serially correlated errors. There were no qualitative differences in the results, and we therefore focus on the specification in (5). The different experiments for size and power, with or without cross-unit cointegration, are obtained by imposing restrictions on the α and β matrices, as detailed below. In all cases the results are based on 5000 replications, and are reported for different values of N (5, 10, 25, 50, 100) and of T (25, 50, 100), to mimic situations often encountered in empirical applications.

3.2. No cross-unit cointegration

The first case we consider is testing for unit roots when the assumptions underlying the panel tests hold, i.e. there are no cross-unit cointegrating relationships. We are interested in evaluating the size and power of the tests in small samples, to establish a benchmark for the case when the independence assumption is violated.

For the size experiments we set $\alpha = \mathbf{0}$ in the DGP, so that X_t and Y_t are independent random walks. For the power experiments we use

$$\alpha = \begin{pmatrix} I_N & I_N \\ 0_N & 0_N \end{pmatrix}, \quad \beta' = \begin{pmatrix} -I_N & 0_N \\ cI_N & 0_N \end{pmatrix},$$

where I_N and 0_N are an $N \times N$ identity matrix and an $N \times N$ matrix of zeros respectively, and $c = 0.9$ or $c = 0.8$. Hence, X_t is an N-variate random walk, while Y_t is made up of N stationary AR(1) processes, with roots equal to c.

The estimated model for each unit is

$$\Delta y_{it} = \gamma_i + \rho_i y_{it-1} + \theta_{i1}\Delta y_{it-1} + \theta_{i2}\Delta y_{it-2} + e_{it}, \tag{6}$$

with $\rho_i = \rho \ \forall \ i$ for *LL* (which uses the pooled panel estimator, and has homogeneous ρ_i under both the null and the alternative hypotheses.)

The choice of lag length reflects the fact that when the lag length is underspecified relative to the DGP the tests become very undersized, while overspecification improves the size, even though the power can deteriorate slightly. The choice of the lag truncation in computing $\widehat{\sigma}^2_{y_i}$ in (2) is also important in determining size, and we follow LL's recommendations for comparability with existing results in the literature. Note further that the model has a unit-specific constant to allow for fixed effects in estimation, while the DGP does not contain any deterministic terms.

The empirical sizes of the unit root tests are reported in Table 1. We can read this table along three directions: for varying N, for varying T, and for both N and T varying. In general, for fixed T, the performance of the $Z_{\overline{LM}}$ test improves when N increases, while for the $Z_{\bar{t}}$ and, in particular, for the *MW* test it deteriorates; for *LL*, the size decreases steadily. For fixed N, the size distortions of all

Table 1. Unit root tests, no cross-unit cointegration - size

N/T		20	50	100
5	$Z_{\overline{LM}}$	0.067	0.073	0.070
	$Z_{\bar{t}}$	0.137	0.083	0.071
	LL	0.091	0.056	0.060
	MW	0.185	0.092	0.067
10	$Z_{\overline{LM}}$	0.066	0.073	0.073
	$Z_{\bar{t}}$	0.132	0.081	0.072
	LL	0.090	0.055	0.061
	MW	0.181	0.097	0.070
25	$Z_{\overline{LM}}$	0.041	0.056	0.055
	$Z_{\bar{t}}$	0.189	0.098	0.069
	LL	0.049	0.039	0.044
	MW	0.325	0.118	0.075
50	$Z_{\overline{LM}}$	0.041	0.052	0.050
	$Z_{\bar{t}}$	0.271	0.109	0.071
	LL	0.042	0.031	0.037
	MW	0.480	0.143	0.081
100	$Z_{\overline{LM}}$	0.031	0.047	0.049
	$Z_{\bar{t}}$	0.382	0.129	0.077
	LL	0.023	0.019	0.029
	MW	0.674	0.176	0.097

tests decrease with T, a feature that is particularly evident for MW. When both N and T vary, we consider the cases where $\frac{N}{T}$ is about 0.25 (($N = 5, T = 20$), ($N = 10, T = 50$), ($N = 25, T = 100$)) or 0.5 (($N = 10$, $T = 20$), ($N = 25, T = 50$), ($N = 50, T = 100$)). The size improvements for larger dimensions are clear, in particular when $\frac{N}{T} = 0.5$. Overall, $Z_{\overline{LM}}$ has very low size distortions when $N > 25$ and $T > 50$, while the other three tests also experience deviations from the nominal level for these quite large values of N and T. It is worth noting that LL performs better when N is small, even for $N < 25$.

The power results are reported in Table 2 for $c = 0.9$.[6] For $T = 20$ and N increasing, $Z_{\bar{t}}$ and MW have the highest power, which is not surprising, given their large size distortions. The $Z_{\overline{LM}}$ and LL tests on the other hand, which have better size properties, have disappointingly low power for large N, *i.e.* less than 40% even for $N = 100$. When $T = 50$, the situation improves substantially, in particular for the $Z_{\overline{LM}}$ test, even for $N = 25$. When $T = 100$ the power is close or equal to one for all the tests, with the exception of LL when $N = 5, 10$. For fixed N, power increases substantially with T (even when $N = 5$), and a similar result holds when N and T vary jointly.

Overall, these figures indicate that the length of the sample period is quite important for achieving high power in a panel context, and that the $Z_{\overline{LM}}$ test not only has good empirical size but also rather high power, particularly when $N > 25$ and $T > 50$.[7]

[6] Results for $c = 0.8$ do not change qualitatively, but power increases faster as N and T increase.

[7] We also considered different values for c across the units. This violates the condition of the LL test on the alternative hypothesis, and indeed the LL test turns out to have systematically lower power than $Z_{\bar{t}}$, $Z_{\overline{LM}}$ and MW in this case.

Table 2. Unit root tests, no cross-unit cointegration - power, roots = 0.9

N/T		20	50	100
5	$Z_{\overline{LM}}$	0.100	0.338	0.870
	$Z_{\bar{t}}$	0.218	0.413	0.905
	LL	0.128	0.161	0.402
	MW	0.252	0.346	0.826
10	$Z_{\overline{LM}}$	0.114	0.523	0.992
	$Z_{\bar{t}}$	0.305	0.690	0.998
	LL	0.133	0.241	0.717
	MW	0.348	0.542	0.986
25	$Z_{\overline{LM}}$	0.156	0.870	1.000
	$Z_{\bar{t}}$	0.553	0.978	1.000
	LL	0.181	0.502	0.987
	MW	0.569	0.890	1.000
50	$Z_{\overline{LM}}$	0.228	0.991	1.000
	$Z_{\bar{t}}$	0.780	1.000	1.000
	LL	0.231	0.800	1.000
	MW	0.793	0.993	1.000
100	$Z_{\overline{LM}}$	0.374	1.000	1.000
	$Z_{\bar{t}}$	0.971	1.000	1.000
	LL	0.333	0.978	1.000
	MW	0.965	1.000	1.000

3.3. Cross-unit cointegration

We now evaluate the size properties of the panel tests when allowing for cross-unit cointegrating relationships while maintaining the hypothesis that all the series are I(1). In this set of DGPs therefore:

$$\alpha = 0.1\begin{pmatrix} -I_N & 0_N \\ 0_N & -I_N \end{pmatrix}, \quad \beta' = \begin{pmatrix} -I_N & I_N \\ 0_N & B_N \end{pmatrix}, \tag{7}$$

$$B = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 \\ \dots & & & & & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 \end{pmatrix}. \tag{8}$$

Hence, all the variables are $I(1)$, but the whole $2N$-dimensional system is driven by less than $2N$ independent random walks. More precisely, the number of independent random walks is equal to the number of zero rows of B, say b. For example, in the formulation in (8), $b = N - 3$. The remaining $2N - b$ stationary roots of the system are equal to 0.9. There exist N within-country cointegrating relationships, $(y_{it} - x_{it})$, $i = 1, \dots, N$, plus $N - b$ cross-

Table 3. Unit root tests, cross-unit cointegration, weak exogeneity - N = 10

q/T		20	50	100
20%	$Z_{\overline{LM}}$	0.055	0.076	0.110
	$Z_{\bar{t}}$	0.157	0.105	0.123
	LL	0.071	0.048	0.070
	MW	0.237	0.113	0.120
50%	$Z_{\overline{LM}}$	0.050	0.091	0.175
	$Z_{\bar{t}}$	0.158	0.128	0.199
	LL	0.070	0.055	0.084
	MW	0.232	0.126	0.174
70%	$Z_{\overline{LM}}$	0.066	0.148	0.347
	$Z_{\bar{t}}$	0.183	0.211	0.398
	LL	0.083	0.071	0.119
	MW	0.255	0.183	0.324
90%	$Z_{\overline{LM}}$	0.062	0.132	0.271
	$Z_{\bar{t}}$	0.177	0.196	0.337
	LL	0.082	0.065	0.116
	MW	0.240	0.162	0.251

q : Percentage of cross-unit cointegration.

unit cointegrating relationships of the type $(x_{it} - x_{i+1t})$, $i = 1, ...N - b$. Moreover, the x_i variables are weakly exogenous in the subsystem for the y_i variables.[8]

The estimated model for each unit remains

$$\Delta y_{it} = \gamma_i + \rho_i y_{it-1} + \theta_{i1}\Delta y_{it-1} + \theta_{i2}\Delta y_{it-2} + e_{it}, \tag{9}$$

and we are interested in evaluating if and how the presence of cross-unit cointegration affects the performance of the panel unit root tests. Note that each y_i is $I(1)$, so that the empirical rejection frequencies of the tests should be around 5%.

The results are reported in Tables 3 to 6, for different values of N and varying percentages of cross-unit cointegration, q, where $q = ((N - b)/N) * 100$. The parameter q typically ranges from 20% to 90%, i.e., from 20% to 90% of the variables across units are related by bivariate co-integrating relationships. Five points are worth making. First, the distortions for $Z_{\bar{t}}$ and MW are substantial in all cases, with values often in the range $20\% - 50\%$ when $N > 25$ and $T = 20$. Second, $Z_{\overline{LM}}$ performs reasonably well for $T = 20$, while LL appears to be quite robust to the presence of cross-unit cointegration, with only minor distortions. Third, focusing on $Z_{\overline{LM}}$ and LL, when T increases the distortions increase. Fourth, the distortions decrease for larger N in the case of LL, and do not increase with N for $Z_{\overline{LM}}$. Fifth, for $N > 25$ the distortions increase as the percentage of cross-unit cointegration increases, while for fewer units the outcome is less clear cut.

[8] Essentially this requires that the cointegrating relationships between y_i and x_i do not affect the x_i variables. See Johansen (1995) for a more precise definition of weak exogeneity in cointegrated systems. We also ran simulations where the DGP was specified in such a way that the x_i variables are not weakly exogenous in the subsystem for the y_i variables. We found that the lack of weak exogeneity did not pose any special problem. Detailed results are available upon request.

Table 4. Unit root tests, cross-unit cointegration, weak exogeneity - N = 25

q/T		20	50	100
25%	$Z_{\overline{LM}}$	0.044	0.081	0.133
	$Z_{\bar{t}}$	0.202	0.137	0.160
	LL	0.051	0.046	0.060
	MW	0.340	0.145	0.161
50%	$Z_{\overline{LM}}$	0.049	0.118	0.246
	$Z_{\bar{t}}$	0.227	0.206	0.310
	LL	0.059	0.052	0.073
	MW	0.359	0.192	0.255
70%	$Z_{\overline{LM}}$	0.053	0.129	0.238
	$Z_{\bar{t}}$	0.227	0.210	0.316
	LL	0.064	0.055	0.078
	MW	0.362	0.188	0.245
90%	$Z_{\overline{LM}}$	0.054	0.134	0.265
	$Z_{\bar{t}}$	0.246	0.235	0.356
	LL	0.069	0.064	0.087
	MW	0.376	0.201	0.266

q : Percentage of cross-unit cointegration

Table 5. Unit root tests, cross-unit cointegration, weak exogeneity - N = 50

q/T		20	50	100
	$Z_{\overline{LM}}$	0.043	0.081	0.108
	$Z_{\bar{t}}$	0.287	0.156	0.159
	LL	0.044	0.039	0.040
	MW	0.500	0.177	0.144
50%	$Z_{\overline{LM}}$	0.051	0.114	0.190
	$Z_{\bar{t}}$	0.317	0.232	0.273
	LL	0.049	0.041	0.051
	MW	0.513	0.219	0.221
70%	$Z_{\overline{LM}}$	0.053	0.147	0.301
	$Z_{\bar{t}}$	0.343	0.298	0.425
	LL	0.051	0.053	0.065
	MW	0.523	0.260	0.326
90%	$Z_{\overline{LM}}$	0.060	0.183	0.365
	$Z_{\bar{t}}$	0.364	0.354	0.504
	LL	0.060	0.060	0.084
	MW	0.535	0.296	0.381

q : Percentage of cross-unit cointegration

Overall, the evidence from our simulations suggests that the presence of cross-unit cointegration can substantially bias upwards the probability of type I error of panel unit root tests - namely the null hypothesis of a unit root is rejected far too often. In the following section we demonstrate the empirical relevance of our result when testing for PPP and show how account may be taken of the effects on inference of possible cross-country cointegration by a suitable adjustment of critical values.

Table 6. Unit root tests, cross-unit cointegration, weak exogeneity - N = 100

N/T		20	50	100
25%	$Z_{\overline{LM}}$	0.037	0.069	0.112
	$Z_{\bar{t}}$	0.407	0.187	0.178
	LL	0.028	0.023	0.032
	MW	0.694	0.212	0.165
50%	$Z_{\overline{LM}}$	0.044	0.115	0.242
	$Z_{\bar{t}}$	0.464	0.290	0.360
	LL	0.036	0.034	0.042
	MW	0.719	0.281	0.298
75%	$Z_{\overline{LM}}$	0.059	0.198	0.444
	$Z_{\bar{t}}$	0.521	0.467	0.631
	LL	0.046	0.053	0.073
	MW	0.745	0.396	0.485
90%	$Z_{\overline{LM}}$	0.064	0.228	0.487
	$Z_{\bar{t}}$	0.540	0.504	0.668
	LL	0.051	0.057	0.075
	MW	0.757	0.416	0.522

q : Percentage of cross-unit cointegration.

4. Testing for PPP

The results in the preceding section shed new light on the use of panel-based unit root tests. If the units are now interpreted as countries, then the amount of cross-country cointegration that may be assumed to exist, will affect the outcomes of the panel tests for a unit root. In this section we demonstrate this proposition empirically within the context of testing the strong version of purchasing power parity (PPP), namely testing whether the real exchange rate (*rer*) is stationary. This question has attracted considerable interest, as discussed in the introduction.

As reported in the data appendix below, we use quarterly data on nominal exchange rates and *cpi*, for the period 1975:1-2002:4, for 18 OECD countries.[9]

Since Papell and Theodoridis (2001) have argued that the choice of the *numeraire* currency is important, in Table 7 we report the country-by-country augmented Dickey-Fuller (ADF) test statistics using both US and Germany in turn as the *numeraire* country. We note that the test fails to reject the null hypothesis of a unit root in the *rer* for each country at any choice of lag length except for France and Korea when Germany is the *numeraire*.

The corresponding panel unit root test results are reported in Table 8. The asymptotic critical values for $Z_{\overline{LM}}$, $Z_{\bar{t}}$ and *LL* are standard normal, while *MW* is distributed as $\chi^2(36)$ (under the null hypothesis of a unit root in the *rer* and the maintained assumption of no cross-country cointegration). Table 9

[9] After 1999 and 2001, the EMU currencies nominal exchange rates are derived on the basis of the euro exchange rate and of the irrevocable parities between the legacy currencies and the euro. This implies of course that in the last part of the sample the series are by construction closely related, which reinforces the importance of our critique.

Table 7. Univariate ADF unit root tests for real exchange rates

	US numeraire				Germany numeraire			
lags	1	2	3	4	1	2	3	4
US	–	–	–	–	−1.83	−1.89	−2.26	−2.47
UK	−2.51	−2.51	−2.63	−2.75	−1.89	−1.85	−2.02	−2.12
Austria	−1.91	−2.00	−2.32	−2.47	−2.14	−2.00	−2.30	−1.91
Belgium	−1.73	−1.77	−2.20	−2.37	−2.11	−2.14	−2.20	−2.62
Denmark	−1.89	−1.97	−2.35	−2.59	−1.56	−1.63	−1.45	−1.70
France	−1.92	−1.96	−2.22	−2.43	−3.11*	−2.99*	−3.80**	−3.35*
Germany	−1.83	−1.89	−2.26	−2.47	–	–	–	–
Nether-lands	−1.96	−2.05	−2.33	−2.53	−1.14	−0.90	−0.66	−1.08
Canada	−0.95	−1.04	−1.82	−1.52	−2.29	−2.24	−2.75	−2.85
Japan	−2.17	−2.02	−2.35	−2.42	−2.45	−2.43	−2.46	−2.28
Finland	−1.59	−1.65	−2.02	−2.11	−1.91	−1.91	−2.20	−2.89*
Greece	−1.64	−1.89	−1.91	−2.59	−1.38	−1.74	−1.48	−1.68
Spain	−1.85	−2.12	−2.40	−2.32	−2.29	−2.45	−2.69	−2.38
Portugal	−1.84	−1.67	−1.97	−2.08	−1.40	−1.45	−2.04	−2.27
Italy	−2.07	−1.93	−2.37	−2.48	−2.01	−1.88	−2.33	−2.07
Switzer-land	−2.26	−2.21	−2.63	−2.70	−2.05	−1.84	−1.84	−1.64
Korea	−1.94	−1.53	−1.95	−1.74	−3.41*	−3.01*	−3.52**	−3.29*
Norway	−2.09	−2.03	−2.22	−2.17	−2.13	−1.95	−2.40	−2.49
Sweden	−1.42	−1.61	−2.02	−1.92	−2.40	−2.38	−2.52	−2.78

*, **% Indicate rejection at the 5% and 1% of the null hypothesis of a unit root.

Table 8. Panel unit root tests for real exchange rates

lags	1	2	3	4
US numeraire				
$Z_{\overline{LM}}$	3.47	2.42	2.43	2.83
$Z_{\bar{t}}$	−3.56	−2.91	−3.00	−3.39
LL	−2.23	−1.38	−1.36	−1.58
MW	66.57	55.74	56.04	59.51
Germany numeraire				
$Z_{\overline{LM}}$	3.16	2.03	2.37	2.90
$Z_{\bar{t}}$	−3.34	−2.63	−2.97	−3.46
LL	−1.73	−0.86	−1.17	−1.50
MW	63.40	51.90	55.19	60.06

reports the 5% critical values for these statistics simulated for $T = 100$, $N = 18$ and varying numbers of bivariate cross-country cointegrating relationships.[10] The rejection region is in the right tail of the density for $Z_{\overline{LM}}$ and MW, and in the left tail for the $Z_{\bar{t}}$ and LL tests.

We may see that when using the critical values appropriate for no cross-country cointegration, the unit root hypothesis is rejected in 13 out of 16 cases (4 tests and 4 different lag-length choices) with US as the *numeraire* and 12

[10] The actual number of observations per country is 108 but the closest avialable correction factors for the IPS and LL tests are tabulated for $T = 100$.

Table 9. 5% finite sample critical values for panel unit root tests (various proportions of bivariate cointegrating relationships; N = 18)*

	0/18	2/18	5/18	9/18	14/18	16/18
$Z_{\overline{LM}}$	1.73	2.07	2.36	3.03	3.64	4.12
$Z_{\bar{t}}$	−1.78	−2.09	−2.34	−2.90	−3.49	−3.92
LL	−1.60	−1.71	−1.94	−1.94	−2.08	−2.20
MW	52.98	56.05	58.27	64.79	69.62	73.88

* N.B. In order to correspond as closely as possible to the dimensions of the empirical panel, $N = 18$ and $T = 100$ are used in the simulations to complete the critical values reported in this table. In additon, varying numbers of variate cross-country cointegrating relationships (as given in the column headings), defined by Eq. (7) and (8), are applied.

out of 16 cases with Germany as the *numeraire*. By contrast, when using critical values adjusted for the presence of cross-country cointegration, these rejection rates decrease. For example, with 14 bivariate cointegrating relationships, the unit root hypothesis is rejected in only 2 out of 16 cases with US as the *numeraire* and never with Germany as the *numeraire* .[11]

The following conclusions emerge. The country-by-country ADF tests fail to reject the hypothesis of a unit root in real exchange rates. In contrast, $Z_{\overline{LM}}$, $Z_{\bar{t}}$ and *MW* reject this hypothesis. This mismatch may be attributed to a bias in the panel tests, that leads to the rejection of the presence of a unit root too often when there are cross-country cointegrating relationships. Indeed, when using corrected critical values, the panel tests fail to reject the unit root hypothesis. The *LL* test appears to be more robust in this respect. This is reflected not only in the relative invariance of the critical values (see Table 9) to increasing amounts of cross-country cointegration, but also in the finding that this test does not reject the presence of a unit root in most cases, in agreement with the country-by-country ADF tests. Finally, these findings are robust to the choice of the *numeraire* country.

5. Conclusions

The simulation results and empirical analysis in our paper demonstrate clearly the importance of taking account of the presence of cross-country cointegrating relationships in interpreting the results of panel unit root tests. Not paying proper heed to such relationships leads to panel unit root tests being grossly over-sized, *i.e.* the unit root null is rejected too often.

Many of the findings in the empirical panel unit root literature on the stationarity of real exchange rates could be interpreted within the light of this critique. As we have shown in our illustration, the correct critical values (in the presence of cross-country cointegration and tailored to the dimensions of our empirical panel of countries) are in many cases substantially larger in absolute value than their commonly used asymptotic equivalents. This is an intuitive finding, since the critical values *need* to be larger in order to control for the size distortions so clearly visible in the relevant Monte Carlo simulations.

[11] We consider only bivariate cointegrating relationships but similar results can be expected when more than two countries are linked in equilibrium.

The *LL* test appears to suffer the least from size distortion in the presence of cross-country cointegrating relationships. This could occur because *LL* is the only pooled test which, by construction, takes some account of the relationships linking the countries. In addition, given that the autoregressive parameter ρ to be estimated is homogeneous under both the null and alternative hypotheses, pooling is not an unreasonable restriction since account can be taken of the heterogeneous deterministic terms by methods such as demeaning the variables unit-by-unit. The remaining tests, by averaging (or aggregating, in the case of *MW*) across test statistics constructed country by country, fail to take any account of these cross-country restrictions or to impose the valid simplifying restriction of homogeneity. Our finding must be evaluated against the background that the DGPs considered here are on the whole those most favourable to the LL testing framework. Additional results, which are available upon request, indicate that when the DGP allows for heterogeneity, the power of the *LL* test is lower than for the other panel tests considered.

Overall, our results provide a serious warning against the " automatic" use of standard panel unit root tests without proper consideration of the underlying assumption of cross-unit independence.

Data appendix

The database used in sect. 4 contains quarterly data on nominal exchange rates and consumer price indices (CPI) from 1975q1 to 2002q4. The source is the IMF's International Financial Statistics (IFS) database. The exchange rates are bilateral USD rates, line rf in the IFS database. For the period after 1999, the USD exchange rates for the OECD countries that are members of the euro area were constructed using the irrevocable conversion rates of the legacy currencies into the euro.

The CPI data are line 64 in the IFS database.

References

Baltagi BH, Kao C (2000) Nonstationary panels, cointegration in panels and dynamic panels: A survery. In: Baltagi B(ed) Advances in econometrics, 15. Elsevier Science, New York, pp. 7–51

Banerjee A (1999) Panel data unit roots and cointegration: An overview. Oxford Bulletin of Economics and Statistics 61: 607–630

Banerjee A, Marcellino M, Osbat C (2000) Some cautions on the use of panel methods for integrated series of macro-economic data. Working Paper 20. European University Institute. [Revised in 2003]

Bayoumi T, MacDonald R (1999) Deviations of exchange rates from purchasing power parity: A story featuring two monetary unions, Staff Papers 46, 1, IMF

Chang Y (2002) Nonlinear IV unit root tests in panels with cross- sectional dependency. Journal of Econometrics 110: 261–292

Engel C (2000) Long-run PPP may not hold after all. Journal of International Economics 57: 243–273

Fisher R (1932) Statistical methods for research workers. Oliver and Boyd, Edimburgh

Frankel JA, Rose AK (1995) Empirical research on nominal exchange rates. In Grossman G M, Rogoff K (eds) Handbook of international economics, vol. 3. Elsevier Science, Amsterdam, pp. 1689–1729

Im KS, Pesaran MH, Shin Y (2003) Testing for unit roots in heterogeneous panels. Journal of Econometrics 115: 53–74
Johansen S (1995) Likelihood-based inference in cointegrated vector autoregressive models. Oxford University Press, Oxford and New York
Jorion P, Sweeney R (1996) Mean reversion in real exchange rates: Evidence and implications for forecasting. Journal of International Money and Finance 15: 535–550
Larsson R, Lyhagen J (2000) Likelihood-based inference in multivariate panel cointegration models, Technical Report 1313, Econometric Society World Congress 2000 Contributed Papers
Levin A, Lin C-F (1992) Unit root tests in panel data: Asymptotic and finite sample propertics. Technical Report 92-93. Department of Economics, University of California at San Diego
Levin A, Lin C-F (1993) Unit root tests in panel data: New results. Discussion Paper 56, University of California at San Diego
Levin A, Lin C-F, Chu C-SJ (2002) Unit root tests in panel data: Asymptotic and finite-sample properties. Journal of Econometrics 108: 1–24
Maddala G, Wu S (1999) A comparative study of unit root tests and a new simple test. Oxford Bulletin of Economics and Statistics 61: 631– 652
Murray CJ, Papell DH (2002) The purchasing power parity persistence paradigm. Journal of International Economics 56: 1–19
Murray CJ, Papell DH (2003) Do panels help solve the purchasing power parity puzzle? mimeo
O'Connell P (1998) The overvaluation of purchasing power parity. Journal of International Economics 44: 1–19
Oh K (1996) Purchasing power parity and unit root tests using panel data. Journal of International Money and Finance 15: 405–418.
Papell DH (2002) The great appreciation, the great depreciation, and the purchasing purchasing power parity hypothesis. Journal of International Economics 57: 51–82
Papell DH (2003) The panel purchasing power parity puzzle. mimeo
Papell DH, Prodan R (2003) Long run purchasing power parity: Cassel or Balassa-Samuelson? mimeo
Papell DH, Theodoridis H (2001) The choice of numerairc currency in panel tests of purchasing power parity. Journal of Money, Credit and Banking 33: 790–803
Pedroni P (1999) Critical values for cointegration tests in heterogeneous panels with multiple regressors. Oxford Bulletin of Economics and Statistics 61 (Special Issue): 653–670
Pedroni P (2001) Purchasing power parity tests in cointegrated panels. The Review of Economics and Statistics 83: 727–731
Pesaran MH, Smith R (1995) Estimating long-run relationships from dynamic heterogeneous panels. Journal of Econometrics 68: 79–113
Rogoff K (1996) The purchasing power parity puzzle. Journal of Economic Literature 34: 647–668