

Abdul Aziz Ali<sup>1</sup> / Kristofer Månsson<sup>2</sup> / Ghazi Shukur<sup>1</sup>

# A wavelet-based variance ratio unit root test for a system of equations

<sup>1</sup> Linnaeus University, Växjö, Sweden<sup>2</sup> Jönköping International Business School, Jönköping University, Jönköping, Sweden, E-mail: kristofer.mansson@ju.se

## Abstract:

In this paper, we suggest a unit root test for a system of equations using a spectral variance decomposition method based on the Maximal Overlap Discrete Wavelet Transform. We obtain the limiting distribution of the test statistic and study its small sample properties using Monte Carlo simulations. We find that, for multiple time series of small lengths, the wavelet-based method is robust to size distortions in the presence of cross-sectional dependence. The wavelet-based test is also more powerful than the Cross-sectionally Augmented Im et al. unit root test (Pesaran, M. H. 2007. "A Simple Panel Unit Root Test in the Presence of Cross-section Dependence." *Journal of Applied Econometrics* 22 (2): 265–312.) for time series with between 20 and 100 observations, using systems of 5 and 10 equations. We demonstrate the usefulness of the test through an application on evaluating the Purchasing Power Parity theory for the Group of 7 countries and find support for the theory, whereas the test by Pesaran (Pesaran, M. H. 2007. "A Simple Panel Unit Root Test in the Presence of Cross-section Dependence." *Journal of Applied Econometrics* 22 (2): 265–312.) finds no such support.

**Keywords:** system of equations, unit roots, wavelets

**DOI:** 10.1515/snde-2018-0005

## 1 Introduction

Testing for unit roots in systems of equations has been an active area of research for at least the last three decades. The principal aim of this research has been to increase the power of unit root tests by utilizing the cross-sectional dimension of multiple time series. In this way, power gains can be made by increasing the overall number of observations while using relatively short time series. This approach is often preferable to the use of long univariate time series, which are likely undergo structural changes.

One of the earliest unit root tests in systems of equations was the test by Levin, Lin, and Chu (2002). This test assumes a common autoregressive parameter for all time series in the equation system and consequently pools the data. The assumption of a common parameter, however, imposes a restriction that limits the use of the test for heterogeneous time series. Im, Pesaran, and Shin (2003) presented the IPS test, which relaxed this assumption and modeled the individual time series using separate linear trends. Their suggested test statistic was the average of the  $t$ -statistics from the individual equations. However, implicit in this method is the assumption that all the time series are of similar length, i.e. that the data is balanced. The test has also been revealed to be sensitive to cross-sectional dependency (see Li and Shukur, 2013, for example).

Another panel unit root test that allows for heterogeneous panels was presented by Maddala and Wu (1999) and Choi (2001). This test combines evidence from several independent tests using their  $p$ -values and has its basis in the method found in Fisher (1932). If  $P_i$  is the  $p$ -value for the  $i$ th unit root test, then  $-2 \sum \log P_i$  has an exact  $\chi^2$  distribution, with degrees of freedom equal to twice the number of the individual tests (and therefore, their  $p$ -values). Maddala's unit root test does not require balanced data, can be conducted on  $p$ -values obtained from any unit root test, and is less sensitive to correlation across time series compared to the IPS unit root test (see Maddala and Wu, 1999).

The tests described above belong to a group of tests referred as the first generation unit root tests in the panel data literature. These tests depend on the assumption that there is no correlation between the individual time series in the equation system – an assumption that rarely holds in practice. Consequently, tests that account for correlation between time series in equation systems have been proposed. These are often referred to as the second generation unit root tests. The cross-sectionally augmented Im, Shin and Pesaran test, hereafter referred to as CIPS, (Pesaran 2007) is perhaps the most popular of the second generation unit root tests. Results from

Kristofer Månsson is the corresponding author.

© 2020, Kristofer Månsson et al., published by Walter de Gruyter GmbH, Berlin/Boston.

This work is licensed under the Creative Commons Attribution 4.0 Public License.

using CIPS on time series of short lengths will be investigated and compared to those from the proposed unit root test.

We suggest a wavelet variance ratio unit root test for a system of equations. Monte Carlo simulations show that the proposed test is powerful and robust to correlation between time series. We derive the limiting distribution of the wavelet variance ratio test statistic in the cases where the alternatives have no deterministic components, as well as when testing against trend stationarity (stationarity around a non-zero mean and time trend). The limiting distribution is presented under the condition that the lengths of the time series increase, but with a fixed number of time series. Results from the Monte Carlo simulations show that the wavelet-based test retains its nominal size for all of the data generating processes (DGPs) considered, and has better power compared to CIPS.

Finally, we demonstrate the usefulness of the test using an empirical application on evaluating the Purchasing Power Parity theory for the Group of 7 countries. Evidence from this evaluation points to different countries following different specifications, with some having stationary exchange rate series.

## 2 Methodology

### 2.1 Variance ratio unit root tests and the wavelet filters

There has been considerable research into testing the random walk and martingale difference hypotheses, mainly in the context of asset prices. Of particular interest is the model where the error term is an uncorrelated process, which is common in financial time series. Consider the Random Walk Model given by,

$$\begin{aligned}y_t &= z_t + \varepsilon_t \\z_t &= \mu + z_{t-1} + \eta_t\end{aligned}$$

where,

$$\varepsilon_t = B(L)\delta_t \text{ and } \eta_t = A(L)\epsilon_t \text{ are stationary processes.}$$

Variance ratio unit root tests use the fact that, for a unit root time series, the variance of the  $k$ th difference of the series is an increasing linear function of the difference,  $k$ . The test statistics of these tests are, therefore, based on estimators of the ratio of variances at different lags to that at lag 1.

Let  $\sigma_q^2 = \text{Var}(y_t - y_{t-q})/q$  and  $\Delta y_t = (y_t - y_{t-1})$ . The relation between  $\sigma_q^2$  and the autocorrelation coefficients of  $\Delta y_t$  are given as (see Cochrane (1988)),

$$\frac{\sigma_q^2}{\sigma_{\Delta y}^2} = 1 + 2 \sum_{k=1}^{q-1} \left(1 - \frac{k}{q}\right) \rho_k,$$

where  $\rho_i$  is the lag  $i$  autocorrelation coefficient of the first differences of the  $\{y_t\}_{t=0}^T$  series. This type of variance ratio unit root test is essentially a specification test using the null hypothesis,

$$H_0 : \rho_k = 0; k = 1, \dots, q.$$

The variance ratio (see Cochrane 1988) is given as follows,

$$\text{VR} = \frac{f_{\Delta y}(0)}{\hat{\sigma}^2(1)}$$

where  $\hat{\sigma}^2(1)$  is an unbiased estimate of the variance at lag 1 and  $f_{\Delta y}(0)$  is the spectral density estimator of  $\Delta y_t$  at the zero frequency. An estimate of  $f_{\Delta y}(0)$  can be based on the sample autocorrelations of  $\Delta y_t$ . When the time series has a unit root, the expected value of the variance ratio should be close to 1 for all lags  $k$ . The variance ratio will be less than 1 when the first differences are correlated, indicating the rejection of the null hypothesis of a serially uncorrelated random walk.

Other variance ratio tests are those suggested by Tanaka (1990) and Kwiatkowski et al. (1992). The test statistic for the variance ratio test given in Kwiatkowski et al. (1992) is,

$$Q_T = \frac{\sum_{t=1}^T Y_t^2 / T^2}{\sum_{t=1}^T y_t^2 / T}$$

where  $Y_t = \sum_{i=1}^t y_i$  is the partial sum of the  $\{y_t\}_{t=0}^T$  process.  $\sum_{t=1}^T y_t^2 / T$  estimates the long-run variance which, in the case of serial dependence, can be estimated using semi-parametric kernel based methods e.g. the Newey-West estimator (Newey and West, 1987). When testing against trend stationary alternatives, the test statistic is for the detrended series,

$$\hat{Q}_T = \frac{\sum_{t=1}^T \tilde{Y}_t^2 / T^2}{\sum_{t=1}^T \tilde{y}_t^2 / T}$$

where the detrended series is given as  $\tilde{y}_t = (y_t - \hat{\mu})$ , and  $\hat{\mu}$  is an estimate of the deterministic component – for example, the sample average in the case of the null being of stationarity about a non-zero mean. The test statistic of Kwiakowski et al. given above tests for stationarity i.e. it has its null hypothesis as stationary. Breitung (2002) reversed the roles of the null and alternative hypotheses and proposed using  $\hat{Q}_T$  as a unit root test where the null hypothesis is non-stationarity. Used in this way, its limiting distribution under the null hypothesis (see Breitung (2002), Proposition 3) does not depend on the long-run variance as the long-run variance cancels out in the variance ratio. This removes the need for kernel function selection and tuning parameter optimization necessary for estimating the long-run variances.

In the frequency domain, the use of variance ratios for unit root testing is motivated by the fact that the spectrum of a unit root process peaks at the near zero frequencies, and tails off exponentially. As a consequence, the largest proportion of the variance is found in the lowest frequency bands. Suitable test statistics can, therefore, be based on the relative distribution of the variance with regards to frequency. For this to be feasible, the spectral variance needs to be decomposed in order to obtain the proportions of the variance contributed by the different frequency intervals. The Discrete Wavelet Transform (DWT) is a variance preserving transform, which decomposes the spectral variance on a scale-by-scale basis using filtering operations. The transform outputs two vectors; a vector of the DWT wavelet coefficients, and a vector of its scaling coefficients. The wavelet coefficients describe the changes at each scale, i.e. the details resulting from differences within each scale. The scaling coefficients, on the other hand, describe averages at each scale, i.e. the smooth resulting from averaging at each scale. The scale of the transform, which is inversely related to frequency, refers to the number of the recursive decompositions. Each recursive iteration from the second onwards decomposes the scaling coefficients from the preceding iteration.

The DWT has its filters operate on non-overlapping values, which means that the input time series have to be of dyadic lengths ( $2^k, k = 2, 3, \dots$ ). In contrast, the Maximal Overlap DWT (MODWT) has its filters operate on overlapping values, which makes it possible to handle samples of any size. The MODWT, therefore, extracts more information on the local variation of the time series. Unlike wavelet functions with longer and smoother filters, the Haar MODWT does not suffer from boundary effects, i.e. the loss of coefficients which are subject to circular filtering operations at the end of the time series. The transform also provides a better estimator of the wavelet variance (see Percival 1995) compared to the DWT. For these reasons, we use the Haar MODWT in this paper. For more details on wavelet filters and their properties, we refer to texts by Percival and Walden (2000) and Gençay, Selçuk, and Whitcher (2001).

The Haar MODWT wavelet filter ( $h_{j,l} : l = 0, \dots, L_j - 1, j = 1, 2, \dots$ ) is given as,

$$h_{j,l} \equiv \begin{cases} \frac{1}{2^j} & \text{for } l = 0, \dots, 2^{j-1} - 1 \\ \frac{1}{-2^j} & \text{for } l = 2^{j-1}, \dots, 2^j - 1 \\ 0 & \text{otherwise} \end{cases}$$

$L_j = 2^j$  for this wavelet, and is the length of filter at scale  $j$ . The scaling filter is given as,

$$g_{j,l} \equiv \begin{cases} \frac{1}{2^j} & \text{for } l = 1, \dots, 2^j - 1 \\ 0 & \text{otherwise} \end{cases}$$

The Haar wavelet filter, therefore, approximates a band-pass filter with the nominal pass-band  $[2^{-(j+1)}, 2^{-j}]$  and the Haar scaling filter approximates an ideal low-pass filter with the nominal pass-band  $[0, 2^{-(j+1)}]$ .

The  $j$ th level wavelet and scaling coefficients are defined as follows respectively:

$$\begin{aligned}\tilde{W}_{t,j} &\equiv \sum_{l=0}^{L_j-1} h_{j,l} y_{t-l}, \quad (t = 0, 1, 2, \dots) \\ \tilde{V}_{t,j} &\equiv \sum_{l=0}^{L_j-1} g_{j,l} y_{t-l}, \quad (t = 0, 1, 2, \dots)\end{aligned}$$

A useful property of the Haar MODWT transform is,

$$y_t = \sum_{j=1}^{J_0} \tilde{W}_{t,j} + \tilde{V}_{t,J_0}$$

where  $J_0$  is an arbitrary scale less than or equal to the maximum resolution of the time series. This property implies that the time series itself (not only its variance) can be additively decomposed into its wavelet and scaling coefficients.

For univariate time series, the wavelet variance ratio unit root test introduced by Fan and Gençay (2010) use a normalized version of the test statistic given below,

$$\tilde{S}_{T,1} = \frac{\|\tilde{V}_1\|^2}{\|\tilde{V}_1\|^2 + \|\tilde{W}_1\|^2}$$

The numerator is the contribution to the variance from the first level scaling coefficients, and the denominator is the total variance partitioned into the parts contributed by the scaling and wavelet coefficients of the first scale, respectively. The variances of the scaling and wavelet coefficients are given as,

$$\frac{1}{T-1} \|\tilde{V}_1\|^2 = \frac{1}{T} \sum_{t=0}^{T-1} \tilde{V}_{t,1} \text{ and } \frac{1}{T-1} \|\tilde{W}_1\|^2 = \frac{1}{T} \sum_{t=0}^{T-1} \tilde{W}_{t,1}$$

respectively.

Under the unit root null hypothesis, it can be seen that  $\sum_{t=0}^{T-1} \tilde{W}_{t,1}^2 = O_p(T)$  and it is shown (see Fan and Gençay, 2010) that  $\sum_{t=0}^{T-1} \tilde{V}_{t,1}^2 = O_p(T^2)$ . The variance ratio therefore takes the form,

$$\tilde{S}_{T,1} = \frac{\|\tilde{V}_1\|^2}{\|\tilde{V}_1\|^2 + \|\tilde{W}_1\|^2} = 1 + o_p(1)$$

The limiting distribution of the test statistic, which is a normalized version of  $\tilde{S}_{T,1}$ , is non-standard and the critical values are obtained using Monte Carlo simulations.

Under the alternative hypothesis both  $\sum_{t=0}^{T-1} \tilde{W}_{t,1}^2$  and  $\sum_{t=0}^{T-1} \tilde{V}_{t,1}^2$  are stationary and the ratio  $\tilde{S}_{T,1}$  will be less than 1.

Li and Shukur (2013) proposed using  $\hat{S}_{T,1}$  in a panel data setting. Their test statistic,  $\bar{S}_{NT}$ , is based on averaging the variance ratios of the individual panel units, i.e.  $\bar{S}_{NT} = N^{-1} \sum_{i=1}^N \hat{S}_{T,1i}$ . The test was conducted on cross-sectionally correlated time series as well as series that were decorrelated by wavestrapping. Monte Carlo simulation results showed that the wavelet-based panel unit root test is more powerful than the IPS test in the presence of correlation among the panel units. The test is also more robust to size distortions resulting from cross-sectional dependency compared to the IPS test, but still over-sized.

## 2.2 The wavelet variance ratio unit root test for a system of equations

Consider the system of equations without deterministic terms for simplicity,

$$y_{it} = \phi_i y_{i,t-1} + u_{it}, \quad i = 1, \dots, N; \quad t = 1, \dots, T$$

where  $y_{it}$  is the time series of interest and  $u_{it}$  is a zero mean weakly stationary error term, i.e.  $u_{it} = \sum_{j=0}^{\infty} \psi_j L^j \varepsilon_{it}$  (with finite long-run non-zero variance  $\sigma_{\varepsilon_{it}}^2 \psi(1)^2 < \infty$ , and  $\psi(1) \neq 0$ ),  $i$  indexes the individual equation, and  $t$  indexes time. Also  $\text{Cov}(u_{it}, u_{kt}) = 0$  for  $i \neq k$ , and  $\varepsilon_{it}$  is an iid, zero mean process with variance  $\sigma_{\varepsilon_{it}}^2$ .  $\psi(L)$  is the lag polynomial that relates the response of  $u_{it}$  to  $\varepsilon_{it}$ .

The unit root hypothesis for the system,

$$H_0 : |\phi_i| = 1 \text{ for all } i$$

and the alternative hypothesis is,

$$H_A : |\phi_i| < 1, i = 1, \dots, N_1 \text{ and } |\phi_i| = 1, N_1 + 1, \dots, N$$

Let the matrix of time series be denoted by,

$$\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N]$$

so that the Haar MODWT scaling and wavelet coefficients for the first scale decomposition are given by,

$$\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_N] \text{ and } \mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_N] \text{ respectively,}$$

where  $\mathbf{v}_i$  is the vector of the scaling coefficients of the series  $\mathbf{y}_i$  ( $i = 1, 2, \dots, N$ ), i.e.  $V_{ti,1}, \dots, V_{Ti,1}$  and  $\mathbf{w}_i$  is the vector of the wavelet coefficients of series  $\mathbf{y}_i$  ( $i = 1, 2, \dots, N$ ), i.e.  $W_{ti,1}, \dots, W_{Ti,1}$

A wavelet variance ratio unit root test can be based on the following,

$$\text{VR} = \text{tr} \left( (\mathbf{V}^T \mathbf{V} + \mathbf{W}^T \mathbf{W}) (\mathbf{W}^T \mathbf{W})^{-1} \right)$$

where,

$$\text{tr}(\mathbf{V}^T \mathbf{V} + \mathbf{W}^T \mathbf{W})$$

is the total variance of the system and,

$$\text{tr}(\mathbf{W}^T \mathbf{W})^{-1}$$

is the variance contributed by the first scale wavelet coefficients.

Under the null hypothesis (where all the series in the equation system are  $I(1)$ ),  $\mathbf{V}^T \mathbf{V}$  is a diagonal matrix with diagonal elements being of order  $O_p(T^2)$ . The diagonal elements of  $\mathbf{V}^T \mathbf{V}$  will dominate those of  $\mathbf{W}^T \mathbf{W}$  which are of order of convergence  $O_p(T)$ . The test statistic will, therefore, take on larger values under the null hypothesis compared the values under the alternative. For white noise series, for example,

$$\text{VR} = \text{tr}(\mathbf{I}_N + \mathbf{V}^T \mathbf{V} (\mathbf{W}^T \mathbf{W})^{-1}) = 2N + o_p(1) \text{ as } T \rightarrow \infty$$

since  $\mathbf{V}^T \mathbf{V} = \mathbf{W}^T \mathbf{W}$ .

VR is not bounded under the null hypothesis. A suitably normalized test statistic given as follows,

$$\text{VR}_M = \frac{1}{T} \text{tr}(\hat{\Gamma} (\mathbf{V}^T \mathbf{V} + \mathbf{W}^T \mathbf{W}) \mathbf{W}^T \mathbf{W}^{-1})$$

where  $\hat{\Gamma}$  is a diagonal matrix with the diagonals consisting of the weights  $\hat{v}_{yi,1}/\hat{\omega}_i^2$ , and  $\hat{v}_{yi,1}$  and  $\hat{\omega}_i^2$  are consistent estimates of the wavelet and long-run variances respectively. The two variances, which enter the limiting distribution as nuisance parameters, are consistently estimated as is shown in the Appendix.

The limiting distribution of the test statistic under the null hypothesis is shown in the following theorem whose proof is given in the Appendix.

**Theorem 1**

The limiting distribution of  $VR_M$  under  $H_0$  is given as,

$$VR_M \Rightarrow \sum_{i=1}^N \int_0^1 [W_i(r)]^2 dr$$

where  $W_i(r), i = 1, 2, \dots, N$  are independent standard Brownian motions, and  $N$  is the number of equations.

Since the test statistic is the sum of variance ratios, a Central Limit Theorem could be invoked by normalizing the sum and letting  $N \rightarrow \infty$ , but this is not pursued in this paper as we are mainly interested in the limit only where  $T \rightarrow \infty$ .

Many unit root tests suffer from loss of power when tested against alternatives that are trend stationary. As a consequence of this, efficient detrending methods (see Schmidt and Phillips, 1966) are required to retain power. We use the detrending techniques suggested by Fan and Gençay (2010), and, as in their work, we restrict our scope to the cases where the models specified under the alternative hypotheses have non-zero means and linear trends only.

The model including deterministic components is given as,

$$y_{i,t} = \mu_i + \alpha_i t + \phi_i y_{i,t-1} + u_{it} \quad i = 1, \dots, N; \quad t = 1, \dots, T \quad (1)$$

For equation  $i$ , the null hypothesis,  $H_0 : \phi_i = 1$  is the unit root hypothesis while under  $H_A$ ,  $\phi_i < |1|$  is the hypothesis of stationarity. Following Fan and Gençay (2010), when  $\alpha = 0$ , we consider the demeaned series  $(y_{it} - \bar{y}_{it})$  where  $\bar{y}_{it} = T^{-1} \sum_{t=1}^T y_{it}$  is the individual's average. Similarly, we consider the detrending  $(\tilde{y}_{it} - \bar{\tilde{y}}_i)$ , when  $\alpha \neq 0$ .  $\tilde{y}_{it} = T^{-1} \sum_{t=1}^T (\Delta y_{it} - \bar{\Delta y}_i)$ ,  $\bar{\tilde{y}}_i$  is the sample mean of  $\tilde{y}_{it}$  for individual  $i$ ,  $\Delta y_{it} = y_{it} - y_{i,t-1}$  and  $\bar{\Delta y}_i$  is the sample mean of  $\Delta y_{it}$  for individual  $i$ .

Let the test statistics be denoted by  $VR_M^M$  and  $VR_M^d$  for the cases where  $\alpha = 0$  and  $\alpha \neq 0$  respectively (see Eqn. (1)). The limiting distributions of these statistics are given by Theorem 2 below. The derivations of these limiting distributions, which are also given in the Appendix, are similar to that given for the distribution of the test statistic given in Theorem 1, except that detrended Brownian motions are used.

**Theorem 2**

The limiting distributions of  $VR_M^M$  and  $VR_M^d$  under  $H_0$  are given as,

$$VR_M^M \Rightarrow \sum_{i=1}^N \int_0^1 [W_i^\mu(r)]^2 dr \text{ and}$$

$$VR_M^d \Rightarrow \sum_{i=1}^N \int_0^1 [V_i^\mu(r)]^2 dr$$

respectively.

Where  $\Rightarrow$  denotes convergence in the associated probability measure,  $W^\mu(r) = W(r) - \int_0^1 W(r) dr$ , is the demeaned Brownian motion, and  $V^\mu(r) = V(r) - \int_0^1 V(r) dr$  is the detrended Brownian motion, with  $V(r) = W(r) - rW(1)$ .

**2.3 Comparison unit root test**

The small sample properties of  $VR_M$  are compared to CIPS. Pesaran (2007) constructs the CIPS test based the following model:

$$y_{it} = (1 - \phi_i)\mu_i + \phi_i y_{i,t-1} + u_{it}$$

where the initial values  $y_{i0}$  are fixed. A single common factor with individual specific factor loadings is specified for the error term,

$$u_{it} = \lambda_i f_t + \epsilon_{it}$$

$\epsilon_{it}$ ,  $i = 1, \dots, N, t = 1, \dots, T$ , are zero mean errors with heterogeneous variances  $\sigma_i^2$ . The common factor,  $f_t$ , is assumed to be stationary and serially uncorrelated, and without loss of generality, its variance is fixed at 1, i.e.  $\sigma_f^2 = 1$ . Cross-sectional correlations are introduced by the factor loadings  $\lambda_i$ , themselves random variables.  $\epsilon_{it}$ ,  $\lambda_i$  and  $f_t$  are assumed to be mutually independent.

Pesaran (2007) proposes a test that augments the standard Dickey-Fuller test with the cross-sectional averages, resulting in the following Cross-sectionally Augmented Dickey-Fuller (CADF) estimating equation,

$$\Delta y_{it} = a_i + b_i y_{i,t-1} + c_i \bar{y}_{t-1} + d_i \Delta \bar{y}_t + \epsilon_{it}$$

where  $\bar{y}_t$  are the cross-sectional averages, and lags of  $\Delta y_{it}$  and  $\Delta \bar{y}_t$  may be included to whiten the residuals. The cross-sectional averages are used as proxies for the unobservable common factor.

Letting  $CADF_i$  represent the CADF statistic for equation  $i$ , CIPS is the average of the CADFs over all the equations,

$$CIPS = \frac{1}{N} \sum_{i=1}^N CADF_i$$

The small sample properties of CIPS are studied using Monte Carlo simulation in Pesaran (2007). The test is shown to be robust to size distortions even, in the presence of strong cross-sectional dependence and serial correlation, and has good power properties for sample sizes of between 50 and 100 for the DGPs considered therein.

In the following section, we examine the performance of  $VR_M$ , and make comparisons with that of CIPS for time series of lengths 20–100, using systems of 5 and 10 equations. The size and power of the tests are compared in cases where there is neither cross-sectional dependency nor serial correlation (hereafter called DGP 1), in the presence of weak cross-sectional correlation but no serial correlation (hereafter called DGP 2), in the case where there is strong cross-sectional correlation but no serial correlation (hereafter called DGP 3), and in the case with both strong cross-sectional correlation and strong serial correlation (hereafter called DGP 4). The choice of DGPs follows that of Pesaran (2007).

### 3 Monte Carlo simulations

Monte Carlo simulations were used to study the size and power properties of the two unit root tests in small sample sizes. The design of the Monte Carlo experiments is discussed next.

#### 3.1 Design of the Monte Carlo experiment

Following Pesaran (2007), time series are simulated using the following DGPs

$$\begin{aligned} y_{it} &= (1 - \phi_i) \mu_i + \phi_i y_{i,t-1} + u_{it}, \quad |\phi| < 1 \\ u_{it} &= \gamma_i f_t + v_{it}, \quad f_t \sim \text{iid } N(0, 1) \\ v_{it} &= \rho_i v_{i,t-1} + \epsilon_{it}, \quad |\rho| < 1 \end{aligned}$$

$$\begin{aligned} \epsilon_{it} &\sim \text{iid } N(0, \sigma_i^2) \\ \sigma_i^2 &\sim U[0.5, 1.5] \\ \mu_i &\sim U[0, 0.02] \end{aligned}$$

Cross-sectional correlation is introduced using a single common factor denoted by  $f_t$ , and represents the unobserved common factor effect.

Table 1 shows the experimental factors and the ranges over which they are varied. The nominal test size is held at 5% as per convention.



**Table 1:** Factors that vary for the different DGPs.

Factor	Symbol	Design
Nominal size	$\pi_0$	0.05
Number of iterations	$I$	10,000 (Size) 50,000 (Critical values) 10,000 (Power)
Number of equations	$N$	5, 10
Number of observations	$T$	20, 30, 50, 100
Common factor loadings (Cross-sectional correlation)	$\gamma_i$	0 (No correlation) $\sim U[0, 0.2]$ (Weak) $\sim U[-1, 3]$ (Strong)
AR parameter for serial correlation	$\rho_i$	$\sim U[0.2, 0.4]$
AR parameter for alternatives	$\phi_i$	$\sim U[0.85, 0.95]$

The common factor loadings are sampled from the uniform distribution with parameters  $U[0, 0.2]$  and  $U[-1, 3]$  for weak and strong cross-sectional dependence, respectively. This corresponds to cross-sectional correlations between the equations of 1% and 50% on average, respectively.

## 4 Results and discussions

### 4.1 Empirical test sizes and power

Monte Carlo simulations were conducted for two purposes; to study the small sample performance of  $VR_M$  by comparing it CIPS, and to study the robustness of  $VR_M$  to cross-sectional and serial dependence. We examine these aspects in the cases where testing is against an alternative that has zero mean and no time trend, and in the case where testing is against an alternative that is stationary around a non-zero mean only. The test statistic used in both cases is  $VR_M^M$  since both tests correspond to  $\alpha = 0$  but differ in their specification of  $\mu_i$  (see Eqn. (1))

#### 4.1.1 Case I. No deterministic terms

The 1%, 5% and 10% critical values for  $VR_M^M$  and CIPS test statistics are shown in Table 2. These critical values correspond to the case where no deterministic terms are assumed.

Table 3 shows the test sizes for the the four DGPs given earlier. There is no evidence of size distortions any of the DGPs for both tests.

**Table 2:** Case I (no deterministic components): critical values.

	T	1%		5%		10%	
		$VR_M^M$	CIPS	$VR_M^M$	CIPS	$VR_M^M$	CIPS
N = 5	20	105.55	-2.29	220.86	-1.92	321.06	-1.72
	30	66.35	-2.28	135.69	-1.91	196.64	-1.72
	50	38.35	-2.24	77.86	-1.91	111.35	-1.71
	100	19.09	-2.23	38.25	-1.89	54.62	-1.72
N = 10	20	693.17	-1.97	1185.68	-1.69	1529.02	-1.55
	30	398.98	-1.95	637.69	-1.70	799.70	-1.56
	50	214.04	-1.95	333.68	-1.70	422.6	-1.56
	100	100.18	-1.93	154.13	-1.71	192.14	-1.57

Each individual series generated using the DGP  $y_{it} = y_{i,t-1} + f_t + \varepsilon_{it}$  for  $i = 1, \dots, N; t = 1, \dots, T$  with  $f_t$  and  $\varepsilon_{it} \sim \text{iid } N(0, 1)$ .

For CIPS, the critical values are calculated from the regression of  $\Delta y_{it}$  on  $y_{i,t-1}$ ,  $\bar{y}_{t-1}$  and  $\Delta \bar{y}_t$ . The cross-sectional mean is  $\bar{y}_t = \sum_{i=1}^N y_{it}$ . For the  $VR_M^M$ , the critical values are given by the quantiles of the empirical distribution of the test statistic with no adjustment made for the deterministic components.



The empirical power of the two tests are also displayed in Table 3. For the case where there are no deterministic components, it is clear that the  $VR_M^M$  is more powerful than the CIPS test for all the DGPs and the sample sizes considered, as well as for both equation systems. The power of CIPS increases with sample size for all the DGPs. The increase in power with increasing sample size is slowest for the DGP that has the combination of the strongest cross-sectional dependence and serial correlation (DGP number 4). For the 5 equations system, the highest power achieved by the CIPS test is only 83.7%.

**Table 3:** Case I (no deterministic components): test sizes.

T	DGP 1		DGP 2		DGP 3		DGP 4	
	$VR_M^M$	CIPS	$VR_M^M$	CIPS	$VR_M^M$	CIPS	$VR_M^M$	CIPS
Size	N = 5							
20	0.055	0.054	0.050	0.050	0.050	0.058	0.049	0.054
30	0.049	0.051	0.051	0.051	0.051	0.050	0.048	0.051
50	0.050	0.051	0.050	0.050	0.052	0.054	0.046	0.051
100	0.051	0.049	0.053	0.053	0.508	0.055	0.046	0.049
Size	N = 10							
20	0.050	0.048	0.046	0.055	0.048	0.053	0.047	0.040
30	0.049	0.050	0.051	0.052	0.051	0.057	0.046	0.046
50	0.050	0.051	0.049	0.049	0.053	0.053	0.049	0.035
100	0.047	0.055	0.046	0.051	0.051	0.063	0.044	0.041
Power	N = 5							
20	1.000	0.134	1.000	0.132	1.000	0.142	1.000	0.091
30	1.000	0.232	1.000	0.211	1.000	0.234	1.000	0.134
50	1.000	0.508	1.000	0.496	1.000	0.503	1.000	0.313
100	1.000	0.960	1.000	0.956	1.000	0.952	1.000	0.837
Power	N = 10							
20	1.000	0.194	1.000	0.188	1.000	0.191	1.000	0.101
30	1.000	0.357	1.000	0.362	1.000	0.366	1.000	0.194
50	1.000	0.796	1.000	0.782	1.000	0.788	1.000	0.519
100	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.988

#### 4.1.2 Case II. Non-zero mean and no time trend

The critical values for  $VR_M^M$  and CIPS are given in Table 4. These values correspond to the case where a non-zero mean but no time trend is assumed. For the CIPS test, the estimating equation fits an intercept but no time trend, and for  $VR_M^M$  the data are demeaned prior to performing the unit root test (see the discussion on tests against trend stationary alternatives in page 10).

**Table 4:** Case II (intercept only): critical values.

		1%		5%		10%	
	T	$VR_M^M$	CIPS	$VR_M^M$	CIPS	$VR_M^M$	CIPS
$N = 5$	20	1.07	−2.99	1.22	−2.59	1.32	−2.40
	30	1.00	−2.91	1.17	−2.56	1.27	−2.39
	50	0.96	−2.87	1.12	−2.55	1.22	−2.38
	100	0.92	−2.86	1.09	−2.54	1.19	−2.38
$N = 10$	20	2.58	−2.61	2.84	−2.35	2.98	−2.22
	30	2.41	−2.58	2.67	−2.33	2.83	−2.21
	50	2.28	−2.55	2.56	−2.33	2.71	−2.22
	100	2.20	−2.53	2.47	−2.32	2.64	−2.22

Each individual series generated using the DGP  $y_{it} = y_{i,t-1} + f_t + \varepsilon_{it}$  for  $i = 1, \dots, N; t = 1, \dots, T$  with  $f_t$  and  $\varepsilon_{it} \sim \text{iid } N(0, 1)$ . For CIPS, The critical values are calculated from the regression of  $\Delta y_{it}$  on a constant,  $y_{i,t-1}$ ,  $\bar{y}_{t-1}$  and  $\Delta \bar{y}_t$ . The cross-sectional mean is  $\bar{y}_t = \sum_{i=1}^N y_{it}$ . For the  $VR_M^M$ , the critical values are given by the quantiles of the empirical distribution of the test statistic for the demeaned series.

Table 5 shows the test sizes for the 4 DGPs. Again there is no evidence of size distortions for all the DGPs using both tests.

Table 5 also displays the power of the two tests. Both tests show low power for the smallest sample sizes ( $T = 20, 30$ ) but the power increases with sample size. Both tests show decreasing power when strong cross-sectional and serial correlation are present. For the 5 equations system,  $VR_M^M$  has higher power than the CIPS for sample sizes of larger than 20. For the 10 equations system,  $VR_M^M$  is more powerful than CIPS for sample sizes of  $T = 50$  and  $T = 100$ . For the smaller sample sizes, both tests show similar (but low) power. As expected, the 10 equations system shows more power than the 5 equations system for both tests.  $VR_M^M$  has noticeably higher power than CIPS in the presence of both serial and cross-sectional correlation.

**Table 5:** Case II (intercept only): test sizes and power.

T	DGP 1		DGP 2		DGP 3		DGP 4	
	$VR_M^M$	CIPS	$VR_M^M$	CIPS	$VR_M^M$	CIPS	$VR_M^M$	CIPS
Size	N = 5							
20	0.047	0.048	0.050	0.051	0.045	0.050	0.037	0.050
30	0.048	0.049	0.055	0.047	0.049	0.056	0.042	0.044
50	0.050	0.047	0.049	0.051	0.047	0.049	0.045	0.044
100	0.051	0.052	0.050	0.049	0.053	0.058	0.041	0.051
Size	N = 10							
20	0.047	0.054	0.050	0.050	0.046	0.058	0.049	0.055
30	0.050	0.051	0.045	0.051	0.049	0.060	0.044	0.044
50	0.047	0.051	0.049	0.050	0.044	0.054	0.050	0.044
100	0.051	0.049	0.050	0.053	0.052	0.055	0.048	0.038
Power	N = 5							
20	0.089	0.079	0.089	0.086	0.085	0.089	0.078	0.077
30	0.157	0.111	0.151	0.113	0.157	0.119	0.147	0.087
50	0.368	0.227	0.385	0.224	0.363	0.235	0.350	0.156
100	0.905	0.729	0.918	0.706	0.886	0.722	0.885	0.533
Power	N = 10							
20	0.098	0.093	0.097	0.092	0.092	0.100	0.098	0.073
30	0.178	0.159	0.184	0.165	0.171	0.165	0.170	0.099
50	0.511	0.373	0.535	0.371	0.479	0.391	0.491	0.225
100	0.986	0.964	0.993	0.959	0.976	0.954	0.977	0.842

The power advantage of the wavelet-based unit root test over CIPS could be explained by the differences in effective sample sizes. While the wavelet-based test loses power when the series are demeaned or detrended, CIPS requires the estimation of several parameters for each individual time series, which means that the effective sample size is reduced (hence the loss of power).

## 5 Empirical application

Purchasing Power Parity (PPP) has been heavily researched in international economics because of its central role in building macroeconomic models. There are two different versions of PPP; the absolute PPP, which refers to the situation where the nominal exchange rate between two currencies is equal to the ratio of the price levels of the two corresponding countries, and the relative PPP which takes into account factors such as trade barriers (tariff and non-tariff barriers), transportation costs, and product differentiation across countries. The empirical literature has focused on the relative version the PPP which is the weaker version of the macroeconomic theory. In the relative version, the rate of depreciation of a currency equals the difference in price inflation of that country's currency and the price inflation in the comparative country, making the real exchange rate constant.

The conventional procedure when evaluating PPP is to test the null hypothesis that the real exchange rate series has a unit root against the alternative hypothesis of being stationary. Rejection of the null hypothesis indicates support for the PPP theory. Initial studies using augmented Dickey-Fuller (ADF) unit root tests suggested by Dickey and Fuller (1979) showed little evidence supporting PPP in the long-run. An example of such a study is Taylor (1988) where the conclusions were very unfavorable to PPP as a long-run equilibrium condition. Other examples of such studies include Corbae and Ouliaris (1988), Layton and Stark (1990), Corbae and Ouliaris (1991), and Bahmani-Oskooee (1993, 1995). However, Frankel and Rose (1996) noted that a non-rejection of the null hypothesis may be due to low statistical power of the unit root tests, which is mainly caused by the lack of data. Glen (1992), Lothian and Taylor (1996), and Taylor (2002) among others suggest that longer time series could be used to provide indirect evidence to support PPP. However, these long-span studies also faced

the criticism (see Hegwood and Papell, 1998, for example) that structural breaks or shifts in the equilibrium exchange rates are possibly generated during the long time span, thereby biasing the results. An alternative approach, which can be used to increase the statistical power, is to utilize the cross-sectional dimension of multiple time series. Examples of studies using panel unit root tests are Cheung, Chinn, and Fujii (2006) and Murray and Papell (2002, 2005).

As an empirical application we use the PPP theory and compare the evidence found from using CIPS and  $VR_M^M$ . The data are on the real exchange rates of the Group of 7 (G7) countries (source: Bruegel website: <http://bruegel.org/>) and covered the time span between 1960 and 2015. To avoid the potential bias from a structural break, we consider data from the post-Bretton Woods period (the period after currencies were unpegged from the dollar, spanning from 1972 to 2015) in a separate analysis. The results are shown in Table 6. The data are demeaned so as to study the relative version of PPP. The results confirm those from the simulation study and conclude that  $VR_M^M$  rejects the null hypothesis for both sample periods. No such support for PPP is found using CIPS for either of the sample periods at the 5% significance level.

The CIPS test is the conventional Pesaran (2007) test. The  $VR_M^M$  test is based on the MODWT using the Haar filter. This method, as noted in Section 2 when describing the wavelet filtering, is chosen since it extracts more information than the DWT. The method can also handle samples of any sizes. By using the Haar filter we also avoid loss of information due to boundary coefficients. The  $VR_M^M$  test is, as discussed in Fan and Gençay (2010), is based on the observation by Granger (1966) that the spectral density of trending time series, such as the real exchange rates, are characterized by a significant power in low frequencies followed by exponential decline at higher frequencies. The more powerful wavelet-based test for unit roots suggested by Fan and Gençay (2010) used this notion and the ability of wavelets to decompose the variance of a time series at different frequencies. Capitalizing on the idea of Granger (1966) and the decomposing ability of wavelets Fan and Gençay (2010) constructed the variance ratio test which we generalize to systems of equations. The results from our simulation study and the Fan and Gençay (2010) study indicates that this type of wavelet variance ratio test is more powerful than the traditional parametric options such as the ADF test for univariate time series and the CIPS test for systems of equations. This is the main reason why the  $VR_M^M$  test is able to reject the null hypothesis of a non-stationary system of equations

Table 6: Empirical example.

Full sample		Post-Bretton Woods	
$VR_M^M$	CIPS	$VR_M^M$	CIPS
1.562*	-2.006	1.661*	-2.078

A star indicates significance at the conventional 5% level of significance.

## 6 Summary and conclusions

A unit root test for a system of equations is introduced in this paper. The proposed test extends the wavelet variance ratio unit root test of Fan and Gençay (2010) to multiple equation time series.

Monte Carlo simulations show that the proposed test has higher power compared to CIPS (Pesaran 2007) for time series of short length (between 20 and 100 observations), and systems of 5 and 10 equations. The test is also shown to be robust to cross-sectional dependency and serial correlation for the DGPs considered in this paper.

We demonstrate its usefulness through an empirical application on evaluating the PPP theory for the G7 countries. Evidence from this evaluation points to different countries following different specifications, with some having stationary exchange rate series.

The proposed unit root test is simple to apply and interpret, and could prove to be useful to the practitioner who is faced with a system of 10 or fewer equations and time series of lengths up to 100. For larger systems of equations or systems with longer time series, any of the existing unit root tests should provide adequate power.

## Acknowledgement

The first two authors would like to acknowledge the contribution of the third author, Ghazi Shukur, who passed away during the review period of the manuscript. We would also like to thank the anonymous reviewers for their suggestions.

## A Appendix

Consider the first-order autoregressive model:

$$y_{it} = \phi_i y_{i,t-1} + \mathbf{z}_{it}^T \boldsymbol{\gamma}_i + u_{it}$$

where  $y_{it}$  is the time series of interest,  $i = 1, \dots, N$  are the individual time series,  $t = 1, \dots, T$  indexes time,  $\mathbf{z}_{it}^T$  are the deterministic components, and  $u_{it}$  is a weakly stationary zero mean process with finite long-run variance.

Here we consider the cases with no deterministic terms as well as where the alternative is trend stationary around a non-zero mean and time trend.

*Proof of Theorem Theorem 1*

Consider  $N$  time series that have no cross-sectional correlation but are possibly autocorrelated,

$$\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N]$$

The Haar MODWT scaling and wavelet coefficients for the first scale decomposition are given by,

$$\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_N] \text{ and } \mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_N] \text{ respectively,}$$

where  $\mathbf{v}_i$  is the vector of the scaling coefficients of the series  $\mathbf{y}_i$  ( $i = 1, 2, \dots, N$ ), i.e.  $V_{ti,1}, \dots, V_{Ti,1}$  and  $\mathbf{w}_i$  is the vector of the wavelet coefficients of series  $\mathbf{y}_i$  ( $i = 1, 2, \dots, N$ ), i.e.  $W_{ti,1}, \dots, W_{Ti,1}$ .

The total variance in the system of equations can be expressed in terms of the Haar MODWT coefficients as follows:

$$\text{tr} \left( T^{-1} (\mathbf{V}^T \mathbf{V} + \mathbf{W}^T \mathbf{W}) \right)$$

The contribution to the total variance due to the wavelet and scaling coefficients is given by,

$$\text{tr} \left( T^{-1} (\mathbf{V}^T \mathbf{V}) \right) \text{ and } \text{tr} \left( T^{-1} (\mathbf{W}^T \mathbf{W}) \right), \text{ respectively.}$$

A unit root test statistic can, therefore, be based on the following ratio,

$$\text{VR} = \text{tr} \left( T^{-1} (\mathbf{V}^T \mathbf{V} + \mathbf{W}^T \mathbf{W}) \mathbf{W}^T \mathbf{W}^{-1} \right)$$

Under the unit root null hypothesis, the diagonal elements of  $\mathbf{V}^T \mathbf{V} = O_p(T^2)$  while those of  $\mathbf{W}^T \mathbf{W} = O_p(T)$ . The variance ratio will take on large values. Whereas for stationary processes, both terms are  $O_p(T)$  and the ratio will be small.

Under the null hypothesis,

$$\begin{aligned} & T^{-1} (\mathbf{V}^T \mathbf{V} + \mathbf{W}^T \mathbf{W}) (\mathbf{W}^T \mathbf{W})^{-1} \\ &= \frac{1}{T} \begin{bmatrix} \alpha_{11} & o_p(1) & \cdots & o_p(1) \\ o_p(1) & \alpha_{22} & \cdots & o_p(1) \\ \vdots & & \ddots & \vdots \\ o_p(1) & o_p(1) & \cdots & \alpha_{NN} \end{bmatrix} \end{aligned}$$

as  $T \rightarrow \infty$  since there is no cross-correlation i.e.  $\mathbf{v}_i^T \mathbf{v}_j$  and  $\mathbf{w}_i^T \mathbf{w}_j = o_p(1)$ , and  $\alpha_{ii} = (\mathbf{v}_i^T \mathbf{v}_i + \mathbf{w}_i^T \mathbf{w}_i)(\mathbf{w}_i^T \mathbf{w}_i)^{-1}$ .

Then,

$$\text{tr} \left( T^{-1} (\mathbf{V}^T \mathbf{V} + \mathbf{W}^T \mathbf{W}) (\mathbf{W}^T \mathbf{W})^{-1} \right) = T^{-1} \sum_i^N (\mathbf{v}_i^T \mathbf{v}_i + \mathbf{w}_i^T \mathbf{w}_i) (\mathbf{w}_i^T \mathbf{w}_i)^{-1}$$

For the MODWT transform,

$$\text{VR}_M = \sum_{i=1}^N \frac{T^{-2} \left( \sum_{t=1}^T V_{ti,1}^2 + \sum_{t=1}^T W_{ti,1}^2 \right)}{T^{-1} \left( \sum_{t=1}^T W_{ti,1}^2 \right)},$$

Also, for this wavelet transform,

$$\sum_{t=1}^T V_{ti,1}^2 + \sum_{t=1}^T W_{ti,1}^2 = \sum_{t=1}^T y_{it}^2$$

From the asymptotic theory for unit root processes (see Hamilton (1994) pp. 486) and Continuous Mapping Theorem (CMT) (see Billingsley 1968),

$$\frac{1}{T^2} \sum_{i=1}^N \sum_{t=1}^T y_{it}^2 \Rightarrow \sum_{i=1}^N \omega_i^2 \int_0^1 [W_i(r)]^2 dr \quad (2)$$

where  $\omega_i^2$  is the long-run variance of  $u_{it}$ .

Also, for the Haar MODWT wavelet filter,

$$\frac{1}{T} \sum_{t=1}^T W_{ti,1}^2 \rightarrow E(W_{i,1}^2) \text{ as } T \rightarrow \infty$$

where  $E(W_{i,1}^2)$  is the first scale wavelet variance for series  $i$  (see Percival 1995)

Using the CMT, the limiting distribution of the variance ratio for each individual series is,

$$\frac{T^{-2} \left( \sum_{t=1}^T V_{ti,1}^2 + \sum_{t=1}^T W_{ti,1}^2 \right)}{T^{-1} \left( \sum_{t=1}^T W_{ti,1}^2 \right)} \Rightarrow \frac{\omega_i^2}{E(W_{i,1}^2)} \int_0^1 [W_i(r)]^2 dr$$

where the long-run variance of  $u_{it}$  (for time series  $i$ ) is given by,

$$\omega_i^2 = \lim_{T \rightarrow \infty} E(T^{-1} S_T^2) = \left[ \gamma_{i,0} + 2 \sum_{j=1}^{\infty} \gamma_{i,j} \right]$$

$S_T^2$  is the partial sum process of  $y_{i,t}^2$ ,  $\gamma_{i,j}$  is the lag  $j$  autocovariance for time series  $i$ , and  $\Rightarrow$  is used to denote convergence of the associated probability measure.

The two nuisance parameters in the limiting distribution,  $\omega_i^2$  and  $E(W_{i,1}^2)$ , can be consistently estimated as follows:

1.  $E(W_{i,1}^2)$  is the wavelet variance at unit scale of the Haar MODWT. Its consistent estimator (see Percival (1995)) is given by,

$$\hat{v}_{yi,1} = \frac{1}{T-1} \sum_{t=0}^{T-1} W_{ti,1}^2$$

The wavelet variance estimator for the Haar MODWT avoids boundary effects, which is the loss of coefficients at the ends of time series as a result of circular filtering operations.

2. For the long-run variance,  $\omega_i^2$ , estimation can be made in one of two ways (see Zivot and Wang, 2006, for details on the long-run variance estimation):

a. Parametric approach

For time series  $i$ , since  $\omega_i$  is a linear process, it follows that,

$$\sum_{j=-\infty}^{\infty} \gamma_j = \sigma^2 \left( \sum_{j=0}^{\infty} \psi_j \right)^2 = \sigma^2 \psi(1)^2$$

giving  $\omega_i^2 = \sigma_i^2 \psi(1)^2$ .

When  $u_{it}$  is ARMA( $p, q$ ), then,

$$\psi(1) = \frac{1 + \theta_1 + \dots + \theta_q}{1 + \phi_1 + \dots + \phi_p} = \frac{\theta(1)}{\phi(1)},$$

which gives

$$\omega_i^2 = \frac{\sigma_i^2 \theta(1)^2}{\phi(1)^2}$$

where  $\sigma_i^2$  is the variance of the error of the ARMA model for time series  $i$ .

Making substitutions using the estimates of the parameters of the ARMA( $p, q$ ) process gives a consistent estimate of  $\omega_i^2$ .

A second parametric approach is to approximate the ARMA( $p, q$ ) process with a higher order AR( $p^*$ ) process,

$$u_{it} = \phi_{i,1} u_{i,t-1} + \dots + \phi_{i,p^*} u_{i,t-p^*} + \varepsilon_{it},$$

and then estimate the long-run variance as follows

$$\omega_i^2 = \frac{\sigma_i^2}{\phi^*(1)^2}$$

b. Semi-parametric method using a kernel function:

One possible semi-parametric estimator of the long-run variance is the Newey and West (1987) estimator, which is the weighted covariance function,

$$\omega_i^2 = \hat{\gamma}_{i,0} + 2 \sum_{\ell=1}^L w_{i,\ell} \hat{\gamma}_{i,\ell}$$

where  $w_{i,\ell}$  are the weights for time series  $i$ ,  $\hat{\gamma}_{i,\ell}$  are the autocovariances for time series  $i$ , and  $L$  is the truncation lag or bandwidth parameter, such that  $L = O(T^{1/3})$  (see Andrews 1991).

Newey and West use the Bartlett weights,

$$w_\ell = 1 - \frac{\ell}{L+1}$$

with  $L = \lfloor 4(T/100)^{2/9} \rfloor$ .

The nuisance parameters are eliminated from the limiting distribution by normalizing the variance ratio with the ratio of the consistent estimates of the nuisance parameters,

$$\frac{\hat{v}_{yi,1}}{\hat{\omega}_i^2} \frac{T^{-2} \left( \sum_{t=1}^T V_{ti,1}^2 + \sum_{t=1}^T W_{ti,1}^2 \right)}{T^{-1} \left( \sum_{t=1}^T W_{ti,1}^2 \right)} \Rightarrow \int_0^1 [W_i(r)]^2 dr$$

giving the result,

$$T^{-1} \sum_{i=1}^N (\mathbf{v}_i^T \mathbf{v}_i + \mathbf{w}_i^T \mathbf{w}_i) (\mathbf{w}_i^T \mathbf{w}_i)^{-1} \Rightarrow T^{-1} \sum_{i=1}^N \int_0^1 [W_i(r)]^2 dr$$

The test statistic is, therefore,

$$VR_M = \text{tr} \left( T^{-1} \hat{\Gamma} (\mathbf{V}^T \mathbf{V} + \mathbf{W}^T \mathbf{W}) \mathbf{W}^T \mathbf{W}^{-1} \right)$$

where  $\hat{\Gamma}$  is a diagonal matrix with the main diagonal consisting of the weights  $\hat{v}_{yi,1}/\hat{\omega}_i^2$ .

*Proof of Theorem Theorem 2*

Let  $\tilde{y}_{it}$  represent the time series adjusted for the deterministic components i.e.

$$\tilde{y}_{it} = (y_{it} - \hat{\mu}_i)$$

where  $\hat{\mu}$  is the estimate of the deterministic component. Then, from asymptotic theory for demeaned unit root processes (see Stock (1994), for example)

$$T^{-2} \sum_{t=1}^T \tilde{y}_{it}^2 \Rightarrow \omega_i^2 [W_i^\mu(r)]^2, \text{ where } W_i^\mu(r) = \omega_i \left( W_i(r) - \int_0^1 W_i(u) du \right)$$

so that

$$T^{-2} \sum_{t=1}^T \tilde{y}_{it}^2 \Rightarrow \omega_i^2 \left[ W_i(r) - \int_0^1 W_i(u) du \right]^2$$

where  $\omega_i^2$  is the long-run variance for equation  $i$ .

For the case where the time series are efficiently detrended, the following result holds (see Kwiatkowski et al. (1992), for example)

$$T^{-2} \sum_{t=1}^T \tilde{y}_{it} \Rightarrow \omega_i^2 \left[ \int_0^1 V^\mu(r) dr \right]^2$$

where  $V^\mu(r) = V(r) - \int_0^1 V(u) du$ ,  $V(r)$  is a standard Brownian bridge given as  $V(r) = W(r) - rW(1)$ , and  $W(r)$  is Brownian motion.

The rest of the proof follows that for Theorem 1. Starting with Eqn. (2), replacing  $W_i(r)$  with  $W_i^\mu(r)$  and  $V_i(r)$  with  $V_i^\mu(r)$  in the proof of Theorem 1 for the cases where the series have been demeaned and detrended respectively, leads to the limiting distributions of the two test statistics ( $VR_M^M$  and  $VR_M^d$ ) as given in Theorem 2.

## References

- Andrews, D. W. 1991. "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation." *Econometrica: Journal of the Econometric Society* 59: 817–858.
- Bahmani-Oskooee, M. 1993. "Purchasing Power Parity Based on Effective Exchange Rate and Cointegration: 25 LDCs' Experience With its Absolute Formulation." *World Development* 21 (6): 1023–1031.
- Bahmani-Oskooee, M. 1995. "Real and Nominal Effective Exchange Rates for 22 LDCs: 1971: 1–1990: 4." *Applied Economics* 27 (7): 591–604.
- Billingsley, P. 1968. *Convergence of Probability Measures*. New York: John Wiley & Sons.
- Breitung, J. 2002. "Nonparametric Tests for Unit Roots and Cointegration." *Journal of Econometrics* 108 (2): 343–363.
- Cheung, Y.-W., M. D. Chinn, and E. Fujii. 2006. "The Chinese Economies in Global Context: The Integration Process and its Determinants." *Journal of the Japanese and International Economies* 20 (1): 128–153.
- Choi, I. 2001. "Unit Root Tests for Panel Data." *Journal of International Money and Finance* 20 (2): 249–272.
- Cochrane, J. H. 1988. "How Big is the Random Walk in GNP?" *Journal of Political Economy* 96 (5): 893–920.
- Corbae, D., and S. Ouliaris 1988. "Cointegration and Tests of Purchasing Power Parity." *The Review of Economics and Statistics* 70: 508–511.



- Corbae, D., and S. Ouliaris. 1991. "A Test of Long-run Purchasing Power Parity Allowing for Structural Breaks." *Economic Record* 67 (1): 26–33.
- Dickey, D. A., and W. A. Fuller. 1979. "Distribution of the Estimators for Autoregressive Time Series with a Unit Root." *Journal of the American statistical association* 74 (366a): 427–431.
- Fan, Y., and R. Gençay. 2010. "Unit Root Tests with Wavelets." *Econometric Theory* 26 (05): 1305–1331.
- Fisher, R. 1932. *Statistical Methods for Research Workers* (Edinburgh: Oliver and Boyd, 1925). Edinburgh: Oliver and Boyd.
- Frankel, J. A., and A. K. Rose. 1996. "A Panel Project on Purchasing Power Parity: Mean Reversion Within and Between Countries." *Journal of International Economics* 40: (1): 209–224.
- Gençay, R., F. Selçuk, and B. J. Whitcher. 2001. *An Introduction to Wavelets and Other Filtering Methods in Finance and Economics*. San Diego: Academic Press.
- Glen, J. D. 1992. "Real Exchange Rates in the Short, Medium, and Long Run." *Journal of International Economics* 33 (1–2): 147–166.
- Granger, C. 1966. "The Typical Spectral Shape of an Economic Variable." *Econometrica* 54: 257–287.
- Hamilton, J. D. 1994. *Time Series Analysis*, Vol. 2. Princeton: Princeton University Press.
- Hegwood, N. D., and D. H. Papell. 1998. "Quasi Purchasing Power Parity." *International Journal of Finance & Economics* 3 (4): 279–289.
- Im, K. S., M. H. Pesaran, and Y. Shin. 2003. "Testing for Unit Roots in Heterogeneous Panels." *Journal of econometrics* 115 (1): 53–74.
- Kwiatkowski, D., P. C. Phillips, P. Schmidt, and Y. Shin. 1992. "Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root: How Sure are we that Economic Time Series have a Unit Root?" *Journal of Econometrics* 54 (1–3): 159–178.
- Layton, A. P., and J. P. Stark. 1990. "Cointegration as an Empirical Test of Purchasing Power Parity." *Journal of Macroeconomics* 12 (1): 125–136.
- Levin, A., C.-F. Lin, and C.-S. J. Chu. 2002. "Unit Root Tests in Panel Data: Asymptotic and Finite-sample Properties." *Journal of econometrics* 108 (1): 1–24.
- Li, Y., and G. Shukur. 2013. "Testing for Unit Roots in Panel Data Using a Wavelet Ratio Method." *Computational Economics* 41 (1): 59–69.
- Lothian, J. R., and M. P. Taylor. 1996. "Real Exchange Rate Behavior: The Recent Float From the Perspective of the Past two Centuries." *Journal of Political Economy* 104 (3): 488–509.
- Maddala, G. S., and S. Wu. 1999. "A Comparative Study of Unit Root Tests with Panel Data and a New Simple Test." *Oxford Bulletin of Economics and statistics* 61 (S1): 631–652.
- Murray, C. J., and D. H. Papell. 2002. "The Purchasing Power Parity Persistence Paradigm." *Journal of International Economics* 56 (1): 1–19.
- Murray, C. J., and D. H. Papell. 2005. "Do Panels Help Solve the Purchasing Power Parity Puzzle?" *Journal of Business & Economic Statistics* 23 (4): 410–415.
- Newey, W. K., and K. D. West. 1987. "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation-consistent Covariance Matrix." *Econometrica* 55: 703–708.
- Percival, D. B. 1995. "On Estimation of the Wavelet Variance." *Biometrika* 82 (3): 619–631.
- Percival, D. B., and A. T. Walden. 2000. *Wavelet Methods for Time Series Analysis* (Cambridge Series in Statistical and Probabilistic Mathematics). Cambridge: Cambridge University Press.
- Pesaran, M. H. 2007. "A Simple Panel Unit Root Test in the Presence of Cross-section Dependence." *Journal of Applied Econometrics* 22 (2): 265–312.
- Schmidt, P., and P. C. Phillips. 1966. "LM Tests for a Unit Root in the Presence of Deterministic Trends." *Oxford Bulletin of Economics and Statistics* 34: 150–161.
- Stock, J. H. 1994. "Unit Roots, Structural Breaks and Trends." In *Handbook of Econometrics*, edited by R. Engle and D. McFadden, Vol. 4, chapter 46, 2752–2753. North Holland: Elsevier located in Amsterdam.
- Tanaka, K. 1990. "Testing for a Moving Average Unit Root." *Econometric Theory* 6 (04): 433–444.
- Taylor, M. P. 1988. "An Empirical Examination of Long-run Purchasing Power Parity Using Cointegration Techniques." *Applied Economics* 20 (10): 1369–1381.
- Taylor, A. M. 2002. "A Century of Purchasing Power Parity." *Review of Economics and Statistics* 84 (1): 139–150.
- Zivot, E., and J. Wang. 2006. *Modelling Financial Time Series with S-Plus*. New York: Springer-Verlag.