Linnaeus University Dissertations
Nr 309/2018

Abdul Aziz Ali

On the use of wavelets in unit root and cointegration tests

LINNAEUS UNIVERSITY PRESS

On the use of wavelets in unit root and cointegration tests

Linnaeus University Dissertations

No 309/2018

On the use of wavelets in unit root and cointegration tests

Abdul Aziz Ali

LINNAEUS UNIVERSITY PRESS

On the use of wavelets in unit root and cointegration tests
Doctoral Dissertation, Department of Economics and Statistics, Linnaeus University, Växjö, 2018

ISBN: 978-91-88761-26-2 (print), 978-91-88761-27-9 (pdf)
Published by: Linnaeus University Press, 351 95 Växjö
Printed by: DanagårdLiTHO, 2018

Abstract

Ali, Abdul Aziz (2018). *On the use of wavelets in unit root and cointegration tests,* Linnaeus University Dissertations No 309/2018, ISBN: 978-91-88761-26-2 (print), 978-91-88761-27-9 (pdf). Written in English.

This thesis consists of four essays linked with the use of wavelet methodologies in unit root testing and in the estimation of the cointegrating parameters of bivariate models.

In papers I and II, we examine the performance of some existing unit root tests in the presence of error distortions. We suggest wavelet-based unit root tests that have better size fidelity and size-adjusted power in the presence of conditional heteroscedasticity and additive measurement errors. We obtain the limiting distribution of the proposed test statistic in each case and examine the small sample performance of the tests using Monte Carlo simulations.

In paper III, we suggest a wavelet-based filtering method to improve the small sample estimation of the cointegrating parameters of bivariate models. We show, using Monte Carlo simulations, that wavelet filtering reduces the small sample estimation bias.

In paper IV, we propose a wavelet variance ratio unit root test for a system of equations. We obtain the limiting distributions of the test statistics under different specifications of the deterministic components of the estimating equations. We also investigate the small sample properties of the test by conducting Monte Carlo simulations. Results from the Monte Carlo simulations show that the test has good size fidelity for small sample sizes (of up to 100 observations per equation, and up to 10 equations), and has better size-adjusted power for these sample sizes, compared the Cross-sectionally Augmented Dickey-Fuller test.

Keywords: Time series; unit root; variance ratio; wavelets

To my family, and their families…

Acknowledgements

I would like to thank all the members of the teaching, research and administrative staff of the Department of Economics and Statistics at Linnaeus University, who have been cordial and helpful colleagues.

I am grateful to my main supervisor Prof. Ghazi Shukur for taking me as his student, for his guidance and supervision, and for supporting the flexible studying that fit around my family circumstances. Ghazi's enthusiasm for sharing his knowledge and wisdom in life, his generosity with his time, and his genuine interest in my wellbeing and that of my family, makes me feel very fortunate to be his student.

Thanks to my assistant supervisor Prof. Thomas Holgersson. I have had peace of mind knowing that Thomas has always been there, ready to help.

I am ever so grateful to Dr. Kristofer Månsson. Kristofer's guidance during the second half of my studies was critical to completing this work. Thanks for all the advice and for enduring me.

I would also like to thank Drs. Yushu Li and Abdulla Almasri for reviewing my work and for their comments and suggestions. Dr. Håkan Locking and Prof. Pär Sjölander always took interest in my work. Thanks for your questions and comments.

Thanks to those who I shared the office space with; Deliang, Abdulaziz and Chizheng. Although I was not in the office as often as I would have wanted to, I enjoyed their company and comradery whenever we met.

Finally, thanks to my siblings and their families who I have not connected with as much as I would have liked to lately. To my wife Balsam, and daughters Amani and Dalia, thanks for letting me to do this. I owe you so much time, and more.

To everyone else who I have not mentioned in person, thanks for your help and support, for your friendship, kindness and warmth, and for sharing your knowledge and your precious time.

Contents

2

2

Chapter 1

Introduction

1.1 Summary

This thesis consists of four essays linked with the use of wavelet methodologies in the analysis of non-stationary time series, focusing on small sample sizes. In two of the essays, the performance of unit root tests has been examined in the presence of conditionally heteroscedastic and additive measurement errors. Wavelet filters that remove the periodic components, which correspond to these error distortions are suggested together with wavelet variance ratio unit root test statistics. In both cases, we obtain the limiting distribution of the test statistics theoretically, and use Monte Carlo simulation for the assessment of size and size-adjusted power of the suggested unit root tests. In the third essay we examine the small sample bias present in the estimation of the cointegrating coefficient of bivariate cointegration models. The results from simulation studies, using a variety of Data Generating Processes (DGPs) show that the wavelet filters, when used in conjunction with three standard estimation methods, considerably improve the signal-to-noise ratios in bivariate cointegration models, thereby reducing the small sample bias. In the fourth essay we suggest a wavelet variance ratio unit root test for a system of equations. We obtain the limiting distribution of the proposed test statistic and use Monte Carlo simulation to show that the test is robust to cross-equation correlation, retains its nominal size, and has good size-adjusted power in finite sample sizes.

1.2 Unit roots and unit root testing

An important consideration when modeling time series is the effect of a current shock on the long-run mean of the series. When the effect of a current shock is transient, the series will revert to its long-run mean or trend function and is said to be stationary. When the effect of a current shock is permanent, however, the series will not have a tendency to revert to its long-run mean, and will be characterized by a stochastic trend. Such a series is said to be non-stationary, integrated, or a unit root process. A difference-stationary time series is a series that can be made stationary by differencing. The degree of differencing required to achieve stationarity is referred to as the order of integration. The order of integration also refers to the number of unit roots in a time series; specifically, an $I(0)$ time series is stationary, while an $I(d)$ time series has d unit roots. It has been observed that the majority of economic time series are non-stationary, and in some cases present with unit roots occurring seasonally (monthly, quarterly, yearly, or at some other frequency), in what is called integrated seasonal time series.

While the relationships between stationary time series can be modeled using the method of Ordinary Least Squares (OLS), the modeling of non-stationary series using OLS may in some cases give misleading inferences. When time series are not cointegrated, OLS results in spurious regression (giving high R^2). Knowledge of whether time series have stochastic trends is therefore an important first step in modeling the long-run relationships between time series. Correspondingly, unit root tests are intended for this purpose.

The unit root tests of the Dickey-Fuller framework (Dickey and Fuller, 1979, 1981) use autoregressive estimating equations. The original Dickey-Fuller unit root test, also called the standard Dickey-Fuller test (Dickey and Fuller, 1979), assumes independently and identically distributed (iid) errors. Most macroeconomic variables have serially correlated errors and do not therefore satisfy this assumption. The standard Dickey-Fuller test has therefore been modified to cope with non-iid errors in two ways:

1. By modifying the estimating equation

   The Augmented Dickey-Fuller (ADF) test (Dickey and Fuller, 1981) uses the following estimating equation,

   $$\Delta y_t = \mu_t + \beta t + \gamma y_{t-1} + \sum_{j=1}^{k} \alpha_j \Delta y_{t-j} + \varepsilon_t$$

   where Δy_t is the first difference, y_{t-1} is the lagged series, t is the time trend and k is the truncation lag parameter (the number of lags of first differences required to remove serial correlation from the residuals). The ADF tests the null hypothesis of a unit root i.e., $H_0 : \gamma = 0$.

   Rejecting the null hypothesis leads to the conclusion that the series is stationary.

   Some of the limitations of the ADF test are that the test is known to suffer from diminished power as a result of augmenting the estimating equation with lagged first differences (see Schwert, 1989; Agiakoglou and Nwebold, 1992, for example). The test also presents low power in distinguishing between trend-stationary and non-stationary drifting time series. In addition to power deterioration, the test suffers from size distortions in the case of heteroscedastic errors in small sample sizes (see Kim and Schmidt, 1993; Cook, 2006). Augmenting the estimating equation of the standard Dickey-Fuller test with lags of the first differences is a parametric approach.

2. By modifying the test statistic

   The unit root test of Phillips (1987), and Phillips and Perron (1988) (PP test) modifies the test statistic instead of the estimating equation. The modification is a non-parametric adjustment, which ensures that the limiting distribution of the test statistic becomes the same as that of the standard Dickey-Fuller test. While the non-parametric modification works well in removing the effects of heteroscedasticity of unspecified form in large samples, this test also suffers from size distortions when used on samples of finite sizes. This is especially the case for time series, which have errors with a Moving-Average (MA) structure, as evidenced in many macroeconomic variables (see Schwert, 1987). MA errors processes can result from additive measurement error contamination, or from testing univariate time series implied by vector autoregressive DGPs (see Cappuccio and Lubian, 2016).

   Both the parametric and non-parametric approaches require the selection of optimal lag lengths. The ADF test requires selection of the optimal augmentation lag, while the PP test requires the selection of the optimal bandwidth for the estimation of the long-run variance. The choice of these tuning parameters also adds to the uncertainty with which models are estimated. An alternative to using long autoregressions is to use unit root tests that are based on the ratio of variances or partial sums (see Breitung, 2002, for example). We look at variance ratio unit root tests next.

1.3 Variance ratio unit root tests

Variance ratio unit root tests are based on the ratio of the variance of the series under the $I(0)$ and $I(1)$ possibilities. The variance of the partial sum of a unit root process increases linearly as a function of time. Consider the case of a unit root process with iid errors and non-stochastic starting value, y_0.

$$y_q = y_0 + \sum_{t=1}^{q} \varepsilon_t$$

then,

$$\sigma^2_{\Delta_1 y_t} = \sigma^2_{\varepsilon_t}$$
$$\text{and}$$
$$\sigma^2_{y_q} = q\sigma_{\varepsilon_t}$$

where Δ_1 is $y_t - y_{t-1}$.

Also,

$$\Delta_q y_t = y_t - y_{t-q} = \sum_{i=0}^{q-1} \Delta_1 y_{t-i}$$

which shows is that the variance of the qth order difference is q times the variance of the first order difference. The variance ratio statistic,

$$\mathrm{VR}(q) = \frac{1}{q}\frac{\mathrm{Var}\left(\Delta_q y_t\right)}{\mathrm{Var}\left(\Delta_1 y_t\right)}$$

can be used as a test statistic to test the null hypothesis that the time series is a random walk with serially uncorrelated errors. Under the null hypothesis, the ratio is equal to 1 while under the alternative hypothesis the series is either a unit root process, has serially correlated errors, or both. The empirical version of the ratio can therefore serve both as a unit root and specification test. The test can easily be generalized to test time series with different trend specifications by detrending the series prior to conducting the unit root.

Lo and MacKinlay (1988) show that the suitably normalized variance ratio test statistic has a standard normal limiting distribution under the null hypothesis. The condition for this is that q is fixed as $T \to \infty$ so that $(q/T) \to 0$.

Examples of variance ratio unit root tests are those suggested by Tanaka (1990) and Kwiatkowski et al. (1992) among others. The test statistic for the variance ratio unit root test given in Kwiatkowski et al. (1992) is,

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

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 a semi-parametric kernel based method e.g., the Newey-West estimator of Newey and West (1987). When testing against trend stationary alternatives, the test statistic for the detrended time series is,

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

where the deterended 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 is used when the null hypothesis specifies stationarity about a non-zero mean. The test statistic of Kwiatkowski et al. (1992) given above tests for stationarity i.e., has its null hypothesis as stationary. Breitung (2002) reverses the roles of the null and alternative hypotheses and proposes using $\hat{\varrho}_T$ 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 because the long-run variance cancels out in the variance ratio. Not having to estimate the long-run variance accords the test an advantage over the PP and other unit root tests that require estimation of long-run variances.

Sargan and Bhargava (1983) generalized the Durbin-Watson test statistic for serial correlation (Durbin and Watson, 1950, 1951) and used it as a unit root test. The test has its basis in whether the residuals from an OLS regression follow a random walk. It uses the ratio of the variance of the residuals of the first difference equation to that of the levels equation. Bhargava (1986) used this testing framework to suggest the test statistic,

$$R_1 = \frac{\sum_{t=2}^{T}(y_t - y_{t-1})^2}{\sum_{t=1}^{T}(y_t - \overline{y})^2}$$
$$\text{with } \overline{y} = \frac{1}{T}\sum_{t=1}^{T} y_t$$

The random walk hypothesis is rejected for large values of the R_1.

A second test statistic discussed in Stock (1999) is the following variance ratio,

$$MSB = \frac{\frac{1}{T^2}\sum_{t=1}^{T} y_{t-1}^2}{s^2}$$

where s^2 is an estimate of the long-run variance, which may be estimated using a kernel estimation method such as that of Newey and West (1987).

All variance ratio unit root tests are motivated by the fact that the variance of the partial sum of an $I(1)$ process is $O_p(T)$ while that of an $I(0)$ process is $O_p(1)$. With suitable normalization, their test statistics take on small values under the null hypothesis.

The variance ratio unit root tests proposed in this thesis are underpinned by the same priciples. The proposed unit root tests are motivated by the fact that the spectrum of a unit root process peaks at the zero frequency, and tails off exponentially (Granger, 1966). As a consequence, the largest proportion of the variance of the process is found in the lowest frequency bands. Suitable unit root test statistics can therefore be based on the relative distribution of the variance of the time series with regards to its frequency content. For this to be feasible, the spectral variance needs to be decomposed on a frequency basis with the use of suitable bandpass filters. The proportions of the total variance contributed by the periodic components corresponding to relevant frequencies can thus be compared. Fan and Gençay (2010) introduced the wavelet variance ratio unit root test based on this principle. We generalize their unit root test to a system of equations, and use the wavelet filter and variance ratio to construct tests with size fidelity in the presence of error distortions.

1.4 Cointegration

When two or more non-stationary time series have a linear combination that is stationary, the time series are said to be cointegrated (Engel and Granger, 1987). Cointegration implies a long-run equilibrium relationship between variables. Typical examples of cointegration include the Permanent Income Hypothesis, which implies cointegration between consumption and income, the Fisher Hypothesis, which implies cointegration between nominal interest rates and inflation, and the Purchasing Power Parity, which implies cointegration between the nominal exchange rate and foreign and domestic prices. The statistical formulation for cointegration is succinctly captured in what follows.

Consider, the bivariate case, where y_t and x_t are both $I(1)$ and ε is $I(0)$.

$$y_t = \beta_0 + \beta_1 x_t + \nu_t \tag{1.1}$$

$$\Delta y_t = \varepsilon_t \tag{1.2}$$

Both y_t and x_t are driven by the common stochastic trend,

$$\sum_{t=1}^{T} \varepsilon_t$$

and β represents the long-run equilibrium, also called the cointegrating coefficient or parameter.

An important property of OLS estimation (which is also carried over to estimators based on its modifications) is consistency of the estimator of the cointegrating parameter. The estimator converges at a rate of T instead of the usual $\sqrt{T}$, which is the case when the time series are stationary. This is easy to see:

The OLS estimator of β is given as,

$$\begin{aligned}
\hat{\beta} &= \sum_{t}^{T} x_t y_t \left(\sum_{t=1}^{T} x_t^2 \right)^{-1} \\
&= \beta + \sum_{t}^{T} x_t \varepsilon_t \left(\sum_{t=1}^{T} x_t^2 \right)^{-1}
\end{aligned}$$

By the assumption of cointegration, x_t is $I(1)$ and ε_t is $I(0)$. The orders of convergence of the terms in the equation are therefore,

$$\begin{aligned}
&T^{-1} \sum_{t=1}^{T} x_t^2 = O_p(T) \\
&T^{-1} \sum_{t=1}^{T} x_t \varepsilon_t = O_p(1) \\
&\text{so that,} \\
&T(\hat{\beta} - \beta) = O_p(1) \\
&\text{and } (\hat{\beta} - \beta) = O_p(T^{-1})
\end{aligned}$$

This important result, which is known as the superconsistency property of the OLS based estimators, is relied upon to eliminate the small sample estimation bias of the cointegrating parameter. However, endogeneity of the regressors in OLS regression is certain when the variables in a conintegrating relationship are jointly determined, and as a result, the regressors are correlated with the error terms (see Davidson and MacKinnon, 1993, pp. 717). In addition to this, contemporaneous correlation between the errors of the different equations, serial correlation of ε_t, as well as possible cross-equation serial correlation, can all lead to long-run edogeneity. All this results in biased estimation of the cointegrating parameter, and t-ratios that are not normally distributed, even asymptotically. An important factor to consider in studying the small sample bias is the ratio of the standard deviations of ε_t to ν_t (see Eqns. 1.1 and 1.2). This ratio ($\sigma_\varepsilon/\sigma_\nu$), called the signal-to-noise ratio, is known to contribute to the small sample estimation bias of the cointegrating parameter (Phillips and Hansen, 1990a). Low signal-to-noise ratios result in larger bias. This thesis also looks to address the issue of low signal-to-noise ratios using wavelet filtering in conjunction with the Fully Modified-OLS method of Phillips and Hansen (1990a); Phillips and Lorentan (1991), the Dynamic-OLS methods of Saikkonen (1991a) and Stock and Watson (1993), and the Integrated Modified-OLS of Vogelsang and Wagner (2014b). All three methods are designed to give asymptotically normal t-ratios so that the normal distribution based inference is valid for cointegrating parameter.

1.5 Error distortions

This thesis is in part concerned with unit root testing under error distortions. The effects of two types of error distortions (conditional heteroscedasticity and additive measurement error) on the size of unit root tests are studied, and unit root tests that are robust to these distortions are also suggested.

1.5.1 Conditional heteroscedasticity

Because time-varying conditional variance is frequently encountered in financial and economic time series, it is of interest to assess the robustness of unit root tests towards errors with non-constant variances. Generalized Autoregressive Conditionally Heteroscedastic (GARCH) processes (Bollerslev, 1986) are characterized by conditional variances that depend on current and past information. For this reason, they provide good models for error processes of time series that exhibit volatility clustering.

The model for the GARCH(p,q) errors is as follows:

$$\begin{aligned} \sigma_t^2 &= \omega + \sum_{i=1}^{q} \alpha_i u_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2 \\ u_t &= \varepsilon_t \sigma_t \\ \varepsilon_t &\sim \text{iid}(0,1) \end{aligned}$$

Where σ_t^2 is the conditional variance (i.e., $\text{Var}(u_t|u_{t-1})$). When $\beta_i = 0$, then the GARCH (p,q) model reduces to the seminal ARCH(q) model of Engle (1982), the model which was later generalized by Bollerslev (1986) to include dependence on past conditional variances. The dependency in GARCH processes is through their second moments. The u_t^2 can therefore be modeled as ARMA processes with AR parameters $\sum_{i=1}^{p}(\alpha_i + \beta_i)$ and MA parameters $-\sum_{i=1}^{p} \beta_i$. The GARCH(1,1) model is the most commonly used model in financial applications because it provides a parsimonious representation of higher order ARCH processes. Parsimony of this model comes from the inclusion of lagged conditional variance to model periods of sustained high or low volatility. Modeling Periods of sustained volatility would otherwise require ARCH models of high order. The first lag autocorrelation function of u_t^2 for the GARCH(1,1) process can be modeled separately from that of subsequent lags, and the decay in the subsequent lags can be flexibly modeled using different choices of $(\alpha + \beta)$ (see Ruppert, 2010, pp. 399–400). This contributes to the versatility and hence popularity of the GARCH(1,1) model.

1.5.2 Additive measurement errors

Although most economic time series are recorded with some degree of measurement error, testing for unit roots is still routinely done using the tests of the Dickey-Fuller framework, which can potentially result in misleading inferences. These tests are known to be over-sized in the presence of error distortions, which can arise from a number of sources. Dickey-Fuller type tests are known to be sensitive to errors that follow an MA specification (see Schwert, 1989, for example). To understand the manifestation of measurement erorrs a unit root process, consider the following:

A random walk process with measurement error can be represented using an unobserved components model as follows:

$$y_t = y_t^* + \varepsilon_t \tag{1.3}$$

$$y_t^* = \rho y_{t-1}^* + u_t \text{ with } \rho = 1 \tag{1.4}$$

where y_t^* is the true but unobserved time series, $\varepsilon_t \sim \text{iid}(0, \sigma_\varepsilon^2)$ is the measurement error and independent of u_{t-s} for all t and s, and u_t is a zero mean white-noise process with variance σ_u^2. The ratio $\sigma_\varepsilon / \sigma_u$ is called the noise-to-signal ratio. The error u_t drives the unit root process due to its permanent effect on y_t, and therefore future values, y_{t+s}. The noise ε_t, however, only affects y_t but not future values, y_{t+s}. The impact of the noise therefore results in the increase of the variance of y_t from $\text{Var}(y_t) = t\sigma_u^2$ to $\text{Var}(y_t) = t\sigma_u^2 + \sigma_\varepsilon^2$. The measurement error impacts the behavior of y_t in the following way:

starting with the first difference,

$$\begin{aligned} y_t - y_{t-1} &= \omega_t, \\ \text{then } \omega_t &= u_t + \varepsilon_t - \varepsilon_{t-1} \end{aligned} \tag{1.5}$$

and

$$\begin{aligned} \mathrm{E}(\omega_t) &= 0 \\ \mathrm{Var}(\omega_t) &= \sigma_u^2 + 2\sigma_\varepsilon^2 \\ \mathrm{Cov}\,(\omega_t, \omega_{t-s}) &= -\sigma_\varepsilon^2 \text{ for } s = 1 \text{ and } 0 \text{ for } s > 1 \end{aligned} \tag{1.6}$$

ω_t is therefore the MA(1) process, $\omega_t = \epsilon_t - \theta\epsilon_{t-1}$, with ϵ_t being a white noise process, and θ and σ_ϵ^2 satisfying

$$(1+\theta^2)\sigma_\epsilon^2 = \sigma_u^2 + 2\sigma_\varepsilon^2 \quad \text{and} \quad \theta = \frac{\sigma_\varepsilon^2}{\sigma_u^2} \tag{1.7}$$

The size of the moving average parameter therefore determines the noise-to-signal ratio in the measurement error contaminated series. This relationship is given as:

$$\frac{\sigma_\varepsilon}{\sigma_u} = \sqrt{\frac{\theta}{(1-\theta)^2}} \tag{1.8}$$

It can be seen that as $\theta \to 1$, the noise-to-signal ratio $\to \infty$ and $\{y_t\}_{t=1}^T$ behaves like white noise, hence the oversizing of many unit root tests. This relationship is useful, because in the part of this thesis that considers the problem, the limiting distribution of the t-ratio, in the presence of measurement errors (which will be shown to depend on nuisance parameters), will be parametrized in terms of θ for mathematical convenience. θ, as can be seen, is a function of the noise-to-signal ratio. The relationship can also be used to estimate the noise-to-signal ratios in $I(1)$ time series which may be of interest in itself, but is not pursued here.

A wavelet variance ratio unit root test that presents good size properties in the presence of additive measurement error is proposed in this thesis as an alternative to the Dickey-Fuller type tests.

1.6 The Wavelet Transform

Much of this thesis pertains to the use of the filters of the Discrete Wavelet Transform (DWT), so it would be appropriate at this point to give a brief and non-technical overview of this transform for potential readers who may not be familiar with wavelets, and to make accessible the application of the transform as a filtering tool, as well as the basis for unit root tests.

1.6.1 The Fourier and Continuous Wavelet Transforms

The analysis of the frequency content of signals (of which time series are a special case) has traditionally been done using Fourier analysis. We therefore use the Fourier transform as the starting point in our overview of wavelet transform.

The spectral representation of a signal can be obtained by multiplying the signal function with an analyzing function (the Continuous Fourier Transform (CFT), for example) and summing the product over the time range of the signal,

$$X(f) = \int_{-\infty}^{\infty} x(t) e^{-2i\pi f t} dt$$

The signal can be recovered by the reverse operation,

$$x(f) = \int_{-\infty}^{\infty} X(t) e^{-2i\pi f t} dt$$

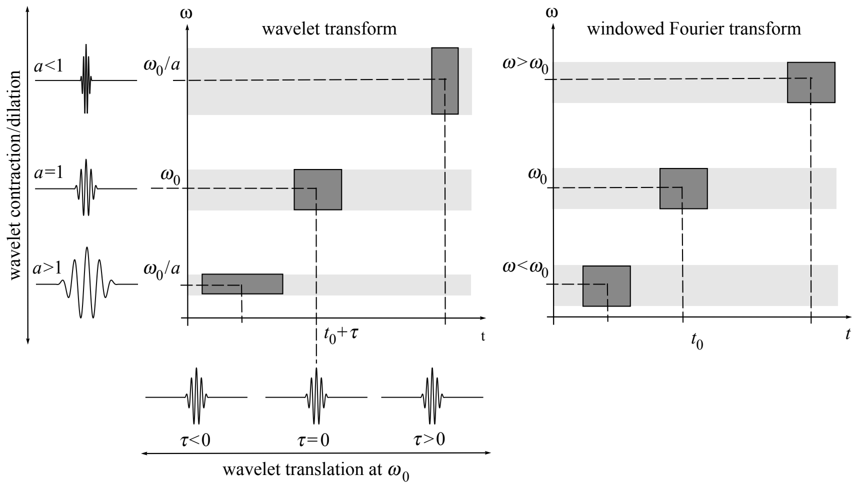
From the given transform, it is apparent that for each frequency, $X(f)$ is a function of $x(t)$ for $t \in (-\infty, \infty)$. As a consequence, no feature of $X(f)$ can be linked to a particular time point. More technically, the function $e^{-2i\pi f t}$, which consists of sinusoids, has an infinite set of points where it equals to zero (infinite support). While the Fourier transform returns the frequency content of the signal, it offers no information on the time where the frequencies occur, and is, therefore, limited in its handling of non-stationary signals – signals which have frequencies that change over time. To localize the frequency content, there is a need to use analyzing functions that have finite support. One way to achieve this is by cutting the signal at equal intervals (windowing). In this way, the spectral representation of the signal can be made to explicitly depend on time. The Short-Time Fourier Transform (STFT) achieves this by multiplying the original signal with windowing functions, which are localized in time about $t = \tau$ (see Eqn 1.9) The spectral representation of $x(t)$ then becomes,

$$X(f, \tau) = \int_{-\infty}^{\infty} x(t) g(t - \tau) e^{-2i\pi f t} dt \tag{1.9}$$

where $g(t - \tau)$ is the windowing function. The Gaussian function $g(t) = exp(-\beta t^2/2)$ is often a candidate windowing function because of its symmetry and smoothness. The ability to resolve features both in time and frequency is called time-frequency resolution. The STFT has a fixed time-frequency resolution because of its constant window size. Windowing results in trade-offs between time and frequency resolution. It is not possible to identify a window of optimal width, which both localizes features and captures the frequencies present in a given time interval. However, some methods give better time-frequency resolution than others.

To overcome the limitations presented by the Fourier transforms, analyzing functions that have compact and flexible support are needed. Wavelets (which are small waves compared to sinusoids), have compact support. They can also be translated and compressed (or dilated) to get the flexibility needed for time-frequency resolution. The localization problem is solved by translating the wavelet, and the frequency resolution is increased by rescaling of the wavelet. Figure 1.1 contrasts the flexible wavelet analyzing function with the fixed window sized STFT. The wavelet function can be translated and scaled to obtain spectral representations that are functions of both frequency (scale) and time.

Figure 1.1: The wavelet and windowed Fourier transforms



Reproduced from Cazelles et al. (2007) with permission. The figure shows how the wavelet function dilates as the scale (a) increases, and how the function can be translated along the time axis (at the angular frequency ω_0 as an example). The windowed Fourier transform (to the right) has a fixed frequency-time resolution.

The Continuous Wavelet Transform is given by the doubly indexed function,

$$\psi_x(s,\tau) = \frac{1}{\sqrt{s}} \int_{-\infty}^{\infty} x(t)\psi\left(\frac{t-\tau}{s}\right) dt$$

where τ is the translation (shifting) factor and s is the scale. Higher scales (lower frequencies) correspond to stretched wavelets, and lower scales correspond to compressed wavelets. The factor $(\sqrt{s}^{-1})$ is required for normalization of the wavelet energy (squared norm of the output coefficients). The normalization is such that the squared norm of the data is equal to the squared norm of the output coefficients, so that variance is preserved.

Important properties of wavelets are that their average value in the time domain is to zero, which implies that, if the function departs from zero, then it must be oscillatory,

$$\int_{-\infty}^{\infty} \psi(t)dt = 0,$$

that the energy[1] is unity,

$$\int_{-\infty}^{\infty} |\psi(t)|^2 dt = 1$$

and that wavelets must satisfy the admissibility condition,

$$\int_{-\infty}^{\infty} \frac{|\Psi(f)|^2}{|f|} df < \infty$$

where $\Psi(f)$ is the Fourier transform of $\psi(t)$. The condition of finite integral is satisfied if $\Psi(0) = 0$. Also, $\Psi(f) \to 0$ as $|f| \to \infty$, which means that wavelets must posses band-pass like properties because they are able to remove periodic components at both the low and high frequencies.

The choice of a wavelet analyzing function depends on several properties of the function. One criterion is the number of vanishing moments. Moments of a function are defined as,

$$m_x = \int_{-\infty}^{\infty} f(x)x^k dx$$

The k^{th} moment vanishes if the integral equals zero.

The implications for the wavelet transform is that, as the number of vanishing moments increase, polynomials of lesser order are no longer identified by the wavelet function i.e., when the wavelet's $(k+1)$ moments are equal to zero, all the polynomial signals of up to order k have zero wavelet coefficients. Wavelets with high numbers of vanishing moments also have longer support and better approximate complicated functions.

[1] The energy of a function is the squared function integrated over its domain (see Gençay et al., 2001, p. 102)

A related property is the regularity of a wavelet. If a function is r-time continuously differentiable at x_0 and r is an integer $(r \geq 0)$, then the regularity of the function is r. The higher the regularity of a wavelet function, the smoother the wavelet. Correspondingly, wavelets of low regularity such as the Haar family of wavelets, give less smooth approximations. Wavelets with high numbers of vanishing moments tend to have higher regularity as well.

We conclude this general section on wavelets by looking at the relationship between scale and frequency. This relationship can only be defined in a broad sense because wavelet functions do not have a single frequency. For that reason, we define a pseudo-frequency and relate it to scale. Let C_f (called the center frequency) be defined as the dominant frequency of the wavelet. The center frequency is the frequency maximizing the Fourier transform of the wavelet function. The wavelet pseudo frequency can then be defined as,

$$F_c = \frac{C_f}{s\delta t}$$

where C_f is the center frequency, s is the scale, and δt is the sampling interval. We see that if the wavelet is dilated by a factor s, then the frequency C_f scales to C_f/s. There is, therefore, an inverse relationship between the scale and frequency for wavelets.

1.6.2 The Haar Wavelet Transform

The Haar CWT was first used by Haar (1909). Because the function meets all the criteria for a wavelet function, it later became known as the Haar wavelet. It is the simplest of all wavelet functions, with the following wavelet and scaling functions:

Wavelet function,

$$\psi(t) = \begin{cases} 1 & 0 \leq t < \frac{1}{2} \\ -1 & \frac{1}{2} \leq t < 1 \\ 0 & \text{otherwise.} \end{cases}$$

and scaling function,

$$\phi(t) = \begin{cases} 1 & 0 \leq t < 1 \\ 0 & \text{otherwise.} \end{cases}$$

The Haar function is defined on $\mathbb{R}$ for every pair of integers $j, k \in Z^+$ as,

$$\psi_{j,k}(t) = 2^{j/2}\psi\left(2^j t - k\right)$$

The function has compact support in the interval $I_{j,k} = \left[k2^{-j}, (k+1)2^{-j}\right]$, has zero average,

$$\int_{-\infty}^{\infty} \psi_{j,k}(t)dt = 0$$

and unit norm

$$\int_{-\infty}^{\infty} \psi_{j,k}(t)^2 dt = 1$$

Haar functions are pairwise orthogonal,

$$\int_{-\infty}^{\infty} \psi_{j,k}(t)\psi_{j',k'}(t)dt = 0 \text{ if } j \neq j' \text{ or } k \neq k'$$

The orthogonality property is easily evident because, either any two supporting intervals are disjoint, or that the wavelet with the smaller support is fully contained within one half of the wavelet with the larger support, which is constant. The narrower wavelet subsequently averages to zero over the constant support. Any continuous function defined on the interval $[0,1]$ can be approximated by linear combinations of $1, \psi(t), \psi(2t), \psi(4t), \ldots, \psi(2^j t)$, and their translations. This combined with the properties given earlier implies that Haar wavelets form orthonormal bases for square-integrable functions in $[0,1]$.

1.6.3 The Discrete Wavelet Transform (DWT)

The Continuous Wavelet Transform (CWT) requires an infinite number of wavelets because of continuously shifting continuously scalable functions. This results in highly redundant wavelet coefficients. In practice, only few coefficients are required to capture most of the variation in a signal, so the redundancy is removed by sampling the signal and using a discrete wavelet transform for analysis. Many time series, including economic time series, are indeed sampled at discrete time points and require a transform that can handle discrete data.

The Haar DWT

To illustrate the working of the filters of the Haar DWT, consider the time series $\{y_t\}_{t=0}^{T}$. The wavelet coefficients at the unit scale are the wighted and adjacent but non-overlapping differences i.e.,

$$W_{1t} = \frac{(y_{2t} - y_{2t-1})}{\sqrt{2}}$$

or in the general form,

$$W_{1t} = \sum_{l=-\infty}^{\infty} h_l y_{2t-l} \tag{1.10}$$

where,

$$h_l = \begin{cases} 2^{-1/2} \text{ for } l = 0 \\ -2^{-1/2} \text{ for } l = 1 \\ 0 \quad \text{otherwise.} \end{cases}$$

Similarly, the scaling coefficients are given as,

$$V_{1t} = \frac{(y_{2t} + y_{2t-1})}{\sqrt{2}}$$

and in the general form,

$$V_{1t} = \sum_{l=-\infty}^{\infty} g_l y_{2t-l} \tag{1.11}$$

where,

$$g_l = \begin{cases} 2^{-1/2} \text{ for } l = 0 \\ 2^{-1/2} \text{ for } l = 1 \\ 0 \quad \text{otherwise.} \end{cases}$$

Equations 1.10 and1.11 represent filtering operations with filter coefficients $\{h_l\}_{-\infty}^{\infty}$ and $\{g_l\}_{-\infty}^{\infty}$ respectively. From the properties of wavelets discussed earlier,

$$h_0 + h_1 = 0 \text{ and } h_0^2 + h_1^2 = 1$$

also,

$$\sum_{l=0}^{2} h_l h_{l+2n} = 0$$

for all non-zero integers n.

The Haar DWT offers variance preservation (the variance of its coefficients are equal to the variance of the time series), and parsimony i.e. the most important features are captured in just a few coefficients. However, this property, which makes the wavelet ideal in applications such as image compression for example, becomes a disadvantage when the wavelet is used for time series analysis. The requirement of mandatory dyadic sample sizes limits its applications in time series analysis. For this reason, we use the Haar Maximal Overlap DWT for time series analysis in this thesis.

1.6.4 The Haar Maximal Overlap DWT (MODWT)

The MODWT has several advantages over the DWT. It can handle samples of any size, and is invariant to circular shifting of the original time series. An important property, which is useful for the work covered in this thesis, is that the MODWT wavelet variance estimator is superior its DWT counterpart (see Percival, 1985). We will use the wavelet variance estimator in estimating the variance contribution made by the (stationary) periodic components of high frequencies, in the construction of variance ratio unit root tests. Because the transform is non-decimated, it produces as many coefficients as the input data, and therefore captures local variation better, although at the expense of some redundancy.

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

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

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

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

The Haar MODWT wavelet filter $\tilde{h}_{j,l}$ therefore approximates a band-pass filter with the nominal pass band $[2^{-(j+1)}, 2^{-j}]$ and the its scaling filter $\tilde{g}_{j,l}$ approximates an ideal low-pass filter with the nominal pass band $[0, 2^{-(j+1)}]$. The reason for only considering frequencies in the range $[0, 1/2]$ is as follows:

An important characteristic of any linear filter is its frequency response function. The squared gain function is the square of the modulus of the response function, and provides information about frequencies that are passed and frequencies that are attenuated by the filter. The squared gain function is an even and symmetric function of frequency, and has period equal to unity. It is therefore sufficient to consider only the frequency range $[0, 1/2]$ when evaluating the band-pass filters.

The jth level wavelet and scaling coefficients as defined as follows:

$$\tilde{W}_{j,t} \equiv \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} y_{t-l}, \quad (t = 0, 1, 2, 3 \ldots)$$
$$\tilde{V}_{j,t} \equiv \sum_{l=0}^{L_j-1} \tilde{g}_{j,l} y_{t-l}, \quad (t = 0, 1, 2, 3 \ldots)$$

Summation of the coefficients gives,

$$\sum_{j=1}^{J_0} \tilde{W}_{j,t} + \tilde{V}_{J_0,t} = \sum_{l=0}^{L_j-1} \left(\sum_{j=1}^{J_0} \tilde{h}_{j,l} + \tilde{g}_{J_0,l} \right) y_{t-l}$$

where J_0 is a scale less than or equal the maximum resolution for the data. For the Haar MODWT wavelets, we have by the definition of its filters,

$$\sum_{j=1}^{J_0} \tilde{h}_{j,l} + \tilde{g}_{J_0,l} \equiv \begin{cases} 1 & \text{for } l = 0, \\ 0 & \text{for } l \neq 0 \end{cases}$$

and as a result of this,

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

The time series can therefore be additively decomposed into the wavelet and scaling coefficients of the Haar MODWT. The significance of this is that the dynamics of the periodic components at any frequency band can be removed (filtered) by simply discarding the wavelet coefficients corresponding to that frequency band. We make use of this property in constructing the wavelet variance ratio unit root tests.

The motivations for using the Haar MODWT function in this thesis are several fold – the Haar MODWT wavelet has the simplest orthonormal wavelet basis, is conceptually simple and reversible, without the boundary effects characteristic of longer and smoother wavelets. It is also computationally fast and lends itself to use in extensive simulations as is the case with the work covered in this thesis. The wavelet has a superior wavelet variance estimator to the DWT, can handle samples of any size, and data can additively be decomposed into its scaling and wavelet coefficients. The wavelet function also decomposes the total variance of a time series on a scale-by-scale basis. Because the focus of this thesis is in evaluating the performance of unit root tests in time series of small sample sizes, the Haar MODWT has advantages over filters with longer support for this purpose.

1.7 Critical values, test size and power

Critical values, empirical test sizes, and power functions for the different statistical tests that are presented in this thesis are obtained using Monte Carlo simulation. When the actual sizes of test statistics are different from their nominal sizes, direct comparisons of power estimates gives misleading results. Over-sized tests show higher empirical power compared to correctly sized (or under-sized) tests when the power comparisons are not adjusted for differences in the test sizes. Below, we show how the empirical test sizes, critical values and size-adjusted power functions are obtained in this thesis.

Most unit root tests reject a one-sided null hypothesis for small (negative) values of the test statistic. For this reason, we illustrate the calculation of the empirical test size and power using the one-sided critical value, $-c$.

For a test statistic $T(\boldsymbol{y})$ with a critical value $-c$ based on nominal size

α, the Monte Carlo procedure uses n_0 independent samples under the null hypothesis and n_1 samples under the alternative hypothesis. The estimated size of the test is then given by,

$$\hat{p}_0 = \frac{1}{n_0} \sum_{i=1}^{n_0} I(T(\boldsymbol{y})_{0i} \leq -c)$$

$T(\boldsymbol{y})_{0i}, i = 1, \ldots, n_0$ are the values of the test statistic for each sample under the null hypothesis, and $I(\cdot)$ is the indicator function for a rejection of the null hypothesis at each iteration. $I(T(\boldsymbol{y})_{0i})$ equals 1 if $T(\boldsymbol{y})_{0i} \leq -c$ and 0 otherwise. Confidence limits can be set for p_0 based on the proportions of rejections of the null hypothesis (see Edgerton and Shukur, 1999, for example).

The empirical power under the alternative is given analogously,

$$\hat{p}_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} I(T(\boldsymbol{y})_{1i} \leq -c)$$

When making power comparisons for tests that have different sizes, an adjustment has to be made to ensure that the tests have the same size, α. The adjustment requires new critical values to be generated using Monte Carlo simulations. The critical values are based on the α quantile of the empirical distribution under the null hypothesis i.e., α quantile of the empirical distribution of $T(\boldsymbol{y})_{0i}$. For a finite population of $T(\boldsymbol{y})_{0i}$ with equally probable values, and indexed $T(\boldsymbol{y})_{01} \ldots T(\boldsymbol{y})_{0n}$ from lowest to highest, the Monte Carlo critical value, $\hat{c}$, is the value $T(\boldsymbol{y})_{0i}$, which corresponds to the αn_0. The size-adjusted power is therefore given as,

$$\hat{p}_1^* = \frac{1}{n_1} \sum_{i=1}^{n_1} I(T(\boldsymbol{y})_{1i} \leq -\hat{c})$$

In this thesis $n_0 = 5n_1$ (with $n_1 = 10\ 000$) has been used. This is done to get the low valeus of $\mathrm{Var}(p_1^*)$ using a reasonable number of iterations. Ratios of up to $n_0 = 10n_1$ have been suggested in the literature (see Zhang and Boos, 1994).

Chapter 2

Outline of papers

2.1 Paper I

In this paper, a wavelet variance ratio unit root test is proposed for time series with GARCH(1,1) error processes. The unit root test is applied to time series that have been filtered from their high frequency periodic components using the Maximal Overlap Discrete Wavelet Transform (MODWT) filter. Both the filtering and the unit root test are based on the same MODWT analyzing function, which allows the filtering and unit root testing to be performed in one step. The empirical test sizes and size-adjusted power functions of the proposed unit root test are compared to four other unit root tests – namely, the Augmented Dickey Fuller test (ADF), the Phillips and Perron (PP) test, the wavelet variance ratio test of Fan and Gençay (2010) ($\widetilde{\mathrm{FG}}_1$), and the modified Dickey-Fuller test of Li and Shukur (2011). The PP and Li and Shukur's tests are designed to cope with heteroscedasticity, while the ADF test is used as a benchmark because of its accessibility and extensive use. $\widetilde{\mathrm{FG}}_1$ is included in the comparison to assess whether meaningful gains are made from filtering the time series from it highest frequency periodic components. The same basic simulation design as used by Sjölander (2008) is used in this paper. The Design consists of fifteen GARCH(1,1) error processes covering the range of Data Generating Processes (DGPs) likely to be found in empirical work.

The proposed unit root test is shown to have better performance compared to the four alternative tests and sample sizes considered. All the tests are biased for the DGPs that have very strong volatility, but the proposed unit root test is the least over-sized even in this case. The two wavelet variance ratio tests have higher size-adjusted power compared to the three DF-type tests (ADF, PP and Li and Shukur's tests). $(\widetilde{\mathrm{FG}}_1)$, used when the conditional heteroscedasticity is neglected, has only marginally higher size-adjusted power than the proposed unit root test. The two tests have identical power in the interesting near unit root region $(\rho \geq 0.95)$ for all sample sizes and DGPs.

The proposed unit root test has therefore the best size properties and matches the test with the highest size-adjusted power for DGPs that are near unit root. The limiting distribution of the proposed test statistic is also derived for completeness.

2.2 Paper II

Although many economic time series are recorded with some degree of measurement error, testing for unit roots is still routinely done using the Dickey-Fuller and related tests, which can potentially result in misleading inferences. In this paper, we examine the effect of measurement errors on the commonly used Dickey-Fuller unit root test (Dickey and Fuller, 1979). We show that the Dickey-Fuller test is severely over-sized in the presence of measurement errors, both asymptotically via the limiting distribution of its t-statistic, and in finite sample sizes using Monte Carlo simulation. As a remedy to this problem we propose a modification to the wavelet variance ratio unit root test introduced by Fan and Gençay (2010). We obtain the limiting distribution of the test statistic and show that the test has superior size and size-adjusted power compared to the Dickey-Fuller test for finite sample sizes, and better size compared to Fan and Gençay's variance ratio test. In addition, an empirical application using the new test on inflation rates of four Asian countries is used to demonstrate its robustness to the possible presence of measurement errors.

2.3 Paper III

In this simulation based paper, we suggest a simple way of reducing the small sample bias that arises in the estimation of the conintegrating parameters of bivariate models. We use the Maximal Overlap Discrete Wavelet Transform (MODWT) wavelet filters to improve the signal-to-noise ratios (ratio of the standard deviation of the errors that drive the unit root process to that of the standard deviation of the additive noise process) thereby reducing the small sample bias resulting from edogeneity of the regressors. The wavelet filters are used in conjunction with Ordinary Least Squares, the Fully Modified Ordinary Least Squares of Phillips and Hansen (1990b); Dynamic Ordinary Least Squares (Saikkonen (1991b), Phillips and Lorentan (1991) and Stock and Watson (1993)); and the Integrated Modified Ordinary Least Squares of Vogelsang and Wagner (2014a), to provide filtered versions of these methods. Monte Carlo simulation shows that wavelet filtering results in estimators with smaller bias for all the methods considered, and is most effective for the smaller sample sizes of $(T = 50, 75)$.

The results from using Fully Modified Ordinary Least Squares on wavelet filtered data are also compared to those obtained by first filtering the data using the HP filter of Hodrick and Prescott (1980) as done in Li et al. (1995). Monte Carlo simulations show that the wavelet filter performed better in DGPs with both contemporaneous and serial correlation.

2.4 Paper IV

In this paper, we suggest a unit root test for systems 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 wavelets-based method has good size fidelity in the presence of cross-equation dependence. The wavelet-based test is also more powerful than the Cross-sectionally Augmented Dickey-Fuller unit root test of Pesaran (2007) for time series with between 20 and 100 observations, using systems of 5 and 10 equations. We also demonstrate the usefulness of the test through an application on evaluating the Purchasing Power Parity theory for the G7 countries and find support for the theory, whereas the test by Pesaran (2007) finds no such support.

Bibliography

Agiakoglou, C. and Nwebold, P. (1992). Empirical evidence on dickey-fuller type tests, *Journal of Time Series Analysis* **13**: 471–483.

Bhargava, A. (1986). On the theory of testing for unit roots in observed time series, *Review of Economic Studies* **53**(3): 369–384.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity, *Journal of econometrics* **31**(3): 307–327.

Breitung, J. (2002). Nonparametric tests for unit roots and cointegration, *Journal of econometrics* **108**(2): 343–363.

Cappuccio, N. and Lubian, D. (2016). Unit root tests: The role of the univariate models implied by multivariate time series, *Econometrics* **4**(21): 1–11.

Cazelles, B. et al. (2007). Time-dependent spectral analysis of epidemiological time-series with wavelets, *J. R. Soc. Interface* **4**(2): 626–636.

Cook, S. (2006). The robustness of modified unit root tests in the presence of GARCH, *Quantitative Finance* **6**(4): 359–363.

Davidson, R. and MacKinnon, J. (1993). *Estimation and Inference in Econometrics*, Oxford University Press.

Dickey, D. A. and 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.

Dickey, D. A. and Fuller, W. A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root, *Econometrica: Journal of the Econometric Society* **49**: 1057–1072.

Durbin, J. and Watson, G. (1950). Testing for serial correlation in least squares regression, *Biometrika* **37**: 409–428.

Durbin, J. and Watson, G. (1951). Testing for serial correlation in least squares regression, *Biometrika* **38**: 159–178.

Edgerton, D. and Shukur, G. (1999). Testing autocorrelation in a system perspective, *Econometric Reviews* **18**(4): 343–386.

Engel, R. and Granger, C. (1987). Co-integration and error correction: Representation, estimation, and testing, *Econometrica* **55**(2): 251–276.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation, *Econometrica: Journal of the Econometric Society* **50**(4): 987–1007.

Fan, Y. and Gençay, R. (2010). Unit root tests with wavelets, *Econometric Theory* **26**(05): 1305–1331.

Gençay, R., Selçuk, F. and Whitcher, B. J. (2001). *An introduction to wavelets and other filtering methods in finance and economics*, Academic press.

Granger, C. (1966). The typical spectral shape of an economic variable, *Econometrica* **34**(1): 150–161.

Haar, A. (1909). Zur theorie der orthogonalen funktionensysteme, *Georg-August Universität Göttingen* .

Hodrick, R. J. and Prescott, E. C. (1980). Postwar US Business Cycles: An Empirical Investigation, *Carnegie Mellon University discussion paper, no. 451* .

Kim, K. and Schmidt, P. (1993). Unit root tests with conditional heteroskedasticity, *Journal of Econometrics* **59**(3): 287–300.

Kwiatkowski, D., Phillips, P. C., Schmidt, P. and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?, *Journal of econometrics* **54**(1-3): 159–178.

Li, Y., Maddala, G. and Rush, M. (1995). New small sample estimators for cointegration regression: Low-pass spectral filter method, *Economics Letters* **47**(2): 123–129.

Li, Y. and Shukur, G. (2011). Wavelet improvement of the over-rejection of unit root test under garch errors: An application to swedish immigration data, *Communications in Statistics-Theory and Methods* **40**(13): 2385–2396.

Lo, A. W. and MacKinlay, C. (1988). Stock market prices do not follow a random walk: Evidence from a simple specification test, *The Rivew of Financial Studies* **1**(1): 41–66.

Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelationconsistent covariance matrix, *Econometrica* **55**: 703–708.

Percival, D. (1985). On estimation of the wavelet variance, *Biometrika* **82**(3): 619–631.

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.

Phillips, C. B. and Hansen, B. E. (1990a). Statistical inference in instrumental variables regression with i(1) processes, *Review of Economic Studies* **57**(1): 99–125.

Phillips, C. B. and Lorentan, M. (1991). Estimating long-run economic equilibria, *Review of Economic Studies* **58**(3): 407–436.

Phillips, P. C. B. (1987). Time series regression with a unit root, *Econometrica* **55**(2): 277–301.

Phillips, P. C. and Hansen, B. E. (1990b). Statistical Inference in Instrumental Variables Regression With I(1) Processes, *The Review of Economic Studies* **57**(1): 99–125.

Phillips, P. C. and Perron, P. (1988). Testing for a unit root in time series regression, *Biometrika* **75**(2): 335–346.

Ruppert, D. (2010). *Statistics and Data Analysis for Financial Engineering*, Springer.

Saikkonen, P. (1991a). Asymptotically efficient estimation of cointegration regressions, *Econometric Theory* **7**(1): 1–21.

Saikkonen, P. (1991b). Asymptotically Efficient Estimation of Cointegration Regressions, *Econometric Theory* **7**(1): 1–21.

Sargan, J. and Bhargava, A. (1983). Testing residuals from least squares regression for being generated by the gaussian random walk, *Econometrica* **51**(1): 153–174.

Schwert, G. W. (1987). Effects of model specification on tests for unit roots in macroeconomic data, *Journal of Monetary Economics* **20**(1): 73–103.

Schwert, G. W. (1989). Tests for unit roots: A monte carlo investigation, *Journal of Business & Economic Statistics* **7**(2): 147–159.

Sjölander, P. (2008). A new test for simultaneous estimation of unit roots and GARCH risk in the presence of stationary conditional heteroscedasticity disturbances, *Applied Financial Economics* **18**(7): 527–558.

Stock, J. (1999). A class of tests for integration and cointegration. In: Engel and White. Causality and Forcasting. A Festchrift for Clive W. J. Granger, pp. 135–157.

Stock and Watson (1993). A simple estimator of cointegrating vectors in higher order integrated systems, *Economertica* **61**: 783–820.

Tanaka, K. (1990). Testing for a moving average unit root, *Econometric Theory* **6**(04): 433–444.

Vogelsang, T. J. and Wagner, M. (2014a). Integrated Modified OLS Estimation and Fixed-b Inference for Cointegrating Regressions, *Journal of Econometrics* **178**(2): 741–760.

Vogelsang, T. and Wagner, M. (2014b). Integrated modified ols estimation and fixed-b inference for cointegrating regressions, *Journal of Econometrics* **178**(2): 741–760.

Zhang, J. and Boos, D. (1994). Adjusted power estimates in monte carlo experiments, *Communications in Statistics - Simulation and Computation* **23**(1): 165–173.