

Herwartz, Helmut; Siedenburg, Florian

Working Paper

A new approach to unit root testing

Economics Working Paper, No. 2009-06

Provided in Cooperation with:

Christian-Albrechts-University of Kiel, Department of Economics

Suggested Citation: Herwartz, Helmut; Siedenburg, Florian (2009) : A new approach to unit root testing, Economics Working Paper, No. 2009-06, Kiel University, Department of Economics, Kiel

This Version is available at:

<https://hdl.handle.net/10419/28394>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

a new approach to unit root testing

by Helmut Herwartz and Florian Siedenburg



A new approach to unit root testing

Helmut Herwartz and Florian Siedenburg

Christian-Albrechts-University of Kiel

This draft: February 2009

Abstract A novel simulation based approach to unit root testing is proposed in this paper. The test is constructed from the distinct orders in probability of the OLS parameter estimates obtained from a spurious and an unbalanced regression, respectively. While the parameter estimate from a regression of two integrated and uncorrelated time series is of order $O_p(1)$, the estimate is of order $O_p(T^{-1})$ if the dependent variable is stationary. The test statistic is constructed as an inter quantile range from the empirical distribution obtained from regressing the standardized data sufficiently often on controlled random walks. GLS detrending (Elliott et al., 1996) and spectral density variance estimators (Perron and Ng, 1998) are applied to account for deterministic terms and residual autocorrelation in the data. A Monte Carlo study confirms that the proposed test has favorable empirical size properties and is powerful in local-to-unity neighborhoods. Testing for PPP for a sample of G6 economies, the proposed test yields results in favor of PPP for half of the sample economies while benchmark tests obtain at most one rejection of the random walk null hypothesis.

JEL Classification: C22, C12

Keywords: Unit root tests, simulation based test, simulation study, GLS detrending.

Corresponding author: F. Siedenburg, Institute for Statistics and Econometrics, Christian-Albrechts-University of Kiel, Olshausenstr. 40-60, D-24118 Kiel, Germany.

Email: fsiedenb@economics.uni-kiel.de. Phone: +49-431-8802225.

1 Introduction

Since the work of Granger and Newbold (1974) it is known that spurious correlations may arise if a least squares regression is fitted to uncorrelated time series which are integrated (at least) of order one ($I(1)$). To avoid this, separating between stationary and integrated series by means of unit root tests is a central aspect of time series econometrics. Dickey and Fuller (1979) (DF henceforth) show that for $I(1)$ processes, the t -ratio from a first order autoregression converges to a nonstandard limiting distribution which can be expressed as a functional of a Brownian motion. Accordingly, the DF unit root test is conducted by comparing this t -ratio with simulated critical values drawn from the limiting distribution. Since then, the literature on unit root testing has been rapidly expanding. Major issues involve coping with residual autocorrelation (Said and Dickey, 1984, Phillips and Perron, 1988) and improving the power features of the tests (e.g. Elliott et al., 1996). An alternative approach to unit root testing has been proposed by Stock (1999)¹. Instead of directly testing the value of the autoregressive parameter, the so-called class of M type tests exploits the fact that the sum of squares of an integrated process is of higher order in probability than the sum of squares of a stationary process. Perron and Ng (1996) and Ng and Perron (2001) suggest modified variants of the M tests which perform well in terms of small type one errors under general forms of residual autocorrelation while retaining good power properties. Fully Nonparametric approaches to unit root testing which are robust against violations of standard assumptions have been proposed e.g. by Breitung and Gouriéroux (1997) and Aparico et al. (2006).

In this paper we present a novel approach to unit root testing. We start by noting that the parameter from a spurious regression converges to a nondegenerated limiting distribution (Phillips, 1986). In contrast, the parameter of an unbalanced regression of an $I(0)$ variable on an $I(1)$ regressor can be shown to converge towards zero at the rate $O_p(T^{-1})$. We show that a consistent unit root test can be based on this distinction. In particular, regressing the appropriately scaled data sufficiently

¹The paper dates back to 1990.

often on a random walk controlled by the analyst yields a sample of random variables from the limiting distribution derived by Phillips (1986). Viable test statistics can then be constructed from ranges of that random variable, which have a nondegenerated distribution under H_0 but degenerate to a one point distribution under H_1 . We conduct a simulation study to assess the empirical properties of the proposed procedure. To preview the results, it turns out that the simulation based testing approach on average offers most precise size estimates compared with ADF- and M -type tests. In large samples, the proposed test achieves higher local power than the standard ADF test but is outperformed by the ADF^{GLS} and the M test. However, there are finite sample scenarios with residual autocorrelation where the proposed test yields the highest power estimates among those tests that are characterized by correct empirical rejection frequencies under H_0 . As an empirical illustration we test for long run PPP in a sample of G6 economies. Two variants of the proposed test yield up to three rejections of the unit root null hypothesis whereas the benchmark tests obtain at most one rejection.

The remainder of the paper proceeds as follows: the testing idea is presented in Section 2. The Monte Carlo simulations are documented in Section 3. Section 4 contains the empirical application. Finally, Section 5 concludes and discusses further potential applications of the proposed test procedure.

2 The simulation based range unit root test

Consider the issue of testing for a unit root ($\rho = 1$) in the time series $\{y_t\}$, generated by

$$y_t = d_t + u_t, \quad u_t = \rho u_{t-1} + v_t, \quad t = 1, \dots, T, \quad (1)$$

where $d_t = z_t' \psi$ and z_t is a vector of deterministic components and ψ is a parameter vector. In (1), the error term $v_t = \sum_{j=0}^{\infty} \delta_j e_{t-j}$ with $\sum_{j=0}^{\infty} j |\delta_j| < \infty$ allows for general forms of serial correlation and $\{e_t\}$ is white noise with variance σ_e^2 .

Stock (1999) proposed the so-called class of M -type unit root and cointegration tests, which are basically derived from the observation, that the sum of squares of

time series of order $I(1)$ and $I(0)$ converge at different rates, namely $O_p(T^2)$, and $O_p(T)$, respectively. Hence, suitable standardizations can be applied to consistently discriminate between $I(1)$ and $I(0)$ processes. For instance, Stock (1999) suggests the statistic

$$MZ_\alpha = (y_T^2 - Ts^2) \left(2T^{-1} \sum_{t=1}^T y_t^2 \right)^{-1}, \quad (2)$$

where s^2 is a consistent estimate of the long run variance

$$\sigma^2 = \lim_{T \rightarrow \infty} T^{-1} E \left(\sum_{t=1}^T v_t^2 \right).$$

2.1 The testing principle

Our proposed simulation based approach is similar to the M -type statistics, in that we try to discriminate between integrated and stationary processes by means of their stochastic order rather than based on an autoregressive parameter estimate as e.g. in the widely used DF test. However, unlike the M -type tests of Stock (1999), the proposed test is based on the stochastic order of the slope parameter from a *spurious regression*. Since the work of Granger and Newbold (1974) and Phillips (1986) it is well known that the OLS estimator $\hat{\beta}$ from the regression

$$y_t = \beta x_t + \varepsilon_t, \quad (3)$$

is not consistent for the true parameter $\beta = 0$ if y_t and x_t are two uncorrelated random walks. To make statements about the limiting distribution of $\hat{\beta}$, we make the following assumptions

Assumption 1 (\mathcal{A}_1) *Let $x_t = x_{t-1} + \epsilon_t$ and y_t generated as in (1) with $d_t = 0$ and $v_t = e_t$. The innovations ϵ_t and e_t are independent white noise, i.e. $\epsilon_t \sim iid(0, \sigma_\epsilon^2)$, $e_t \sim iid(0, \sigma_e^2)$ and $Cov[\epsilon_t, e_t] = 0$.*

Assumption 2 (\mathcal{A}_2) *Initial values are given by $y_0 = x_0 = 0$.*

Under the unit root null hypothesis $\rho = 1$ and (\mathcal{A}_1) - (\mathcal{A}_2) , it follows from the results in Phillips (1986) that the limiting distribution of $\hat{\beta} = (\sum_{t=1}^T x_t^2)^{-1} \sum_{t=1}^T x_t y_t$ is given

by

$$\widehat{\beta} \xrightarrow{d} \sigma_e/\sigma_v \left[\int_0^1 W_x(a)^2 da \right]^{-1} \int_0^1 W_x(a) W_y(a) da, \quad (4)$$

which implies that $\widehat{\beta} \sim O_p(1)$. If $\{y_t\}$ is a weakly stationary process under the alternative hypothesis, $H_1 : \rho < 1$, $\widehat{\beta} \xrightarrow{p} 0$ at the rate of T and, hence, $\widehat{\beta} \sim O_p(T^{-1})$. The testing idea is to exploit the distinct orders in probability under H_0 and H_1 , respectively. In the unit root case, the non-degenerated limiting distribution implies some finite range for a sample of R realizations of $\{\widehat{\beta}_r\}_{r=1}^R$, while in the stationary case, the range of the degenerated distribution of $\{\widehat{\beta}_r\}_{r=1}^R$ becomes arbitrarily small for $T \rightarrow \infty$. Hence, a test statistic could be constructed as the range of the sequence of regression coefficients $\{\widehat{\beta}_r\}_{r=1}^R$, obtained from regressing the data sufficiently often, say R times, on simulated random walks that are controlled by the econometrician, i.e.

$$y_t = \beta x_t^{(r)} + \varepsilon_t^{(r)}, \quad t = 1, \dots, T, \quad r = 1, \dots, R,$$

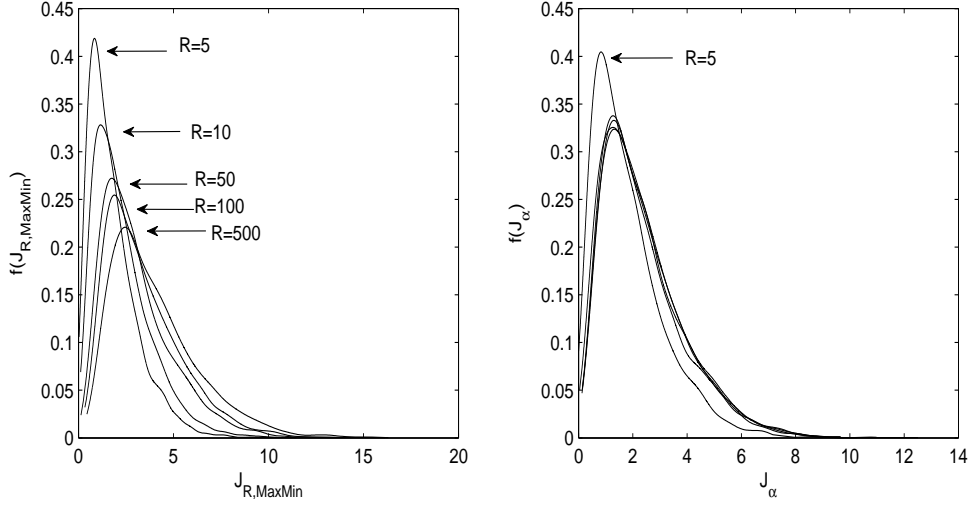
with $x_t^{(r)}$ as defined in \mathcal{A}_1 . Denote this test statistic as $J_{R,MaxMin}$,

$$J_{R,MaxMin} = |\widehat{\beta}_R^{max} - \widehat{\beta}_R^{min}|, \quad (5)$$

and $\widehat{\beta}_R^{max}$ and $\widehat{\beta}_R^{min}$ are obtained as $\widehat{\beta}_R^{max} = \max\{\widehat{\beta}_r\}_{r=1}^R$, $\widehat{\beta}_R^{min} = \min\{\widehat{\beta}_r\}_{r=1}^R$.

However, this statistic asymptotically depends on the choice of R since $J_{R,MaxMin}$ is based only on the two extremal points of the distribution of $\{\widehat{\beta}_r\}$. Increasing the sample size of $\{\widehat{\beta}_r\}$ increases the likelihood of observing larger values of $J_{R,MaxMin}$. Alternative statistics which are independent of R can be obtained by using some inter quantile range of $\widehat{\beta}_R$. Test statistics are then constructed as $J_\alpha = |\widehat{\beta}_{(1-\alpha/2)} - \widehat{\beta}_{\alpha/2}|$, where $\widehat{\beta}_\alpha$ refers to the $\alpha\%$ quantile of the empirical distribution of $\{\widehat{\beta}_r\}$. Figure 1 displays the limiting distribution of $J_{R,MaxMin}$ and $J_{0.1}$ for different values of R , where the distribution functions are based on 100000 replications. The time dimension is chosen reasonably large with $T = 1000$ in order to ensure convergence of the estimated regression coefficients to the asymptotic distribution given in (4). However, unreported results show that a very similar picture emerges even for a time dimension as small as $T = 50$.

Figure 1: Empirical PDFs conditional on $R, T = 1000$



Notes: Empirical pdfs are approximated by a Gaussian kernel density estimator with bandwidth chosen according to Silverman's rule of thumb.

The left hand side graph of Figure 1 confirms that the limiting distribution of $J_{R,MaxMin}$ depends on R . The variance of the distribution increases with R and the mode of the distribution shifts to the right as R increases. On the other hand, as shown in the right hand side graph, the limiting distribution of $J_{0.1}$ displays convergence to a unique distribution even for relatively small values of R .

Under assumption (\mathcal{A}_1) , the limiting distribution of J_α depends on three parameters: the innovation standard deviations of x_t and y_t , σ_ϵ and σ_e , and the chosen quantile α . As σ_ϵ and α are controlled by the econometrician, σ_e is the only nuisance parameter. Since σ_e can be estimated consistently from the data, the observed process can be standardized such that simulated critical values for a test based on J_α generated with $\sigma_e = 1$ remain valid.

2.2 Initial conditions and deterministic terms

The limiting distribution of $\widehat{\beta}$ as given in (4) and, hence, the limiting distribution of J_α , crucially depends on assumptions (A1) and (A2). Thus, we have to account for the more realistic cases of y_t being a random walk with drift and initial values $y_0 \neq 0$. An easy solution to account for non-zero initial values is to subtract the first observation from the data. Moreover, OLS or GLS detrending schemes can be used to eliminate non-zero mean or drift terms. The OLS detrended series is obtained as

$$\widehat{y}_t = y_t - z_t' \widehat{\psi}, \quad (6)$$

where $\widehat{\psi}$ is the OLS estimate from the regression

$$y_t = z_t' \psi + e_t, \quad e_t \sim iid(0, 1), \quad (7)$$

and z_t collects the deterministic terms, i.e. $z_t = 1$ in the intercept only case or $z_t = (1, t)'$ in the trend case, respectively. Elliott et al. (1996) demonstrate that conducting DF type unit root tests on GLS detrended data leads to significant power gains compared with the standard DF test. Similarly, Ng and Perron (2001) show that GLS detrending improves the power of M -type unit root tests. GLS detrending proceeds by constructing quasi-differences of the data, i.e.

$$y_t^{\bar{\alpha}} = y_t - \bar{\alpha} y_{t-1}, \quad y_0^{\bar{\alpha}} = y_0, \quad \text{and} \quad t = 1, \dots, T,$$

with $\bar{\alpha} = 1 + \bar{c}/T$, where \bar{c} denotes the local-to-unity parameter. The detrended vector of deterministic components $z_t^{\bar{\alpha}}$ is constructed in analogy, i.e.

$$z_t^{\bar{\alpha}} = z_t - \bar{\alpha} z_{t-1}, \quad z_0^{\bar{\alpha}} = z_0, \quad \text{and} \quad t = 1, \dots, T.$$

The local-to-unity parameter \bar{c} is chosen according to Elliott et al. (1996), such that $\bar{c} = -7$ if $z_t = 1$ and $\bar{c} = -13.5$ if $z_t = (1, t)'$, respectively. The detrending parameter estimate $\widetilde{\psi}$ is then the OLS estimator from the regression

$$y_t^{\bar{\alpha}} = z_t^{\bar{\alpha}'} \psi + e_t, \quad e_t \sim iid(0, 1), \quad (8)$$

and the GLS detrended series is

$$\widetilde{y}_t = y_t - z_t' \widetilde{\psi}, \quad t = 1, \dots, T. \quad (9)$$

2.3 Consistent estimation of the long run variance

As outlined in Section 2.1, a consistent estimate of the innovation variance σ_e^2 is required to implement the proposed unit root test. Moreover, to account for (potential) serial dependence in the innovations e_t , it appears preferable to consider serial correlation robust estimators, as, for example, kernel based estimators (Newey and West, 1987). However, Perron and Ng (1996) show that the use of kernel based estimators as in Phillips and Perron (1988) often leads to poor empirical size features of the tests. Instead, they advocate to use an autoregressive spectral density estimator at frequency zero of v_t , denoted s_{AR}^2 , in the construction of the M -statistic. Consider the augmented DF regression

$$\Delta y_t = d_t + \beta_0 y_{t-1} + \sum_{j=1}^k \beta_j \Delta y_{t-j} + e_{tk}, \quad (10)$$

where k denotes the (previously determined) lag order. The estimator s_{AR}^2 is then constructed as

$$s_{AR}^2 = \hat{\sigma}_k^2 / (1 - \hat{\beta}(1))^2, \quad (11)$$

with $\hat{\sigma}_k^2 = (T - k)^{-1} \sum_{t=k+1}^T \hat{e}_{tk}^2$, $\hat{\beta}(1) = \sum_{j=1}^k \hat{\beta}_j$, where $\hat{\beta}_i$ and $\{\hat{e}_{tk}\}$ are obtained from regression (10). Ng and Perron (2001) demonstrate that the finite sample size of the M -tests can be further improved by estimating s_{AR}^2 from the regression

$$\Delta \tilde{y}_t = \beta_0 \tilde{y}_{t-1} + \sum_{j=1}^k \beta_j \Delta \tilde{y}_{t-j} + e_{tk}, \quad (12)$$

where \tilde{y}_t denotes GLS detrended data as outlined in Section 2.2. To distinguish both estimators, we denote them as $s_{AR,OLS}^2$ and $s_{AR,GLS}^2$ if based on regression (10) or (12), respectively.

2.3.1 Lag length selection

It is long recognized that underfitting of k in (10) leads to severe size distortions of most unit root tests. Ng and Perron (2001) point out that even the comparatively liberal Akaike Information Criterion (AIC) tends to select too small a lag order if

there is a negative moving average (MA) root in the innovation process. Therefore, they propose the *modified* AIC (MAIC). It is given as

$$\text{MAIC}(k) = \ln(\hat{\sigma}_k^2) + \frac{2(\tau_T(k) + k)}{T - k_{\max}}, \quad (13)$$

where $\ln(\hat{\sigma}_k^2)$ is the natural logarithm of $\hat{\sigma}_k^2$ and $\tau_T(k) = (\hat{\sigma}_k^2)^{-1} \hat{\beta}_0^2 \sum_{t=k_{\max}+1}^T y_{t-1}^2$ and $\hat{\sigma}_k^2 = (T - k_{\max})^{-1} \sum_{t=k_{\max}+1}^T \hat{e}_{tk}^2$. The chosen lag length is then obtained as $k_{MAIC} = \arg \min_k (\text{MAIC})$ with admissible values of $k \in [0, \lfloor 12(T/100)^{1/4} \rfloor]$, where $\lfloor q \rfloor$ denotes the integer part of q . While Ng and Perron (2001) suggest to apply the MAIC to GLS detrended data, Perron and Qu (2007) recommend to base the information criterion on OLS detrended data (or an ordinary ADF regression), while still employing GLS detrending for the construction of s_{AR}^2 and the test statistics. They argue that this hybrid procedure improves the small sample properties of the considered tests.

2.4 Critical values

Critical values for all considered detrending schemes of the simulation based unit root test are documented in Table 1. Since neither OLS- nor GLS detrending ensures accordance with (A2), we also generate critical values for tests based on detrended data from which the first observation has been subtracted so that $\hat{y}_1 = 0$ and $\tilde{y}_1 = 0$, respectively.

Critical values are generated using 100000 Monte Carlo replications and setting $\sigma_e = \sigma_\epsilon = 1$, $\alpha = 10\%$ and $R = 50$. Values in parentheses denote critical values obtained for detrended data from which the first observation has been subtracted.

3 Finite sample properties

We analyze the finite sample properties of the J_α test by means of a Monte Carlo study. We simulate data according to model (1) for $t = -49, \dots, -1, 0, 1, \dots, T$ and discard the pre-sample values. Besides the benchmark scenario with $v_t \sim iid N(0, 1)$, we also consider serially dependent innovation processes formalized by

Table 1: Critical values, $J_{0.1}$

T	Intercept			Trend		
	1%	5%	10%	1%	5%	10%
OLS-detrending, $s_{AR,OLS}^2$						
25	0.28 (0.44)	0.40 (0.69)	0.49 (0.87)	0.19 (0.29)	0.25 (0.42)	0.30 (0.53)
50	0.24 (0.40)	0.36 (0.63)	0.44 (0.82)	0.15 (0.24)	0.21 (0.37)	0.25 (0.48)
100	0.22 (0.37)	0.34 (0.61)	0.42 (0.79)	0.13 (0.23)	0.19 (0.35)	0.23 (0.46)
250	0.22 (0.37)	0.32 (0.60)	0.41 (0.79)	0.13 (0.22)	0.18 (0.35)	0.23 (0.45)
500	0.21 (0.36)	0.33 (0.59)	0.41 (0.77)	0.13 (0.22)	0.18 (0.34)	0.22 (0.44)
1000	0.21 (0.36)	0.32 (0.59)	0.41 (0.77)	0.12 (0.21)	0.18 (0.34)	0.22 (0.44)
10000	0.21 (0.36)	0.32 (0.60)	0.41 (0.78)	0.12 (0.21)	0.18 (0.34)	0.22 (0.44)
GLS-detrending, $s_{AR,GLS}^2$						
25	0.34 (0.44)	0.50 (0.68)	0.60 (0.87)	0.21 (0.28)	0.28 (0.41)	0.33 (0.50)
50	0.33 (0.40)	0.50 (0.63)	0.63 (0.81)	0.19 (0.23)	0.26 (0.34)	0.31 (0.43)
100	0.34 (0.37)	0.53 (0.61)	0.67 (0.79)	0.18 (0.21)	0.26 (0.31)	0.32 (0.39)
250	0.35 (0.37)	0.56 (0.60)	0.73 (0.79)	0.18 (0.20)	0.27 (0.29)	0.33 (0.36)
500	0.35 (0.36)	0.57 (0.59)	0.74 (0.77)	0.19 (0.19)	0.28 (0.29)	0.34 (0.35)
1000	0.36 (0.36)	0.58 (0.59)	0.75 (0.77)	0.18 (0.19)	0.27 (0.28)	0.34 (0.35)
10000	0.36 (0.36)	0.60 (0.60)	0.78 (0.78)	0.19 (0.19)	0.28 (0.28)	0.34 (0.34)

Notes: Data is generated according to assumptions (\mathcal{A}_1) -(\mathcal{A}_2) with $e_t, \epsilon_t \sim iid N(0,1)$. Results based on 100000 replications and $R = 50$. The variance estimator s_{AR}^2 (eq. 11) is constructed with $k = 0$. Values in parentheses denote critical values obtained for data with $\hat{y}_1 = 0$ and $\tilde{y}_1 = 0$, respectively.

means of moving average and autoregressive innovation structures

$$MA(1) : \quad v_t = \Theta e_{t-1} + e_t, \quad e_t \sim iid N(0,1), \quad (14)$$

and

$$AR(1) : \quad v_t = \Theta v_{t-1} + e_t, \quad e_t \sim iid N(0,1), \quad (15)$$

respectively. Both cases are simulated for parameter values $\Theta \in \{-0.8, -0.5, 0.5, 0.8\}$ to capture a wide range of correlation patterns. The random walk $\{x_t\}$ needed for the construction of J_α is generated according to (\mathcal{A}_1) -(\mathcal{A}_2), as before. As already mentioned, due to the different detrending schemes we have four variants of the J_α test at hand. We denote them as $J_{\alpha,1}$ - $J_{\alpha,4}$, referring to the statistics based on OLS detrended data, OLS detrended data with $\hat{y}_1 = 0$, GLS detrended data and GLS detrended data with $\tilde{y}_1 = 0$, respectively. To assess the relative performance of the proposed simulation based unit root test, we additionally include three commonly used unit root tests in our simulation, namely the standard ADF-test, the ADF-GLS test of Elliott et al. (1996) and the \bar{M}^{GLS} test proposed by Ng and Perron

(2001). The lag length is selected for all tests according to Perron and Qu (2007) by the MAIC based on OLS-detrended data. Empirical size is evaluated under the null hypothesis of $\rho = 1$ and the nominal significance level is 5%, however, results remain qualitatively unchanged if other significance levels are chosen. We base empirical size estimates on simulated critical values for all tests since exact critical values are not tabulated in the literature for some of the tests. Exact critical values are generated from 100000 replications of model (1) under the null hypothesis with white noise error terms. Size adjusted rejection frequencies are calculated under the local alternative $H_1 : \rho = 1 - \bar{c}/T$, where $\bar{c} = 7$ and $\bar{c} = 13.5$ in the intercept and trend case, respectively. Table 2 and 3 list rejection frequencies under the null hypothesis while size adjusted local power results are documented in Tables 4 and 5.

3.1 Rejection frequencies in the unit root case

Rejection frequencies obtained for the intercept case with white noise innovations illustrate that the new proposed test achieves a high degree of size control. None ($J_{\alpha,1}$), one ($J_{\alpha,2}$, $J_{\alpha,4}$) or two ($J_{\alpha,3}$) significant deviations from the nominal significance level can be observed while in contrast, both variants of the ADF test as well as the \bar{M}^{GLS} test are characterized by much more significantly distorted empirical size estimates. In all of these cases, empirical rejection frequencies are below the nominal level and as low as 1.4% for $T = 25$ (ADF). These downward biases are presumably induced by spuriously included lags in the test regressions or in the construction of the spectral density variance estimator. Size distortions are generally more pronounced for small sample sizes which is in line with Cheung and Lai (1995) who demonstrate that the critical values of the ADF statistic exhibit a nonlinear dependence on k which vanishes for increasing T . If the simulations are based on the correct lag length (i.e. $k = 0$, unreported) empirical rejection frequencies vary insignificantly around the nominal level for all tests. A similar picture emerges if the tests are constructed to allow for trending data. While all variants of the simulation based test yield only two significant size distortions, the ordinary ADF statistic leads to rejection frequencies significantly smaller than 5% for all sample sizes T .

Finally, the ADF^{GLS} and the \bar{M}^{GLS} statistics obtain underrejections of H_0 in four out of six scenarios.

If the random walk innovations are generated by an MA structure with negative coefficient, rejection frequencies are much less precise for all statistics than under independent innovations. Especially for (large) negative MA coefficients ($\Theta = -0.8$) and $T < 100$, rejection frequencies between 23% and 39% can be observed under H_0 . This indicates that the automatic, sample size dependent choice of k_{max} might be too restrictive in small samples with strong negative MA error processes. For larger T , rejection frequencies of most tests successively tend to approach the nominal level from above. A notable exception can be observed for the \bar{M}^{GLS} statistic. While it yields excess rejection probabilities for $T \in [25, 50]$ the test tends to be rather conservative in large samples with rejection frequencies ranging between 2.1% ($T = 250$) and 3.6% ($T=1000$). This finding is in line with results reported in Ng and Perron (2001). The simulation based quantile range tests do not offer significant improvements nor drastic deteriorations compared with the existing procedures in the cases of negative MA correlations. Generally, all versions of the J_α statistic tend to overreject H_0 under negative MA errors. The magnitude of the upward size distortions depends positively on the strength of the negative MA root and inversely on T . Ceteris paribus, size distortions are more pronounced if the tests are constructed to allow for trending data.

Positive MA dynamics tend to induce rejection frequencies of less than the nominal level for both variants of the ADF statistic. On the other hand, the J_α statistics and the \bar{M}^{GLS} statistic perform relatively well with some minor undersizing observed only for small time dimensions. All statistics appear to be less sensitive to the magnitude of the (positive) MA coefficient compared to the case of a negative parameter. Similarly, differences between the intercept and trend cases are less pronounced compared with the scenario of negative MA innovations.

If the random walk innovations follow an AR(1) process (Table 3), the J_α statistics outperform the benchmark tests in many cases. In the intercept only case with negative AR coefficients, all benchmark tests underreject H_0 , mostly by a significant

margin. For instance consider a time dimension of $T = 100$. The \bar{M}^{GLS} statistic obtains rejection frequencies of 0.9% ($\Theta = -0.8$) and 3.5% ($\Theta = -0.5$). Similar rejection frequencies are documented for both variants of the ADF statistic. In contrast, all variants of the J_α statistic yield rejection frequencies close to the nominal level for $T = 100$ and at most some moderate oversizing for $T = 25$. Among the J_α statistics, both variants computed from data with initial conditions $\hat{y}_1 = 0$ and $\tilde{y}_1 = 0$ yield smallest size distortions. If the tests are implemented to account for a linear time trend, the general picture changes in so far, as the observed overrejections of the J_α statistics for $T = 25$ are somewhat more pronounced and the benchmark statistics also display overrejections of H_0 for $\Theta = -0.5$ and $T = 25$. Yet, overall, the J_α statistics yield most accurate size features for negative residual autocorrelation, formalized by an AR structure.

If the random walk innovations follow an AR(1) process with positive coefficient, both variants of the ADF statistic yield significantly too small rejection frequencies for most combinations of T and Θ . In many cases rejection frequencies are around or below 3.5% for reasonably large time dimensions such as $T \in [100, 250]$. On the other hand, the \bar{M}^{GLS} statistic tends to overreject H_0 for positive AR processes and small time dimensions. If $\Theta = 0.5$ empirical sizes of up to 8.2% (trend case) are observed for $T = 50$. If $\Theta = 0.8$, size distortions are even more pronounced. In the intercept case, empirical rejection frequencies between 7.1% ($T = 100$) and 27.8% ($T = 25$) are documented while in the trend case, these are between 11.6% ($T = 100$) and 35.2% ($T = 25$). As in the case of negative AR correlation, the J_α statistics remain most robust in terms of empirical rejection frequencies close to the nominal level. Significant size distortions are mostly restricted to small time dimensions and less pronounced than those obtained by the benchmark statistics. For moderate positive autocorrelation ($\Theta = 0.5$) and $T = 25$, the J_α statistics obtain rejection frequencies between 3.2% and 4.1% (intercept case) and between 2.0% and 2.6% (trend case), respectively. If $\Theta = 0.8$, the J_α statistics yield upward biased size estimates for $T = 25$, however, this bias is much smaller and vanishes faster than the bias observed for the \bar{M}^{GLS} statistic.

To conclude this section, the conducted Monte Carlo study confirms that the simulation based J_α unit root tests are very competitive in terms their implied rejection frequencies under the unit root null hypothesis. The only exception is a scenario of (strong) negative MA processes, however, in this case none of the considered benchmark statistics yields fully satisfactory results either. Across all considered scenarios it turns out that subtracting the first observation from the data leads to more precise size estimates of the J_α tests, while the chosen detrending scheme has at most marginal impact on rejection frequencies under H_0 .

3.2 Size adjusted local power

Size adjusted local power estimates for *iid* and MA innovations are documented in Table 4. Entries in *italic* indicate that these power estimates are based on substantial size adjustment, as the corresponding rejection frequencies under H_0 are not covered by the 95% confidence interval around the nominal level. Hence, these power estimates should be interpreted cautiously, since rejection frequencies are unreliable under H_0 .

In the intercept only case, the results document that the proposed J_α tests yield (size adjusted) power estimates of roughly 30% for the largest time dimension $T = 1000$. The most notable distinction among the J_α tests can also be observed for $T = 1000$, where $J_{\alpha,1}$ (i.e. the statistic based on OLS detrending without initial observation adjustment) yields up to 5 percentage points lower power estimates than all other variants. Among the latter, there are no marked power differentials, though.

Compared with the standard ADF test, the J_α tests display moderate power advantages varying between 3-5% for $T \in [25, 500]$ and more sizable advantages of up to 10 percentage points for $T = 1000$. The ADF^{GLS} test and the \bar{M}^{GLS} tests achieve the highest (size adjusted) local power. On average both tests offer a positive power differential compared with the J_α tests of around 5 to 6 percentage points in the large sample case ($T = 1000$). Serially correlated innovations reduce local power estimates of all tests for small time dimensions. However, it is noteworthy

that the J_α tests appear to be less affected by this adverse effect compared with the \bar{M}^{GLS} statistic. Consider, for instance, the MA case with $\Theta = 0.8$ and $T = 50$. In this scenario, the J_α statistics yield about 50% higher rejection frequencies than the \bar{M}^{GLS} statistic.

If tests are implemented to account for trending data, the ADF^{GLS} and \bar{M}^{GLS} statistics remain most powerful in large samples. However, the power differential compared with the standard ADF test is less pronounced as in the intercept case, resembling a result of Elliott et al. (1996). In contrast to the intercept case, $J_{\alpha,2}$ is now the least powerful variant among the J_α tests with up to 7 percentage points lower rejection frequencies. As before, residual serial correlation reduces local power estimates in small samples.

Table 5 lists local power estimates for data generated with AR(1) innovations. The most notable differences compared with the case of MA innovations can be observed for small sample ($T \in [25, 50]$) scenarios with positive AR coefficients where size adjusted rejection frequencies are substantially depressed. In particular, for $\Theta = 0.8$ and $T = 25$, local power estimates vary around 1% for the J_α statistics, between 2.8% and 4.3% for the ADF statistics and are as low as 0.5% for the \bar{M}^{GLS} statistic. For large time dimensions, the main conclusions drawn from the results in Table 4 persist.

To summarize local power estimates, it turns out that for large sample sizes the ADF^{GLS} and \bar{M}^{GLS} statistics are the most powerful among the considered tests. Moreover, if statistics are computed to account for a linear time trend, GLS detrending appears to be preferable to OLS detrending in the construction of the J_α statistics.

4 Empirical illustration: PPP of G6 economies

4.1 Theoretical background

The purchasing power parity (PPP) hypothesis is an important building block of many international macroeconomic models. Strong form PPP postulates that one

basket of goods has the same price across different countries when expressed in a common currency. The real exchange rate of country i at time t is defined as

$$Q_{it} = \frac{E_{it}P_{it}}{P_t^*}, \quad (16)$$

where E_{it} denotes the nominal exchange rate and P_{it} and P_t^* are the domestic and foreign (numeraire) price levels, respectively. Under strong form PPP, one would expect that $Q_{it} = 1$ in all periods t . However, there are many practical reasons why strong form PPP is too hard an assumption (see e.g. Wagner, 2008 and the references therein). Hence, empirical analyses of PPP usually test for prevalence of weak form long run PPP which can be characterized as real exchange rates, fluctuating around a stationary mean. We follow the literature in formalizing the model in log terms and using the United States as numeraire country. Then, the real exchange rate of country i is given as

$$q_{it} = e_{it} + p_{it} - p_{us,t}^*, \quad (17)$$

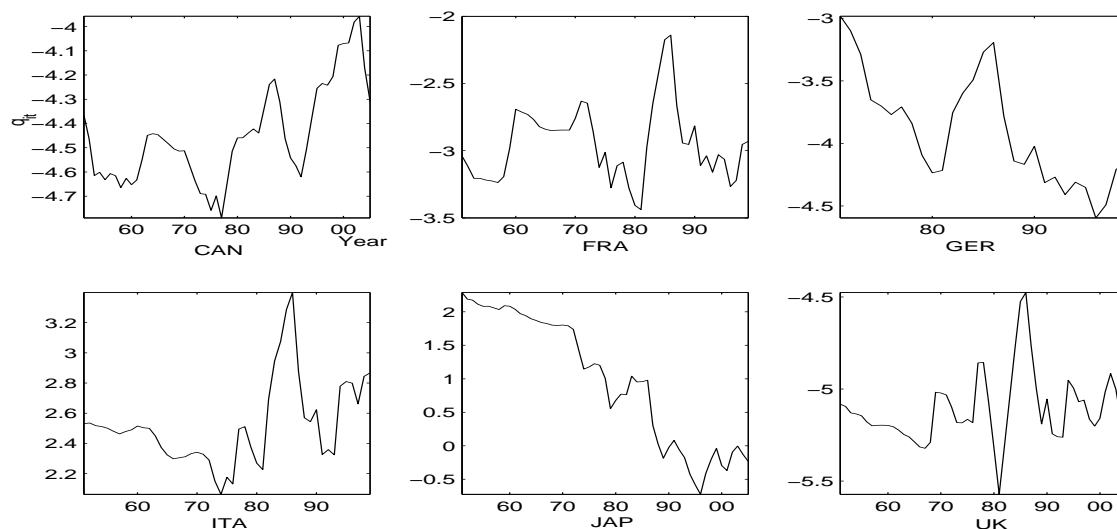
with lower case letters denoting logs of the variables defined in 2. There is a vast literature on the empirical validity of PPP (see Taylor and Taylor, 2004 for a survey). Most studies based on univariate unit root tests conclude that PPP does not hold. Since it has been argued that the inability of rejecting the unit root null hypothesis might be due to low power of univariate unit root tests, panel techniques have been employed which generally yield results more in favor of long run PPP. More recently however, it has repeatedly been pointed out that these results obtained by first generation panel unit root tests might have been induced by neglected cross sectional dependence and that evidence based on appropriate second generation panel unit root tests yield less support for long run PPP to hold (e.g. Wagner, 2008).

We acknowledge that a thorough investigation of PPP should make use of (second generation) panel unit root tests. However, we use this small scale empirical illustration to highlight differences of the analyzed test procedures in an empirical context.

4.2 Data

We obtain annual data on price levels and exchange rates from the Penn World Tables (PWT), Mark 6.2. Data spans from 1950-2004 and the base year is 2000. However, for those economies adopting the Euro in 1999, we restrict the sample to the observations prior to the introduction of the Euro. In the case of Germany, price level data availability only starts in 1970. Figure 2 displays the log real exchange rates of the G6 economies.

Figure 2: Real exchange rates of G6 economies



Notes: Figures display the logs of bilateral real exchange rates versus the US-\$.

4.3 Results

Construction of the test statistics proceeds in analogy to the Monte Carlo study and the same simulated critical values are used for the test decisions. Table 6 lists the empirical results which document some notable differences between the various tests. Based on the standard ADF test, according to our simulation study the least powerful among the considered tests, we cannot reject the unit root null hypothesis for any of the G6 economies at the 5% significance level. If we base inference on

either of the $J_{\alpha,1}$, ADF^{GLS} or \bar{M}^{GLS} statistics, we obtain evidence in favor of long run PPP at least for the United Kingdom. Based on $J_{\alpha,3}$, rejection of H_0 is also found for France. Finally, according to the initial observation adjusted statistics $J_{\alpha,2}$ and $J_{\alpha,4}$, long run PPP additionally holds for Italy, and hence, for half of the economies under consideration.

The mixed evidence on PPP among the G6 economies obtained by the alternative unit root tests points out that the proposed simulation based testing principle can add additional insights in empirical applications of unit root tests.

5 Conclusions

In this paper, we introduce a new approach to unit root testing. The underlying motivation for the new test is that the regression coefficient from a spurious regression (i.e. a regression of two uncorrelated random walks) has a nondegenerated limiting distribution. In contrast, if the dependent variable is stationary, the regression coefficient converges to zero in probability. To construct a feasible test statistic, we propose to run a sequence of regressions of the data on simulated random walks with Gaussian innovations. Test statistics can then be obtained as some inter quantile ranges of the resulting empirical distribution. If appropriately scaled, these statistics have an invariant limiting distribution under the null hypothesis, while they converge to zero at the rate T^{-1} under the alternative hypothesis. We implement variants of these statistics based on the range between the 5 and 95 percentile of the simulated distribution. To account for higher order serial correlation, we apply the autoregressive spectral density variance estimator proposed by Perron and Ng (1998) in conjunction with the modified Akaike information criterion (MAIC) (Ng and Perron, 2001) to obtain a nuisance parameter free test statistic.

By means of a Monte Carlo study we assess the finite sample properties of the new test. It turns out that it has favorable size properties for most of the considered data generating processes, especially for relatively small time dimensions. In contrast to standard ADF tests, removal of deterministic terms by means of GLS

detrending does not substantially improve finite sample power features of the test in the intercept case but appears to be preferable in the trend case. In terms of size adjusted local power, it turns out that the proposed test is more powerful than the standard ADF test in the intercept only case, while it is slightly less powerful than the ADF-GLS test of Elliott et al. (1996) and the \bar{M}^{GLS} test of Ng and Perron (2001) in large samples. However there are some scenarios of small samples with residual autocorrelation in which the proposed test yields highest power among those tests which achieve reasonable rejection frequencies under H_0 . In an empirical illustration on PPP among G6 economies, we show that the proposed test tends to support long run PPP for more economies than the benchmark tests.

A number of interesting issues are open for future research. First and foremost, the analytical derivation of the of the proposed test's limiting distribution deserves further consideration. Furthermore, it is not clear if the analyzed statistics are the most efficient implementation of the proposed testing principle. One could, for instance, consider alternative regression designs, use different inter quantile ranges to construct test statistics or apply other variance estimators or pre-whitening schemes to cope with residual serial correlation. Moreover, it should be straightforward to apply the proposed testing idea to the fields of stationarity and cointegration testing as well as to expand it to the panel case. Especially the latter appears promising, considering the relatively good performance of the proposed test in small samples. Another important issue for further research is to analyze in how far the new approach copes with violations of standard assumptions, such as outliers, breaks in the intercept or trend function as well as nonstationary volatility.

References

- Aparico, F., Escribano, A., Sipolis, A. E., 2006. Range unit-root (rur) tests: robust against nonlinearities, error distributions, structural breaks and outliers. *Journal of Time Series Analysis* 27 (4), 545–576.
- Breitung, J., Gouriéroux, C., 1997. Rank tests for unit roots. *Journal of Econometrics* 81 (1), 7–27.
- Cheung, Y.-W., Lai, K. S., 1995. Lag order and critical values of the Augmented Dickey-Fuller test. *Journal of Business and Economic Statistics* 13 (3), 277–280.
- Dickey, D. A., Fuller, W. A., 1979. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association* 74, 427–431.
- Elliott, G., Rothenberg, T. J., Stock, J. H., 1996. Efficient tests for an autoregressive unit root. *Econometrica* 64 (4), 813–836.
- Granger, C. W. J., Newbold, P., 1974. Spurious regressions in econometrics. *Journal of Econometrics* 2 (2), 111–120.
- Heston, A., Summers, R., Aten, B., 2006. Penn world tables version 6.2, Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania.
- Newey, W. K., West, K. D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55 (3), 703–708.
- Ng, S., Perron, P., 2001. Lag length selection and the construction of unit root tests with good size and power. *Econometrica* 69 (6), 1519–1554.
- Perron, P., Ng, S., 1996. Useful modifications to some unit root tests with dependent errors and their local asymptotic properties. *Review of Economic Studies* 63 (3), 435–463.

- Perron, P., Ng, S., 1998. An autoregressive spectral density estimator at frequency zero for nonstationarity tests. *Econometric Theory* 14 (5), 560–603.
- Perron, P., Qu, Z., 2007. A simple modification to improve the finite sample properties of Ng and Perron’s unit root test. *Economics Letters* 94 (1), 12–19.
- Phillips, P. C. B., 1986. Understanding spurious regressions in econometrics. *Journal of Econometrics* 33 (3), 331–340.
- Phillips, P. C. B., Perron, P., 1988. Testing for a unit root in time series regression. *Biometrika* 75 (2), 335–346.
- Said, S. E., Dickey, D. A., 1984. Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika* 71 (3), 599–607.
- Stock, J. H., 1999. A class of tests for integration and cointegration. In: Engle, R. F., White, H. (Eds.), *Cointegration, Causality and Forecasting. A Festschrift in Honour of Clive W.J. Granger*. Oxford University Press, Oxford.
- Taylor, A. M., Taylor, M. P., 2004. The purchasing power parity debate. *Journal of Economic Perspectives* 18 (4), 135–158.
- Wagner, M., 2008. On ppp, unit roots and panels. *Empirical Economics* 35 (2), 229–249.

Table 2: Empirical size

T	Intercept							Trend						
	$J_{\alpha,1}$	$J_{\alpha,2}$	$J_{\alpha,3}$	$J_{\alpha,4}$	ADF	ADF^G	\bar{M}^G	$J_{\alpha,1}$	$J_{\alpha,2}$	$J_{\alpha,3}$	$J_{\alpha,4}$	ADF	ADF^G	\bar{M}^G
$v_t \sim iid N(0,1)$														
25	4.8	4.7	4.1	4.1	1.4	2.2	3.5	4.4	4.6	4.6	4.5	3.2	3.4	3.9
50	4.6	4.1	3.8	3.8	2.4	2.2	2.7	4.1	3.8	3.9	4.0	2.5	2.5	3.1
100	4.7	4.4	4.6	4.5	3.1	3.5	3.8	4.5	4.4	4.6	4.8	3.1	3.5	3.8
250	5.2	4.7	4.7	5.1	4.4	4.3	4.2	4.3	4.3	4.0	4.3	3.4	3.8	4.0
500	5.6	4.9	5.0	5.2	4.8	4.6	4.8	5.0	4.7	5.2	5.1	4.3	4.5	4.8
1000	5.0	4.4	4.4	4.6	4.7	4.2	4.3	5.0	4.9	4.8	5.2	4.0	5.1	5.1
MA case: $\Theta = -0.8$														
25	39.2	23.4	36.3	23.0	28.8	32.9	32.3	44.7	22.8	44.4	23.8	48.6	51.3	48.2
50	30.0	20.7	23.9	19.4	16.9	18.9	16.5	47.8	24.8	43.9	28.6	34.4	36.8	35.5
100	20.9	13.7	12.7	11.8	8.6	9.4	5.7	34.0	17.2	24.2	20.1	16.3	17.2	13.4
250	12.7	9.4	8.4	8.4	4.8	6.9	2.1	17.7	9.7	10.7	10.3	6.4	6.9	2.1
500	10.6	7.8	7.5	7.5	5.8	6.9	2.7	12.5	7.8	8.8	8.8	6.0	6.4	1.2
1000	7.4	5.9	5.9	6.0	5.7	6.3	3.6	8.9	7.5	7.3	7.8	5.9	7.0	1.5
MA case: $\Theta = -0.5$														
25	15.5	10.9	14.1	10.3	10.7	13.4	14.5	19.8	12.6	19.4	12.9	21.3	24.4	25.1
50	10.9	8.2	8.3	7.6	6.0	6.5	6.8	15.0	9.8	14.2	10.4	9.5	10.6	12.1
100	9.1	6.7	6.6	6.3	3.5	4.8	4.5	11.0	8.2	10.1	8.8	5.0	6.1	6.4
250	7.0	6.0	5.9	5.9	4.2	5.1	4.9	7.4	5.7	6.0	6.1	4.1	4.9	4.5
500	7.0	6.0	5.8	5.5	4.7	5.3	5.3	7.0	5.8	6.5	6.2	4.9	5.5	5.5
1000	5.6	5.1	4.8	5.0	4.8	4.9	4.8	6.1	5.8	5.8	6.2	4.6	5.7	5.7
MA case: $\Theta = 0.5$														
25	3.2	3.6	3.1	3.3	1.0	0.5	5.0	2.1	2.4	1.7	2.0	0.6	0.2	3.3
50	4.2	4.3	3.6	4.1	1.5	1.1	4.1	2.4	3.0	2.5	3.0	0.4	0.2	3.5
100	5.3	4.9	4.9	4.9	2.4	2.7	4.6	4.6	4.7	4.7	4.7	1.4	1.7	5.1
250	5.1	5.0	4.9	5.0	3.3	3.5	4.6	4.4	4.7	4.2	4.3	2.3	2.3	4.0
500	6.0	5.3	5.4	5.2	3.9	4.5	4.9	5.4	5.1	5.6	5.2	3.5	3.5	4.8
1000	5.0	4.8	4.4	4.8	4.4	4.2	4.5	5.4	5.1	5.1	5.5	3.9	4.5	5.3
MA case: $\Theta = 0.8$														
25	4.1	4.6	4.2	4.4	0.9	0.4	8.4	2.0	2.4	2.0	2.2	0.5	0.1	5.7
50	5.0	4.8	4.4	4.7	1.1	1.0	5.6	3.5	4.3	3.9	3.8	0.4	0.3	7.4
100	6.0	5.5	5.1	5.3	1.5	1.9	5.2	5.0	5.5	4.6	4.8	1.0	0.9	6.5
250	5.9	5.8	5.4	5.5	2.2	3.2	5.5	5.8	5.3	4.9	5.0	1.5	1.7	5.5
500	6.3	5.9	5.5	5.3	3.6	4.2	5.4	6.3	5.7	6.1	6.0	2.6	3.3	6.1
1000	5.1	5.1	4.5	4.9	3.7	4.1	4.8	6.0	5.8	5.7	5.8	3.6	4.2	6.5

Notes: Statistics $J_{\alpha,1}$ - $J_{\alpha,4}$ refer to the new simulation based tests as defined above, while ADF^G denotes the ADF-GLS test of Elliott et al. (1996) and \bar{M}^G the M -type test of Ng and Perron (2001) based on GLS detrending. To facilitate interpretation of the Tables, bold entries indicate rejection frequencies which are not covered by the 95% confidence interval $[4.4\%, 5.6\%]$ around the nominal 5% level constructed as $\alpha \pm 1.96\sqrt{\alpha(1-\alpha)/5000}$, $\alpha = 0.05$. Rejection frequencies under the null hypothesis are calculated for data generated according to model (1) with $d_t = 0$ and $\rho = 1$. MA and AR error processes are generated by (14) and (15), respectively. 5000 replications are generated throughout and test statistics $J_{\alpha,1}$ - $J_{\alpha,4}$ are based on $R = 50$. For all statistics, the lag length is chosen according to the $MAIC$ applied to OLS demeaned or detrended data.

Table 3: Empirical size, continued

T	Intercept							Trend						
	$J_{\alpha,1}$	$J_{\alpha,2}$	$J_{\alpha,3}$	$J_{\alpha,4}$	ADF	ADF^G	\bar{M}^G	$J_{\alpha,1}$	$J_{\alpha,2}$	$J_{\alpha,3}$	$J_{\alpha,4}$	ADF	ADF^G	\bar{M}^G
AR case: $\Theta = -0.8$														
25	7.4	5.5	5.8	5.4	1.9	2.7	1.7	8.8	6.2	8.3	6.1	3.0	3.2	1.7
50	5.1	4.2	4.1	4.3	2.1	2.2	0.5	6.2	4.8	5.2	4.5	1.9	2.0	0.3
100	5.3	4.6	4.7	4.5	2.2	3.4	0.9	5.7	4.9	5.4	5.1	2.5	2.6	0.3
250	5.2	4.6	4.4	4.8	3.6	3.8	2.3	4.6	4.5	4.2	4.6	2.9	3.6	1.2
500	5.7	4.7	4.9	5.1	4.4	4.3	3.8	5.2	4.8	5.2	5.0	4.1	4.3	2.6
1000	5.0	4.5	4.5	4.7	4.1	4.3	3.9	5.2	4.8	5.0	5.2	3.7	5.0	3.7
AR case: $\Theta = -0.5$														
25	8.4	6.3	7.2	5.6	4.1	4.8	5.6	10.2	7.2	10.1	7.3	7.9	8.7	8.5
50	5.3	4.6	4.6	4.6	2.5	2.1	2.1	6.6	4.8	6.1	5.1	2.4	2.8	2.9
100	5.3	4.5	4.8	4.5	2.3	3.5	3.1	6.0	5.4	5.9	5.4	2.7	3.0	2.5
250	5.4	4.8	4.8	5.0	3.8	3.9	3.8	4.7	4.4	4.4	4.5	3.0	3.7	3.2
500	5.7	5.3	5.1	5.1	4.6	4.4	4.4	5.3	4.7	5.3	5.4	4.2	4.5	4.3
1000	4.9	4.6	4.4	4.5	4.3	4.2	4.2	5.2	5.0	5.1	5.3	3.8	5.0	4.8
AR case: $\Theta = 0.5$														
25	3.2	4.1	3.5	3.8	1.2	0.4	8.8	2.1	2.6	2.0	2.2	0.5	0.0	8.9
50	3.7	4.5	4.0	4.3	2.3	1.8	5.8	2.4	3.5	2.9	2.7	0.7	0.6	8.2
100	4.8	4.8	5.0	4.9	2.9	3.4	5.1	4.4	5.0	4.5	4.7	2.4	3.2	7.1
250	5.2	5.0	5.1	5.0	3.8	4.1	4.8	4.1	4.7	4.1	4.3	3.4	3.6	5.1
500	5.6	5.0	5.1	5.1	4.3	4.6	5.0	5.0	5.0	5.2	4.9	4.0	4.2	5.0
1000	4.8	4.5	4.1	4.7	4.2	4.3	4.3	5.2	5.1	5.0	5.1	4.0	5.0	5.3
AR case: $\Theta = 0.8$														
25	9.6	10.4	11.2	10.6	2.6	2.0	27.8	10.9	10.3	10.4	9.7	1.0	0.4	35.2
50	5.0	6.1	5.6	6.0	3.0	3.0	13.3	5.8	7.2	6.4	6.0	2.6	2.5	26.7
100	4.5	4.7	4.6	4.8	3.3	3.6	7.1	4.1	4.6	3.9	4.2	2.9	3.2	11.6
250	4.1	4.5	4.7	5.1	3.8	4.4	5.6	3.4	4.0	3.4	3.5	3.2	3.4	5.3
500	5.1	5.0	5.2	5.2	4.3	4.7	5.5	4.5	4.8	4.6	4.6	4.3	4.2	5.5
1000	4.5	4.5	4.3	4.6	4.1	4.2	4.5	4.8	5.2	4.8	4.9	4.1	5.0	5.6

Notes: See Table 2

Table 4: Size adjusted local power

T	Intercept							Trend						
	$J_{\alpha,1}$	$J_{\alpha,2}$	$J_{\alpha,3}$	$J_{\alpha,4}$	ADF	ADF^G	\bar{M}^G	$J_{\alpha,1}$	$J_{\alpha,2}$	$J_{\alpha,3}$	$J_{\alpha,4}$	ADF	ADF^G	\bar{M}^G
$v_t \sim iid N(0,1)$														
25	19.0	19.5	21.8	19.1	17.3	26.4	22.5	29.5	19.1	28.7	20.4	33.1	37.3	37.7
50	23.1	23.4	25.7	22.4	17.4	30.6	27.8	28.1	22.1	32.4	24.2	30.4	40.6	39.2
100	23.5	21.0	21.4	21.2	18.6	28.3	27.0	27.9	21.1	28.2	24.1	28.6	34.2	35.0
250	21.3	22.2	23.6	20.4	17.4	27.5	27.7	29.9	21.7	29.5	28.1	28.6	37.4	36.9
500	22.9	23.2	22.7	22.2	17.9	29.2	29.0	27.5	22.5	28.2	29.1	27.6	36.8	35.6
1000	24.9	31.4	30.4	29.5	19.4	36.2	36.9	26.9	23.7	31.0	28.9	31.9	37.0	37.7
MA case: $\Theta = -0.8$														
25	15.2	11.5	15.3	12.1	19.7	18.9	10.2	17.2	11.4	17.8	11.2	27.8	28.3	4.1
50	20.6	16.2	18.1	16.5	20.4	20.4	20.2	26.5	13.7	22.5	16.8	31.0	31.7	17.0
100	21.8	17.0	18.4	17.1	18.9	18.0	18.3	30.2	17.2	24.5	20.0	27.0	26.3	26.3
250	22.2	17.0	17.0	15.9	17.2	16.7	15.2	30.0	17.3	21.1	19.7	23.8	21.9	19.1
500	22.3	20.0	19.7	19.8	17.7	22.3	19.3	28.0	17.9	22.9	23.3	24.5	25.7	20.4
1000	25.9	27.2	26.6	25.9	20.2	30.1	28.7	25.9	17.8	23.7	23.0	26.6	28.7	24.7
MA case: $\Theta = -0.5$														
25	17.9	15.8	19.0	15.7	19.7	20.9	19.8	17.9	15.8	19.0	15.7	19.7	20.9	19.8
50	20.6	19.5	22.8	19.8	15.6	21.2	20.2	20.6	19.5	22.8	19.8	15.6	21.2	20.2
100	22.1	20.2	22.2	19.9	18.2	22.7	22.6	22.1	20.2	22.2	19.9	18.2	22.7	22.6
250	20.5	20.2	20.8	19.6	17.3	24.3	23.4	20.5	20.2	20.8	19.6	17.3	24.3	23.4
500	22.5	21.5	21.9	22.4	18.2	28.1	27.3	22.5	21.5	21.9	22.4	18.2	28.1	27.3
1000	26.1	29.1	30.2	28.8	19.6	35.4	34.6	26.1	29.1	30.2	28.8	19.6	35.4	34.6
MA case: $\Theta = 0.5$														
25	14.2	13.9	15.2	13.3	10.5	22.6	4.7	24.9	18.9	26.6	18.6	25.0	37.5	16.6
50	15.8	16.6	17.8	16.2	9.3	19.8	13.7	18.3	15.5	21.3	16.7	16.0	25.6	6.7
100	20.1	18.0	18.6	18.3	15.3	24.0	21.5	22.5	17.2	21.9	20.1	17.5	23.6	20.1
250	20.5	20.7	21.1	20.6	16.4	24.9	25.1	27.2	19.9	26.4	25.8	24.1	32.4	32.3
500	22.5	21.8	21.4	22.1	16.4	27.4	28.4	25.6	20.8	26.3	27.1	24.2	34.7	34.1
1000	25.1	30.0	30.9	28.1	18.3	36.5	36.7	26.1	23.3	29.1	28.2	27.8	36.7	36.5
MA case: $\Theta = 0.8$														
25	9.5	9.6	10.3	9.3	8.5	16.6	1.1	20.4	15.6	22.3	16.4	19.7	33.2	0.7
50	13.8	15.0	15.5	14.8	9.8	20.0	10.4	15.0	12.4	14.7	13.7	11.3	19.3	2.8
100	18.3	16.1	16.9	15.7	11.8	22.1	16.4	20.4	15.6	20.2	18.3	17.2	24.5	14.0
250	19.7	19.6	19.3	19.3	14.6	22.5	20.0	25.6	18.8	24.4	23.3	20.7	28.8	25.5
500	22.0	21.5	20.9	21.4	15.3	26.3	25.7	24.6	19.4	24.7	25.5	22.3	31.1	30.0
1000	25.0	28.7	30.0	28.2	18.7	35.4	34.1	25.5	22.6	28.7	27.5	25.4	33.6	32.7

Notes: To facilitate interpretation of the Tables, italic entries indicate size adjusted power estimates, corresponding to rejection frequencies under H_0 which are not covered by the 95% confidence interval [4.4%, 5.6%]. For further notes see Table 2.

Table 5: Size adjusted local power, continued

T	Intercept							Trend						
	$J_{\alpha,1}$	$J_{\alpha,2}$	$J_{\alpha,3}$	$J_{\alpha,4}$	ADF	ADF^G	\bar{M}^G	$J_{\alpha,1}$	$J_{\alpha,2}$	$J_{\alpha,3}$	$J_{\alpha,4}$	ADF	ADF^G	\bar{M}^G
AR case: $\Theta = -0.8$														
25	17.2	14.6	18.8	14.1	17.0	19.5	11.2	23.1	12.0	21.6	12.1	27.0	26.9	10.8
50	23.0	19.0	21.8	18.6	16.4	24.3	19.6	28.8	15.8	24.9	18.3	27.2	30.1	18.1
100	23.1	19.7	19.9	20.0	17.3	23.1	21.9	28.3	17.8	23.4	20.4	26.8	29.6	25.3
250	20.9	21.4	22.2	20.0	15.9	25.9	25.6	29.7	20.4	27.6	25.8	27.1	34.3	32.3
500	23.2	23.0	22.9	22.1	17.6	29.2	29.5	27.8	21.0	27.1	26.9	26.5	34.3	32.8
1000	25.6	29.3	29.7	29.1	19.8	35.7	36.1	27.2	23.2	29.8	27.4	30.6	36.2	36.6
AR case: $\Theta = -0.5$														
25	17.5	16.9	19.3	17.3	14.8	18.5	15.2	27.0	15.8	26.3	15.8	27.3	27.2	14.6
50	23.3	22.2	24.1	21.7	16.2	27.5	24.3	28.2	21.0	28.8	22.8	25.9	31.1	23.0
100	23.6	21.6	21.6	20.7	18.4	26.0	24.8	27.4	19.3	26.2	22.3	26.9	31.9	30.4
250	21.1	21.0	21.8	20.3	15.7	27.0	26.2	29.1	21.3	28.0	26.1	26.6	34.8	34.0
500	22.8	21.9	22.3	22.4	17.3	28.5	29.2	27.6	22.6	28.6	27.3	27.5	34.1	33.9
1000	25.7	29.8	30.2	29.3	19.1	36.2	37.0	27.0	23.6	30.4	28.1	30.9	36.6	37.3
AR case: $\Theta = 0.5$														
25	8.6	7.7	8.5	7.2	5.6	11.6	1.2	16.9	13.7	17.6	14.2	17.5	31.4	0.1
50	14.0	14.4	15.4	14.6	8.1	16.6	13.2	9.7	10.0	10.2	10.6	5.3	7.5	2.2
100	20.3	17.7	18.5	17.8	14.9	24.2	23.5	20.6	15.5	20.7	18.7	16.8	22.9	20.1
250	19.7	20.6	20.7	20.1	16.4	24.9	25.4	26.9	19.6	27.2	25.1	24.9	33.4	32.6
500	22.2	22.6	21.9	22.0	17.7	28.8	28.2	26.1	20.9	26.8	28.1	27.4	35.7	35.4
1000	25.5	30.6	31.0	28.6	19.4	37.3	37.0	26.6	23.0	29.3	28.9	29.0	36.1	36.4
AR case: $\Theta = 0.8$														
25	1.4	1.2	0.8	1.2	2.8	4.3	0.5	0.2	0.2	0.2	0.3	3.3	3.1	0.1
50	9.7	9.5	9.9	10.0	9.8	15.2	6.4	5.5	4.8	5.5	5.9	5.2	7.4	1.0
100	15.3	15.2	16.3	14.9	12.3	19.9	16.0	14.8	12.6	16.2	15.1	13.3	16.2	11.0
250	18.3	19.6	20.0	17.7	14.8	22.4	21.5	23.3	17.7	22.4	23.0	20.0	29.1	25.8
500	21.0	21.1	20.3	21.2	16.5	27.9	25.9	23.1	19.2	25.3	25.4	23.2	31.9	31.4
1000	25.7	28.7	30.5	28.3	19.2	35.9	36.2	25.8	21.8	28.2	27.6	26.9	34.2	34.9

Notes: See Table 4.

Table 6: Empirical results

Country	T	k	$J_{\alpha,1}$	$J_{\alpha,2}$	$J_{\alpha,3}$	$J_{\alpha,4}$	ADF	ADF^G	\bar{M}^G
CAN	55	10	2.41	1.96	1.37	2.27	0.31	-0.60	-0.95
FRA	49	0	0.41	0.50	0.29	0.52	-2.28	-2.27	-8.90
GER	29	0	0.85	3.92	1.42	3.85	-2.18	-1.34	-2.19
ITA	49	0	0.41	0.46	0.67	0.45	-1.91	-1.95	-7.51
JAP	55	1	1.82	3.81	2.33	3.57	-0.83	-0.05	-0.03
UK	55	0	0.22	0.31	0.28	0.45	-2.72	-2.75	-13.09

Notes: Bold entries denote significance at the 5% level. Values below T and k refer to the available time series dimension and the chosen lag length, respectively. For further notes see Table 2.