

Josep Lluís Carrion-i-Silvestre · Andreu Sansó

A guide to the computation of stationarity tests

Received: 9 September 2003 / Accepted: 7 April 2005 / Published online: 14 March 2006
© Springer-Verlag 2006

Abstract In this paper we compare ways of computing stationarity tests. We show that whereas some of the procedures recommended lead to inconsistency of the tests, it is still possible to compute a test with good properties in finite sample in terms of empirical size and power. The guidance suggested in the paper is illustrated by testing for the purchasing power parity hypothesis in some developed countries.

Keywords Stationarity test · KPSS test · LMC test · Optimal bandwidth selection · Long-run variance estimation · Purchasing Power Parity

1 Introduction

The literature on unit root and stationarity tests has experienced a huge advance in recent years. On the one hand, the proposals of unit root tests have achieved a higher degree of development with the tests in Ng and Perron (2001). These authors define a set of modified test statistics that show the best performance in finite samples in terms of empirical size and power. Many authors have focused on stationarity tests. In contrast to unit root tests, these tests specify the null hypothesis of stationarity and the alternative of non-stationarity, so they can be seen as the reversal complement of the unit root tests. The relevant proposals in this field are Nabeya and Tanaka (1988), Kwiatkowski et al. (1992)—hereafter, KPSS test—and Leybourne and McCabe (1994, 1999)—henceforth, LMC test. Some authors have advocated the joint application of both kinds of statistics as a way to obtain robust

J. L. Carrion-i-Silvestre
AQR Research Group, Departament d'Econometria, Estadística i Economia Espanyola,
Universitat de Barcelona, Spain

A. Sansó
Department d'Economia Aplicada, Universitat de les Illes Balears, Spain

J. L. Carrion-i-Silvestre (✉)
Av. Diagonal, 690, 08034 Barcelona, Spain
E-mail: carrion@ub.edu

conclusions about the stochastic time series properties—see Amano and Van Norden (1992), Cheung and Chinn (1997), Charemza and Syczewska (1998), Maddala and Kim (1998), Carrion-i-Silvestre et al. (2001), and Gabriel (2003) for a discussion on confirmatory analysis.

However, the main drawback of stationarity tests is the difficulty entailed in the estimation of the long-run variance involved in the computation of the tests. While the KPSS test uses a non-parametric estimation for the long-run variance, the LMC test conducts the estimation in a parametric way. Both procedures require the definition of a lag-length parameter, which is shown to have high influence on the finite sample properties of the tests—see Lee (1996) and Chen (2002) for the KPSS test. This feature has recently been pointed out in Caner and Kilian (2001), who indicate that the KPSS and LMC tests show size distortions when the stochastic process is near to non-stationarity. Besides, Engel (2000) warns about the use of the KPSS because of its lack of power. Chen (2002) also points out that the power of the test is affected when the lag length is not appropriate.

We argue that the pitfalls can be mitigated once the long-run variance is properly estimated. However, this is not easy. Thus, in most situations the bandwidth for the spectral window that requires the non-parametric approach is chosen by automatic methods—i.e. Andrews (1991) or Newey and West (1994) procedures—although this leads to the inconsistency—see Choi and Ahn (1995, 1999) and Kurozumi (2002). This feature is of special interest for applied economists, because some of most popular econometric programs do not take this inconsistency problem into account. For instance, the latest versions of Eviews and Stata offer the possibility of computing the KPSS test using an automatic bandwidth selection method that causes the inconsistency. Instead, we could follow the parametric approach. Therefore, comparison of the two approaches is relevant from an empirical point of view.

In this paper we assess the performance of a set of alternatives that can be used when computing the stationarity tests. In this regard, we consider the suggestions in Choi and Ahn (1995, 1999), Kurozumi (2002) and Sul et al. (2005) when implementing an automatic procedure to estimate the long-run variance from a non-parametric point of view. For the LMC test, we can apply an information criterion or the testing procedure defined in Leybourne and McCabe (1999).

Here we provide a guide to the estimation of the long-run variance when testing the null hypothesis of stationarity. The guide given is illustrated with an example based on the purchasing power parity (PPP) hypothesis. When reviewing the literature we noticed that stationarity tests are widely applied to the PPP hypothesis framework, but little attention has been paid to the estimation of the long-run variance. Here we focus on two databases depending on the frequency of the time series. First, we examine the annual data set in Taylor (2002), which allow us to test the PPP for 19 countries, covering a century. Second, we analyse the PPP hypothesis using monthly time series from the *International Financial Statistics* database of the IMF. As expected, the properties of the statistics improve as the number of observations increases. Thus, it is interesting to assess the robustness of the analysis using time series that allow us to focus on features such as the long-run and the span of the information.

The outline of the paper is as follows. In Section 2 we describe the stationarity tests and the different approaches to the estimation of the long-run variance. In Section 3 we conduct Monte Carlo simulations in order to assess the finite sample

properties of each procedure. Section 4 illustrates the PPP hypothesis. Finally, Section 5 concludes.

2 Stationarity tests

Let $\{y_t\}$ be the stochastic process with the Data Generation Process (DGP) given by:

$$y_t = \alpha_t + \beta t + u_t \tag{1}$$

$$\alpha_t = \alpha_{t-1} + v_t, \tag{2}$$

$t=1, \dots, T$, where $\{u_t\}$ and $\{v_t\}$ are assumed to be two independent stochastic processes that satisfy the strong-mixing conditions in Phillips (1987) and Phillips and Perron (1988). In this framework the null hypothesis of stationarity implies that $\sigma_v^2=0$ since in this case the non-stationary component α_t collapses to a constant α_0 , and, hence, the time series is generated by stationary components. Consequently, the alternative hypothesis of non-stationarity is given by $\sigma_v^2>0$. As mentioned above, we deal with two different statistics to analyse the stochastic properties of the time series, i.e. the KPSS and the LMC statistics. In what follows we briefly present the main features of these statistics.

The KPSS test is given by:

$$\hat{\eta}_i = \hat{\sigma}^{-2} T^{-2} \sum_{t=1}^T S_t^2, \tag{3}$$

$i=0, 1, 2$, where $S_t = \sum_{j=1}^t \hat{u}_j$ denotes the partial sum process, $\{\hat{u}_t\}$ is the OLS estimated residuals that are obtained after y_t is regressed against the deterministic function $f(t)$. Specifically, the subscript i in Eq. 3 denotes the deterministic function. Thus, we use $i=0$ for the model that does not include any deterministic term, i.e. $f(t)=0$. The model that includes a constant term, $f(t)=\mu$, is denoted by $i=1$, while $i=2$ is for the model with a time trend, $f(t)=\mu+\beta t$. The long-run variance is estimated using the non-parametric estimator:

$$\hat{\sigma}^2 = T^{-1} \sum_{t=1}^T \hat{u}_t^2 + 2T^{-1} \sum_{s=1}^l w(s, l) \sum_{t=s+1}^T \hat{u}_t \hat{u}_{t-s},$$

where $w(s, l)$ denotes the spectral window. While the choice of the kernel is dependant on the preference of practitioners—Kwiatkowski et al. (1992) use the Bartlett kernel, but Hobijn et al. (1998) suggest the Quadratic spectral window—caution should be taken when estimating the spectral bandwidth. Thus, we can find suggestions that can lead to erroneous conclusions. For instance, Lee (1996) uses Andrews (1991) method, while Hobijn et al. (1998) suggest applying the automatic methods in Newey and West (1994) to estimate the bandwidth—henceforth, we denote the estimated bandwidth parameter that is selected using these methods as \hat{l}_{auto} . Unfortunately, Choi and Ahn (1995, 1999) and Kurozumi (2002) advise that

the use of these data-based selection methods provokes the inconsistency of the test. Notwithstanding, some bounds to control the estimated bandwidth can be imposed to avoid such inconsistency. Below we highlight three alternatives.

First, the rule in Choi and Ahn (1995, 1999) establishes that the bandwidth that is estimated using Andrews’ method should be constrained to $\hat{l} = 2$ if $\hat{l}_{auto} \geq T^\delta$, where $\delta=0.7$ for raw series—i.e. for the deterministic specification given by $i=0$ —and $\delta=0.65$ for detrended time series—i.e. for the deterministic specification given by $i=1,2$. Second, Kurozumi (2002) proposes that the bandwidth for the Bartlett kernel should be estimated as:

$$\hat{l} = \min \left\{ 1.1447 \left(\frac{4\hat{a}^2 T}{(1 + \hat{a})^2 (1 - \hat{a})^2} \right)^{1/3}, 1.1447 \left(\frac{4k^2 T}{(1 + k)^2 (1 - k)^2} \right)^{1/3} \right\},$$

where \hat{a} is the estimate of the autoregressive parameter given by Andrews’ method. Evidence drawn from simulations experiments leads us to specify $k=0.7$ or $k=0.8$ to establish a compromise between the empirical size and power of the test. Third, Sul et al. (2005) propose a prewhitened Heteroskedasticity and Autocorrelation Consistent (HAC) estimator for the long-run variance. In the first stage an AR model for the residuals $\{\hat{u}_t\}$ is estimated:

$$\hat{u}_t = \vartheta_1 \hat{u}_{t-1} + \dots + \vartheta_p \hat{u}_{t-p} + \psi_t. \tag{4}$$

After the estimation of Eq. 4 is carried out it is possible to obtain the long-run variance of the estimated residuals in Eq. 4, which is denoted as $\tilde{\sigma}_\psi^2$, through the application of an HAC estimator—for instance, Bartlett or Quadratic Spectral window—to monitor the presence of heteroskedasticity. In the second stage the estimated long-run variance is recolored:

$$\hat{\sigma}^2 = \frac{\tilde{\sigma}_\psi^2}{\tilde{\vartheta}(1)^2},$$

where $\tilde{\vartheta}(1)$ denotes the autoregressive polynomial $\tilde{\vartheta}(L) = 1 - \tilde{\vartheta}_1 L - \dots - \tilde{\vartheta}_p L^p$ evaluated at $L=1$. In order to avoid the inconsistency of the test statistic, Sul et al. (2005) suggest using the following boundary condition rule to obtain the long-run variance estimate:

$$\hat{\sigma}^2 = \min \left\{ T\tilde{\sigma}_\psi^2, \frac{\tilde{\sigma}_\psi^2}{\tilde{\vartheta}(1)^2} \right\}.$$

The application of this rule ensures that the estimated long-run variance is bounded above $T\tilde{\sigma}_\psi^2$. Now, the individual KPSS test statistic is of order $O_p(T)$ under the alternative hypothesis and diverges, so that the test is consistent. In the simulations we consider an AR(1) process for Eq. 4, and $\tilde{\sigma}_\psi^2$ is obtained using the Quadratic spectral window as depicted in Andrews (1991), Andrews and Monahan (1992) and Sul et al. (2005).

Regarding the parametric approach, Leybourne and McCabe (1994) specify the dynamics of $\{y_t\}$ in Eq. 1 using an AR(p) representation. Thus, the DGP of $\{y_t\}$ is given by:

$$\phi_p(L)y_t = \alpha_t + \beta t + u_t, \tag{5}$$

and Eq. 2, where the roots of $\phi_p(L)$ lie outside the unit circle. The LMC test is given by:

$$\hat{\xi}_i = \hat{\sigma}_u^{-2} T^{-2} \sum_{t=1}^T S_t^2 \tag{6}$$

$i=0,1,2$, where $\{\hat{u}_t\}$ is the OLS estimated residuals that are obtained after the filtered variable y_t^* is regressed against $f(t)$ and $\hat{\sigma}_u^2 = T^{-1} \sum_{t=1}^T \hat{u}_t^2$. We denote by y_t^* the variable once the AR(p) model has been filtered out, that is, $y_t^* = y_t - \sum_{j=1}^p \hat{\phi}_j y_{t-j}$, where $\hat{\phi}_j$ are the OLS estimated coefficients in the following over-differenced, second-order in moments equivalent, model:

$$\phi_p(L)\Delta y_t = g(t) + (1 - \theta L)\zeta_t, \tag{7}$$

where $g(t)=\Delta f(t)$ and $\zeta_t \sim \text{iid}(0, \sigma_\zeta^2)$ with $\sigma_\zeta^2 = \sigma_u^2 \theta^{-1}$. Under the null hypothesis $\theta=1$, while the alternative hypothesis implies $0 < \theta < 1$. As stated in Caner and Kilian (2001), Leybourne and McCabe (1994, 1999) estimate the model with the GAUSS-ARIMA Library but defining a grid for the specification of the starting value for θ . Specifically, the grid starts at $\theta = 0$ and ends at $\theta = 1$ with steps of 0.05. The starting values for the autoregressive parameters are set equal to $\phi_j = \theta + 0.1, j=1, \dots, p$. They select the set of parameter estimates that corresponds with the highest log-likelihood. McCabe and Leybourne (1998) showed that the ML estimation of θ in Eq. 7 is T -consistent and that the AR parameters are asymptotically equivalent to those obtained from the OLS estimation of Eq. 5 in levels.

The order of the autoregressive model in Eq. 7 can be chosen using two different procedures. Once a maximum number of lags (p_{\max}) has been specified, the first procedure uses an information criterion to select p . The Akaike (AIC) and Schwarz bayesian (SC) criteria are the most extended ones. In the simulation experiments we have also specified the modified Schwarz criterion recently proposed in Liu et al. (1997) (LWZ). The second approach, which is based on the general-to-specific principle, relies on the testing procedure described in Leybourne and McCabe (1999). In brief, let y_t be a stochastic process with the DGP given in Eq. 7. Then, under the null hypothesis of $H_0: \phi_p=0$ the test $Z(p) = T^{1/2}$

$\hat{\phi}_p \hat{\theta} \rightarrow_d N(0, 1)$, where $\hat{\phi}_p$ and $\hat{\theta}$ are estimated from Eq. 7.

Both the KPSS and LMC tests have the same limiting distributions, which are shown to belong to the family of the Cramér-von Mises distributions—see Harvey (2001). The asymptotic critical values can be found in Kwiatkowski et al. (1992) for the models $i=1, 2$ and Hobijn et al. (1998) for the model $i=0$, and finite sample critical values are available in Sephton (1995) for the models $i=1, 2$.

3 Monte Carlo evidence

In this section we conduct Monte Carlo simulations to assess how the stationarity tests described above behave in practice. The DGP that is used to analyse the empirical size is given by:

$$(1 - \phi_1 L)(1 - \phi_2 L)y_t = \varepsilon_t,$$

with $\varepsilon_t \sim \text{iid } N(0,1)$. The set of autoregressive parameters that has been specified is $\phi_1 = \{0, 0.5, 0.7, 0.8, 0.9, 0.95, 0.96, 0.97, 0.98, 0.99, 1\}$ and $\phi_2 = \{0, 0.5, -0.5\}$. Note that these values allow us to analyse both the empirical size and power of the test, since we allow one of the roots to be unity. The sample size has been set equal to $T = \{100, 300, 600\}$ and $n = 2,000$ replications for each experiment are conducted. The nominal size is set at the 5% level and the finite sample critical values are drawn from the response surfaces in Sephton (1995). The GAUSS codes that have been used in the simulations are available from the authors upon request.

The results of the simulations are reported in Tables 1 and 2 for an AR(1) process. The labels in the table are as follows: *K* for the rule in Kurozumi (2002)—with $k=0.7$ —*CA* for the rule in Choi and Ahn (1995, 1999), *SPC* for the rule in Sul et al. (2005), *Z(p)* for the procedure in Leybourne and McCabe (1999)—with a maximum order for the parametric correction of $p_{\max}=3$ and the 10% significance level for the lags to be included in the auxiliary regression (7)—*AIC* for the Akaike information criterion—with a maximum order for the parametric correction of $p_{\max}=3$ —and *known* when we assume that the order of the autoregressive model is $p=0$ for $\phi=0$ and $p=1$ elsewhere.¹

Looking at the results for the specification that includes a constant term in the regression, Table 1 shows that the SPC method is correctly sized except when the autoregressive parameter is very close to one. The use of the AIC criterion together with this procedure results in a size that is almost identical to the one where the order of the autoregressive model is known. All other tests are oversized with respect to the SPC method, especially the CA and LMC methods. Note that the value of $\phi=1$ gives us a first measure of the power of the tests. In this case the lower power of the SPC test can be explained by the size distortions of the others. Similar conclusions are reached irrespective of the deterministic specification (Table 2). Therefore, in order to save space, we continue the analysis of empirical size and power focusing on the methods with correct size, i.e. the KPSS test implementing the SPC method to estimate the long-run variance: KPSS (SPC).

Table 3 reports the rejections frequencies of the KPSS (SPC) test for AR(2) processes. Once again, the test has the correct size except when one of the roots is very close to one. Note that the test has virtually no power when a trend is included in the auxiliary regression, there is a negative root and the sample is not large enough (say $T=100$). Nevertheless, as the sample size grows the power increases.

Concerning the power of the test, in the previous tables we saw that KPSS (SPC) has a reasonable power, especially in large samples, for pure random walks and I(1) processes with positive autocorrelated shocks. Another DGP commonly used to study the power of the stationarity tests is given by the structural Eqs. 1–2 with $\sigma_v^2 > 0$ —see Kwiatkowski et al. (1992), Leybourne and McCabe (1994), and

¹ Similar results were obtained with the BIC and LWZ information criteria.

Table 1 Empirical size of the KPSS and LMC tests allowing for a constant term

T=100							
ϕ	KPSS				LMC		
	K	CA	SPC		Known	Z(p)	AIC
			Known	AIC			
0	0.054	0.054	0.057	0.058	0.061	0.059	0.064
0.5	0.060	0.060	0.041	0.044	0.069	0.068	0.074
0.7	0.084	0.083	0.045	0.031	0.115	0.118	0.128
0.8	0.078	0.089	0.029	0.032	0.153	0.147	0.185
0.9	0.146	0.412	0.042	0.031	0.317	0.325	0.487
0.95	0.232	0.697	0.180	0.176	0.530	0.495	0.743
0.96	0.283	0.745	0.236	0.264	0.649	0.556	0.801
0.97	0.313	0.811	0.306	0.310	0.638	0.639	0.867
0.98	0.388	0.838	0.394	0.422	0.678	0.683	0.881
0.99	0.442	0.887	0.456	0.471	0.722	0.713	0.904
1	0.537	0.879	0.561	0.597	0.743	0.708	0.907
T=300							
0	0.057	0.057	0.048	0.050	0.054	0.056	0.057
0.5	0.065	0.065	0.049	0.039	0.065	0.057	0.057
0.7	0.095	0.095	0.056	0.038	0.076	0.084	0.086
0.8	0.100	0.096	0.034	0.045	0.089	0.102	0.106
0.9	0.131	0.138	0.028	0.032	0.141	0.144	0.143
0.95	0.275	0.768	0.069	0.058	0.268	0.279	0.350
0.96	0.299	0.839	0.129	0.107	0.359	0.340	0.465
0.97	0.336	0.908	0.265	0.255	0.428	0.422	0.588
0.98	0.461	0.963	0.414	0.433	0.594	0.592	0.767
0.99	0.583	0.983	0.627	0.615	0.733	0.719	0.868
1	0.758	0.989	0.799	0.815	0.828	0.795	0.914
T=600							
0	0.054	0.054	0.049	0.040	0.056	0.054	0.054
0.5	0.059	0.059	0.048	0.050	0.046	0.054	0.053
0.7	0.081	0.081	0.051	0.041	0.064	0.075	0.070
0.8	0.064	0.061	0.060	0.035	0.060	0.054	0.055
0.9	0.123	0.083	0.051	0.046	0.085	0.085	0.084
0.95	0.214	0.550	0.040	0.033	0.134	0.140	0.133
0.96	0.262	0.796	0.049	0.058	0.167	0.174	0.180
0.97	0.334	0.927	0.111	0.113	0.293	0.238	0.274
0.98	0.457	0.980	0.359	0.334	0.446	0.411	0.508
0.99	0.640	0.994	0.658	0.659	0.654	0.657	0.777
1	0.828	1	0.903	0.904	0.874	0.876	0.935

AR(1) process. The DGP is given by $(1-\phi L)y_t = \varepsilon_t$, with $\varepsilon_t \sim \text{iid } N(0, 1)$. The maximum order for the parametric correction is set equal to $p_{\max}=3$ when using the $Z(p)$ and AIC information criterion. The column labelled as *known* specifies $p=1$ when $\phi>0$ and $p=0$ when $\phi=0$. Two thousand replications are conducted. The nominal size is set at the 5% level of significance and the critical values are from the response surfaces in Sephton (1995)

Table 2 Empirical size of the KPSS and LMC tests allowing for a time trend

T=100							
ϕ	KPSS				LMC		
	K	CA	SPC		Known	Z(p)	AIC
			Known	AIC			
0	0.047	0.047	0.059	0.060	0.056	0.061	0.060
0.5	0.083	0.083	0.037	0.047	0.058	0.059	0.062
0.7	0.074	0.074	0.029	0.031	0.060	0.061	0.078
0.8	0.078	0.078	0.023	0.023	0.068	0.097	0.120
0.9	0.126	0.359	0.014	0.020	0.209	0.188	0.348
0.95	0.265	0.624	0.064	0.082	0.355	0.329	0.580
0.96	0.236	0.648	0.105	0.090	0.394	0.370	0.626
0.97	0.283	0.685	0.131	0.149	0.441	0.416	0.668
0.98	0.299	0.732	0.155	0.163	0.484	0.449	0.712
0.99	0.345	0.749	0.183	0.206	0.482	0.454	0.725
1	0.331	0.756	0.196	0.198	0.495	0.477	0.725
T=300							
0	0.051	0.051	0.052	0.049	0.050	0.049	0.051
0.5	0.072	0.072	0.043	0.038	0.047	0.046	0.049
0.7	0.096	0.096	0.035	0.035	0.065	0.073	0.068
0.8	0.082	0.080	0.038	0.034	0.059	0.057	0.058
0.9	0.144	0.132	0.027	0.023	0.065	0.068	0.073
0.95	0.343	0.674	0.050	0.039	0.175	0.176	0.240
0.96	0.386	0.795	0.094	0.105	0.233	0.217	0.338
0.97	0.460	0.875	0.182	0.196	0.317	0.311	0.451
0.98	0.521	0.928	0.329	0.339	0.408	0.384	0.583
0.99	0.637	0.960	0.442	0.444	0.569	0.538	0.729
1	0.695	0.966	0.542	0.547	0.608	0.592	0.782
T=600							
0	0.050	0.050	0.044	0.049	0.051	0.048	0.049
0.5	0.076	0.076	0.037	0.055	0.065	0.064	0.067
0.7	0.065	0.065	0.046	0.045	0.060	0.060	0.059
0.8	0.088	0.088	0.046	0.038	0.053	0.053	0.052
0.9	0.137	0.089	0.032	0.028	0.065	0.060	0.058
0.95	0.292	0.489	0.036	0.031	0.077	0.080	0.070
0.96	0.360	0.740	0.056	0.030	0.115	0.109	0.111
0.97	0.454	0.909	0.140	0.093	0.194	0.186	0.190
0.98	0.607	0.973	0.316	0.295	0.366	0.338	0.411
0.99	0.732	0.993	0.624	0.593	0.549	0.510	0.638
1	0.834	0.997	0.891	0.774	0.709	0.680	0.801

AR(1) process. The DGP is given by $(1-\phi L)y_t = \varepsilon_t$, with $\varepsilon_t \sim \text{iid } N(0, 1)$. The maximum order for the parametric correction is set equal to $p_{\max}=3$ when using the Z(p) and AIC information criterion. The column labelled as *known* specifies $p=1$ when $\phi>0$ and $p=0$ when $\phi \geq 0$. Two thousand replications are conducted. The nominal size is set at the 5% level of significance and the critical values are from the response surfaces in Sephton (1995)

Table 3 Empirical size of the KPSS (SPC method)

Panel A: Constant term						
ϕ_1	$\phi_2=0.5$			$\phi_2=-0.5$		
	T			T		
	100	300	600	100	300	600
0	0.041	0.043	0.046	0.057	0.046	0.053
0.5	0.034	0.046	0.047	0.054	0.053	0.050
0.7	0.031	0.040	0.042	0.050	0.047	0.044
0.8	0.051	0.041	0.040	0.027	0.043	0.047
0.9	0.331	0.092	0.040	0.020	0.029	0.048
0.95	0.621	0.485	0.272	0.047	0.018	0.025
0.96	0.695	0.607	0.424	0.088	0.028	0.023
0.97	0.743	0.720	0.635	0.129	0.071	0.024
0.98	0.785	0.858	0.826	0.193	0.191	0.114
0.99	0.826	0.917	0.945	0.263	0.358	0.374
1	0.861	0.970	0.989	0.377	0.628	0.783
Panel B: Time trend						
0	0.044	0.041	0.046	0.077	0.056	0.052
0.5	0.036	0.039	0.046	0.094	0.053	0.049
0.7	0.036	0.039	0.041	0.047	0.042	0.043
0.8	0.052	0.038	0.041	0.027	0.038	0.050
0.9	0.290	0.084	0.034	0.022	0.029	0.030
0.95	0.534	0.492	0.321	0.021	0.013	0.024
0.96	0.577	0.637	0.515	0.025	0.015	0.023
0.97	0.602	0.762	0.699	0.036	0.033	0.019
0.98	0.648	0.837	0.869	0.042	0.078	0.065
0.99	0.691	0.892	0.951	0.046	0.190	0.272
1	0.698	0.920	0.977	0.055	0.260	0.499

AR(2) process. The DGP is given by $(1-\phi_1L)(1-\phi_2L)y_t = \varepsilon_t$, with $\varepsilon_t \sim iid N(0, 1)$. The order for the parametric correction is set equal to $p_{max}=3$ using the AIC information criterion to estimate the order of the autoregressive model. Two thousand replications are conducted. The nominal size is set at the 5% level of significance and the critical values are from the response surfaces in Sephton (1995)

Lee (1996) among others. Here in the experiments we set $\beta=0$ and $u_t = \varepsilon_t \sim iid N(0, \sigma_\varepsilon^2)$. Note that in this case the reduced form of the DGP can be written as:

$$(1 - L)y_t = (1 - \tau L)\eta_t,$$

where η_t is a white noise process, $\tau = 1 + \frac{1}{2}r - \frac{1}{2}\sqrt{(r+4)r}$ and $r = \sigma_v^2/\sigma_\varepsilon^2$ is the signal-to-noise ratio. Therefore, y_t is an IMA(1,1) process where the MA parameter depends on the signal-to-noise ratio $\sigma_v^2/\sigma_\varepsilon^2$. Table 4 shows the empirical power of the KPSS (SPC) test for several values of the signal-to-noise ratio $\sigma_v^2/\sigma_\varepsilon^2$. For low values of $\sigma_v^2/\sigma_\varepsilon^2$ the corresponding MA parameter τ is very close to one, so there is quasi-cancellation of the roots and the process looks similar to a white noise process in finite samples. For these values, the power of the KPSS (SPC) does not

Table 4 Power of the KPSS (*SPC* method)

Panel A: Constant term				
$\sigma_v^2/\sigma_\varepsilon^2$	$T=100$	$T=300$	$T=600$	τ
0.0001	0.057	0.164	0.372	0.990
0.001	0.155	0.553	0.821	0.969
0.01	0.454	0.584	0.440	0.905
0.05	0.328	0.215	0.273	0.800
0.1	0.208	0.255	0.381	0.730
0.5	0.259	0.500	0.658	0.500
1	0.376	0.591	0.754	0.382
2	0.442	0.682	0.827	0.268
5	0.538	0.744	0.866	0.146
10	0.550	0.783	0.897	0.084
Panel B: Time trend				
0.0001	0.055	0.072	0.183	0.990
0.001	0.082	0.335	0.681	0.969
0.01	0.278	0.745	0.714	0.905
0.05	0.433	0.334	0.178	0.800
0.1	0.347	0.165	0.104	0.730
0.5	0.108	0.139	0.316	0.500
1	0.106	0.239	0.451	0.382
2	0.118	0.347	0.585	0.268
5	0.142	0.466	0.696	0.146
10	0.170	0.524	0.751	0.084

increase monotonically. When the signal-to-noise ratio is greater than 0.05, the power rises monotonically and can be considered adequate.

4 Additional evidence concerning the purchasing power parity hypothesis

Empirical economic research has devoted increasing interest to testing the PPP hypothesis. Some relevant references in the field are Meese and Rogoff (1988), Abuaf and Jorion (1990), Froot and Rogoff (1995), Rogoff (1996), Oh (1996) and Papell (1997), to mention a few. All these contributions use unit root or cointegration techniques, in a time series or panel data framework, in order to assess whether PPP holds. In this paper we contribute to the evidence available for this topic using the stationarity tests described above.

Stationarity tests have been extensively applied in PPP hypothesis testing, mainly because PPP appears as the null hypothesis of these tests—see, for instance, Corbae and Ouliaris (1991). Taylor (2001) indicates that the PPP hypothesis should be tested using a procedure that specifies the null hypothesis of stationarity, because this hypothesis is well established in the economist's priors. Some authors adopt this idea and rely on stationarity statistics to test the PPP hypothesis. For

instance, Culver and Papell (1999) used the KPSS test and found evidence in favour of the PPP in the post-Bretton Woods era with quarterly data. Engel (2000) uses the KPSS test when testing for the PPP, although he warns of the low power found in the simulations. Finally, Caner and Kilian (2001) also apply the KPSS and LMC tests to the PPP, but advising of the size distortion problems.

We have shown that the application of stationarity tests is not exempt from problems, since size distortions can lead to erroneous conclusions. Fortunately, these pitfalls can be mitigated if the long-run variance is properly estimated. In this section we follow the recommendations given in the previous section and apply the KPSS test to two different sets of time series using the procedure developed in Sul et al. (2005) to estimate the long-run variance. The first data set refers to the 19 countries considered in Taylor (2002). These data allow us to compute annual real exchange rate for Argentina (1884), Australia (1870), Belgium (1880), Brazil (1889), Canada (1870), Denmark (1880), Finland (1881), France (1880), Germany (1880), Italy (1880), Japan (1885), Mexico (1886), Netherlands (1870), Norway (1870), Portugal (1890), Spain (1880), Sweden (1880), Switzerland (1892) and the United Kingdom (1870) using the United States as the numerarie—see Taylor (2002) for additional details on the data sources. The starting date of the time series is irregular—the first year of the time period is given between parentheses above—but all of them end at 1996. Therefore, the data cover a century. The second data set is given by the post-Bretton Woods era monthly time series drawn from the *International Financial Statistics* database of the IMF, which covers the period from January 1973 to December 1998. This data set covers Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Japan, Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland and United Kingdom.

Table 5 reports the results for the ADF-GLS, modified tests—modified Sargan–Bharvaga (MSB) and modified Z_t test (MZt)—in Ng and Perron (2001), and the KPSS (SPC) statistic. Although the model that includes the constant term is the deterministic specification that is consistent with the PPP hypothesis, we have also computed the test allowing for a time trend. As pointed out in Taylor (2002), this follows the spirit of the Balassa–Samuelson effect. The order of the autoregressive correction is selected using the modified AIC information criteria (MAIC) proposed in Ng and Perron (2001). These results indicate that the null hypothesis of unit root cannot be rejected by either unit root tests for Canada, Denmark, France, Japan, Norway and Portugal. On the other hand, we find evidence in favour of the PPP for 13 countries, although for Brazil, Mexico and Switzerland we require the specification of a time trend. There are some discrepancies between Taylor's results and ours, since he concludes that the PPP has held in the long run over the twentieth century for his sample of 20 countries. These contradictions might be due to the use of the MAIC instead of the Lagrange multiplier as the lag-length selection procedure. The results for Canada and France, where the order of the autoregressive correction selected by the MAIC criterion is large, are of particular interest.

When looking at the KPSS (SPC) statistic where the specification is given by a constant term, the PPP hypothesis cannot be rejected at the 5% level for 12 countries. If the deterministic component is given by a time trend, the KPSS statistic finds evidence in favour of the PPP at the 5% level for 18 countries. Taking

Table 5 Unit root and stationarity tests for the annual PPP

	Constant					Time trend				
	<i>T</i>	MSB	MZt	ADF-GLS	<i>p</i> KPSS (SPC)	MSB	MZt	ADF-GLS	<i>p</i> KPSS (SPC)	
Argentina	113	0.181 ^b	-2.583 ^a	-2.594 ^a	4 0.085	0.162 ^b	-2.866	-2.764	4 0.046	
Australia	127	0.214 ^b	-2.336 ^b	-2.447 ^b	0 0.352	0.196	-2.534	-2.691	0 0.093	
Belgium	117	0.182 ^b	-2.567 ^b	-2.473 ^b	3 0.520 ^b	0.145 ^b	-3.437 ^a	-3.885 ^a	0 0.058	
Brazil	108	0.580	-0.369	-0.480	4 1.293 ^a	0.166 ^b	-2.907	-2.339	5 0.084	
Canada	127	0.302	-1.566	-1.613	0 0.278	0.230	-1.815	1-1.931	0 0.113	
Denmark	117	0.285	-1.179	-1.026	6 0.151	0.242	-1.901	1-1.762	6 0.114	
Finland	116	0.131 ^a	-3.786 ^a	-4.444 ^a	0 0.233	0.127 ^a	-3.938 ^a	-4.692 ^a	0 0.083	
France	117	0.448	-1.065	-1.120	6 0.446	0.209	-2.182	-2.108	6 0.201 ^b	
Germany	117	0.206 ^b	-2.168 ^b	-2.148 ^b	2 0.181	0.170	-2.918 ^b	-2.850	2 0.032	
Italy	117	0.165 ^a	-3.011 ^a	-3.295 ^a	0 0.067	0.165 ^b	-3.020 ^b	-3.307 ^b	0 0.059	
Japan	112	0.451	-0.021	-0.230	1 3.063 ^a	0.204	-2.363	-2.340	2 0.111	
Mexico	111	0.309	-1.359	-1.305	6 0.321	0.141 ^a	-3.539 ^a	-4.073 ^a	0 0.036	
Netherlands	127	0.178 ^b	-2.494 ^b	-2.523 ^b	1 0.161	0.216	-2.231	-2.330	0 0.093	
Norway	127	0.304	-1.174	-1.111	5 0.229	0.192	-2.588	1-2.506	5 0.100	
Portugal	107	0.354	-1.411	-1.336	6 0.222	0.242	-1.745	1-1.709	4 0.102	
Spain	117	0.232 ^b	-2.141 ^b	-2.095 ^b	3 0.194	0.190	-2.498	-2.422	3 0.100	
Sweden	117	0.198 ^b	-2.312 ^b	-2.301 ^b	2 0.365	0.159 ^b	-3.154 ^b	-3.484 ^a	0 0.056	
Switzerland	105	0.334	-0.519	-0.655	2 0.812 ^a	0.175	-2.741	-2.732	2 0.075	
UK	127	0.203 ^b	-2.333 ^b	-2.297 ^b	4 0.204	0.193	-2.534	-2.474	4 0.067	
Critical values					Critical values					
cv (1%)	0.174	-2.58	-2.58	0.734	0.143	-3.42	-3.42	0.216		
cv (5%)	0.233	-1.98	-1.98	0.466	0.168	-2.91	-2.91	0.148		
cv (10%)	0.275	-1.62	-1.62	0.349	0.185	-2.62	-2.62	0.120		

The U.S. are chosen as the numeraire. The order of the autoregressive correction has been selected using the modified AIC criteria (MAIC) with the maximum set equal to $p_{max}=7$. The superscripts *a*, *b* and *c* denote statistical significance using asymptotic critical values at the 1, 5 and 10% level of significance, respectively, for the unit root statistics. For the KPSS statistic we have used the AIC criterion to select the order of the autoregressive correction with $p_{max}=7$. Regarding the critical values, we report the finite sample critical values drawn from the response surfaces in Sephton (1995). Since the sample size varies, we have decided to report the critical values for the minimum sample size ($T=105$), although similar qualitatively results are obtained when using the particular finite sample critical values for each time serie

into account both the results of the unit root and stationarity tests, we have found evidence in favour of the PPP hypothesis for Argentina, Finland, Germany, Italy, Netherlands, Spain and UK when the deterministic specification is given by a constant term. When allowing for a time trend, we conclude that the PPP holds for Argentina, Belgium, Brazil, Finland, Germany, Italy, Mexico, Sweden and Switzerland. In all, evidence in favour of the PPP hypothesis has been found for 12 countries. The real exchange rates for France, Japan, and Switzerland are characterised as non-stationary for the unit root and stationarity tests when using the constant term as the deterministic component. France is still classified as non-

stationary when the time trend is allowed. For the other countries, the application of the tests results in inconclusive situations.

Therefore, our results contradicts those in Taylor (2002), because while he found evidence in favour of the PPP hypothesis—using either the constant or the time trend—we can only conclude that the PPP holds just for 12 of the 19 countries studied.

The picture changes when analysing the monthly data set (see Table 6). In this case we can find evidence in favour of the PPP hypothesis for Ireland and (mild evidence for) Italy when using the deterministic specification given by a constant term, and for New Zealand when using the time trend. Evidence of non-stationarity is found for Australia, Austria, Belgium, Canada, Denmark, Japan, Portugal, Spain, Sweden and Switzerland when specifying a constant term. The evidence against the PPP hypothesis increases when using the time trend since non-stationarity behaviour is found for Austria, Belgium, Canada, Denmark, Finland, France, Germany, Greece, Italy, Norway, Portugal, Spain and Sweden. For the other cases contradictory conclusions are found.

To sum up, the evidence in favour of the PPP hypothesis depends on the sample period. For the large data set covering a century we found strong evidence in favour of this hypothesis (both types of test coincide in classifying the series as stationary) for: Argentina, Finland, Germany, Italy, Netherlands, Spain and UK and, if a trend is included, for: Belgium, Sweden and Switzerland. Moreover, for this span there is strong evidence against the PPP for Brazil, France, Japan and Switzerland when no trend is allowed but only for France when one is. For the monthly data set covering the last quarter of the century, there is strong evidence favouring the hypothesis only for two countries (Italy and Ireland) for the specification without time trend and New Zealand if this term is included. Strong evidence against the PPP is found for (no time trend in the auxiliary regression): Australia, Austria, Belgium, Canada, Denmark, Japan, Portugal, Spain, Sweden and Switzerland, and, if there is a trend, for Austria, Belgium, Canada, Denmark, Finland, France, Germany, Greece, Italy, Norway, Portugal, Spain and Sweden. Finally, it should be mentioned that contradictory results for some countries for these two data sets and periods could be due to a structural change—see Cheung and Chinn (1997). In this respect our results indicate that those studies aiming to analyse the PPP hypothesis allowing for structural breaks using, for instance, the approach in Kurozumi (2002), should pay special attention to the estimation of the long-run variance in order to control size distortions.

5 Conclusion

The paper has shown the performance of the different choices that an analyst can make when computing stationarity tests. Although the scope of the paper has not covered all the possible stationarity tests that are available in the econometric literature, we have focused on the two tests most widely used, i.e. the KPSS and the LMC tests. The aim of the paper is to after advice on the computation of these tests, assessing the consequences of the different choices on the selection of the spectral bandwidth or the lag length. We conclude that the KPSS test that estimates the long-run variance using the proposal in Sul et al. (2005) is the best choice since this combination shows less size distortion than the LMC test and has reasonable

Table 6 Unit root and stationarity tests for the monthly PPP (1973.1–1998.12)

	Constant				Time trend					
	MSB	MZt	ADF-GLS	<i>p</i> KPSS (SPC)	MSB	MZt	ADF-GLS	<i>p</i> KPSS (SPC)		
Australia	0.331	0.099	0.095	0	1.494 ^a	0.217	-1.766	-1.763	2	0.093
Austria	0.463	-0.837	-0.844	0	0.623 ^b	0.287	-1.705	-1.725	0	0.206 ^b
Belgium	0.339	-1.458	-1.469	0	0.357 ^c	0.320	-1.555	-1.569	0	0.326 ^a
Canada	0.420	0.162	0.353	13	2.443 ^a	0.198	-2.315	-1.793	13	0.293 ^a
Denmark	0.402	-1.080	-1.088	0	0.371 ^c	0.300	-1.651	-1.668	0	0.248 ^a
Finland	0.462	-1.046	-1.054	0	0.134	0.324	-1.389	-1.406	0	0.126 ^c
France	0.308	-1.591	-1.606	0	0.207	0.275	-1.804	-1.825	0	0.207 ^b
Germany	0.305	-1.607	-1.622 ^c	0	0.282	0.281	-1.776	-1.795	0	0.283 ^a
Greece	0.247 ^c	-1.877 ^c	-1.856 ^c	14	0.465 ^b	0.202	-2.465	-2.368	14	0.364 ^a
Ireland	0.261 ^c	-1.904 ^c	-1.927 ^c	0	0.065	0.246	-2.021	-2.049	0	0.055
Italy	0.273 ^c	-1.825 ^c	-1.846 ^c	0	0.450 ^c	0.266	-1.873	-1.895	0	0.150 ^b
Japan	0.667	-0.243	-0.232	1	2.912 ^a	0.273	-1.705	-1.725	0	0.099
Netherlands	0.308	-1.582	-1.599	0	0.121	0.229	-2.130	-2.168	0	0.065
New Zealand	0.214 ^b	-1.369	-1.011	15	0.374 ^c	0.146 ^b	-3.171 ^a	-2.381	15	0.046
Norway	0.358	-1.389	-1.401	0	0.125	0.269	-1.719	-1.741	0	0.125 ^c
Portugal	0.374	-1.097	-1.105	0	0.850 ^a	0.303	-1.647	-1.662	0	0.495 ^a
Spain	0.514	-0.794	-0.801	0	0.660 ^b	0.330	-1.434	-1.447	0	0.189 ^b
Sweden	0.303	-1.412	-1.423	0	0.489 ^b	0.292	-1.650	-1.666	0	0.251 ^a
Switzerland	0.521	-0.655	-0.663	0	0.674 ^b	0.269	-1.786	-1.809	0	0.108
United Kingdom	0.317	-1.293	-1.305	0	0.291	0.221	-2.260	-2.299	0	0.053
	Critical values				Critical values					
cv (1%)	0.174	-2.58	-2.58	0.741	0.143	-3.42	-3.42	0.217		
cv (5%)	0.233	-1.98	-1.98	0.463	0.168	-2.91	-2.91	0.148		
cv (10%)	0.275	-1.62	-1.62	0.348	0.185	-2.62	-2.62	0.120		

The US is chosen as the numeraire. The order of the autoregressive correction has been selected using the modified AIC criteria (MAIC), with the maximum set equal to $p_{\max} = \text{int}[12(T/100)^{(1/4)}]$. For the KPSS statistic we have used the AIC criterion to select the order of the autoregressive correction with $p_{\max} = \text{int}[12(T/100)^{(1/4)}]$. Regarding the critical values, we report the finite sample critical values drawn from the response surfaces in Sephton (1995)

^a Denotes statistical significance using asymptomatic critical values at the 1% level of significance for the unit root statistics

^b Denotes statistical significance using asymptomatic critical values at the 5% level of significance for the unit root statistics

^c Denotes statistical significance using asymptomatic critical values at the 10% level of significance for the unit root statistics

power. Therefore, the inference on the stochastic properties of the time series should be based on the non-parametric approach when computing stationarity tests.

Finally, our findings are extensive to other fields of applications of the stationarity tests. For instance, our suggestions can be taken into account when computing the KPSS test in the seasonal framework—see Lyhagen (2000) and Phillips and Jin (2002)—when dealing with fractional integration—see Lee and

Schmidt (1996) and Amsler (1999)—when testing the stationarity allowing for structural breaks—see Carrion-i-Silvestre and Sansó (2004), Lee and Strazicich (2001), Busetti and Harvey (2001) and Kurozumi (2002)—or when testing for cointegration—see Harris and Inder (1994), Shin (1994), Bartley et al. (2001) and Carrion-i-Silvestre and Sansó (2005).

Acknowledgements We thank A. M. Taylor for providing us with the data, and the Editor, B. H. Baltagi, and two anonymous referees for helpful comments. Financial support is acknowledged from the Ministerio de Ciencia y Tecnología under grant SEJ2005-08646/ECON and SEJ2005-07781/ECON and Conselleria d'Economia, Hisenda i Innovació del Govern Balear under PRIB-2004-10095, respectively.

References

- Abuaf N, Jorion P (1990) Purchasing power parity in the long run. *J Finance* 45:157–174
- Amano R, Van Norden S (1992) Unit-root tests and the burden of proof. Discussion paper, Bank of Canada
- Amsler C (1999) Size and power: lower tail KPSS tests and anti persistent alternatives. *Appl Econ Lett* 6:693–695
- Andrews DWK (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59:817–858
- Andrews DWK, Monahan JC (1992) An improved heteroskedasticity and autocorrelation consistent autocovariance matrix. *Econometrica* 60:953–966
- Bartley WA, Lee J, Strazicich MC (2001) Testing the null of cointegration in the presence of a structural break. *Econ Lett* 73: 315–323
- Busetti F, Harvey A (2001) Testing for the presence of a random walk in series with structural breaks. *J Time Ser Anal* 22: 127–150
- Caner M, Kilian L (2001) Size distortions of tests of the null hypothesis of stationarity: evidence and implications for the PPP debate. *J Int Money Financ* 20:639–657
- Carrion-i-Silvestre JL, Sansó A (2004) The KPSS test with two structural breaks. Discussion paper, Department of Econometrics, Statistics and Spanish Economy. University of Barcelona
- Carrion-i-Silvestre JL, Sansó A (2005) Testing the null of cointegration with structural breaks. *Oxford Bulletin of Economics and Statistics* Forthcoming.
- Carrion-i-Silvestre JL, Sansó A, Artís M (2001) Unit root and stationarity tests' wedding. *Econ Lett* 70:1–8
- Charemza WW, Syczewska EM (1998) Joint application of the Dickey–Fuller and KPSS tests. *Econ Lett* 61:17–21
- Chen MY (2002) Testing stationarity against unit roots and structural changes. *Appl Econ Lett* 9:459–464
- Cheung YW, Chinn MD (1997) Further investigation of the uncertain unit root in GNP. *J Bus Econ Stat* 15:68–73
- Choi I, Ahn BC (1995) Testing for cointegration in a system of equations. *Econom Theory* 11: 952–983
- Choi I, Ahn BC (1999) Testing the null of stationarity for multiple time series. *J Econom* 88: 41–77
- Corbae D, Ouliaris S (1991) A test of long-run purchasing power parity allowing for structural breaks. *Econ Rec* 67:26–33
- Culver SE, Papell DH (1999) Long-run purchasing power parity with short-run data: evidence with a null hypothesis of stationarity. *J Int Money Financ* 18:751–768
- Engel C (2000) Long-run PPP may not hold after all. *J Int Econ* 57:243–273
- Froot KA, Rogoff K (1995) Perspectives on PPP and long-run real exchange rates. *Handbook of international economics*, vol III. Grossman E, Rogoff K (Eds.). Elsevier, North-Holland, pp. 1647–1688
- Gabriel VJ (2003) Cointegration and the joint confirmation hypothesis. *Econ Lett* 78:17–25
- Harris D, Inder B (1994) A test of the null hypothesis of cointegration. In: Hargreaves CP (ed) *Nonstationary time series analysis and cointegration*, Chap. 5:133–152. Oxford University Press

- Harvey AC (2001) A unified approach to testing for stationarity and unit roots. Discussion paper, Faculty of Economics and Politics. University of Cambridge
- Hobijn B, Franses PHB, Ooms M (1998) Generalizations of the KPSS-test for stationarity. Discussion Paper 9802, Econometric Institute. Erasmus University Rotterdam
- Kurozumi E (2002) Testing for stationarity with a break. *J Econom* 108:63–99
- Kwiatkowski D, Phillips PCB, Schmidt PJ, Shin Y (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. *J Econom* 54:159–178
- Lee J (1996) On the power of stationarity tests using optimal bandwidth estimates. *Econ Lett* 51:131–137
- Lee J, Schmidt P (1996) On the power of the KPSS test of stationarity against fractionally-integrated alternatives. *J Econom* 73:285–302
- Lee J, Strazicich M (2001) Testing the null of stationarity in the presence of one structural break. *Appl Econ Lett* 8:377–382
- Leybourne SJ, McCabe BPM (1994) A consistent test for a unit root. *J Bus Econ Stat* 12:157–166
- Leybourne SJ, McCabe BPM (1999) Modified stationarity tests with data-dependent model selection rules. *J Bus Econ Stat* 17:264–270
- Liu J, Wu S, Zidek JV (1997) On segmented multivariate regressions. *Stat Sin* 7:497–525
- Lyhagen J (2000) The seasonal KPSS statistic. Discussion paper, Department of Economic Statistics. Stockholm School of Economics
- Maddala GS, Kim IM (1998) Unit roots, cointegration and structural change. Cambridge
- McCabe BPM, Leybourne SJ (1998) On estimating an ARMA model with a MA unit root. *Econom Theory* 14:326–338
- Meese R, Rogoff K (1988) What is real? The exchange rate interest differential relation over the modern floating rate period. *J Finance* 43:933–948
- Nabeya S, Tanaka K (1988) Asymptotic theory of a test for the constancy of regression coefficients against the random walk alternative. *Ann Stat* 16:218–235
- Newey WK, West KD (1994) Automatic lag selection in covariance matrix estimation. *Rev Econ Stud* 61:631–653
- Ng S, Perron P (2001) Lag length selection and the construction of unit root tests with good size and power. *Econometrica* 69:1519–1554
- Oh KY (1996) Purchasing power parity and unit root tests using panel data. *J Int Money Financ* 15:405–418
- Papell D (1997) Searching for stationarity: purchasing power parity under the current float. *J Int Econ* 43:313–332
- Phillips PCB (1987) Time series regression with a unit root. *Econometrica* 55:277–301
- Phillips PCB, Jin S (2002) The KPSS test with seasonal dummies. *Econ Lett* 77:239–243
- Phillips PCB, Perron P (1988) Testing for a unit root in time series regression. *Biometrika* 75:335–346
- Rogoff K (1996) The purchasing power parity puzzle. *J Econ Lit* 34:647–468
- Sephton PS (1995) Response surfaces estimates of the KPSS stationary test. *Econ Lett* 47:255–261
- Shin Y (1994) A residual-based test of the null of cointegration against the alternative of no cointegration. *Econom Theory* 10:91–115
- Sul D, Phillips PCB, Choi CY (2005) Prewhitening Bias in HAC Estimation. *Oxf Bull Econ Stat* 67:517–546
- Taylor AM (2001) Potential pitfalls for the purchasing-power—parity puzzle? Sampling and specification biases in mean-reversion tests of the law of one price. *Econometrica* 69:473–498
- Taylor AM (2002) A century of purchasing—power parity. *Rev Econ Stat* 84:139–150