

# Serial Correlation Tests with Wavelets

Ramazan Gençay\*

April 2010, This version: November 2011

## Abstract

This paper offers two new statistical tests for serial correlation with better power properties. The first test is concerned with wavelet-based portmanteau tests of serial correlation. The second test extends the wavelet-based tests to the residuals of a linear regression model.

The wavelet approach is appealing, since it is based on the different behavior of the spectra of a white noise process and that of a weakly stationary process. By decomposing the variance (energy) of the underlying process into the variance of its low frequency components and that of its high frequency components via wavelet transformation, we design tests of no serial correlation against weakly stationary alternatives. The main premise is that ratio of the high frequency variance to that of the overall variance of a white noise process is centered at  $1/2$  whereas the relative variance of a weakly stationary process is bounded in  $(0, 1)$ . The limiting null distribution of our test is  $N(0,1)$ . We demonstrate the size and power properties of our tests through Monte Carlo simulations. Our results are unifying in the sense that Durbin-Watson  $d$  test is a special case of a wavelet-based test.

Keywords: Independence, serial correlation, discrete wavelet transformation, maximum overlap wavelet transformation, variance ratio test, variance decomposition.

JEL No: C1, C2, C12, C22, F31, G0, G1.

---

\*Department of Economics, Simon Fraser University, 8888 University Drive, Burnaby, British Columbia, V5A 1S6, Canada. Ramo Gençay is grateful to the Natural Sciences and Engineering Research Council of Canada and the Social Sciences and Humanities Research Council of Canada for research support. Email: [rgencay@sfu.ca](mailto:rgencay@sfu.ca)

# 1 Introduction

Testing for serial correlation has been long regarded as an important issue for modeling, inference, and prediction.<sup>1</sup> From the perspective of a stochastic process, testing for the absence of dynamic dependencies is often important for modeling and forecast evaluation. In a regression framework, the presence of serial correlation leads to the inconsistency of the ordinary least squares estimators when the regressors contain lagged dependent variables.

One of the well-known portmanteau tests for serial correlation in econometrics is the so-called Box and Pierce (1970) (BP) test. Ljung and Box (1978) is a modified version of the BP, improving its small sample properties. Extension of the BP test to a general class of dependent processes, including non-martingale difference sequences, is proposed by Lobato *et al.* (2002). Escanciano and Lobato (2009) introduce a data-driven BP test to overcome the choice of number of autocorrelations. The main limitations of these tests is that large samples are required to provide a reasonable approximation to the asymptotic distribution of the test statistics when the null is true. Hence, our first objective is to improve the small sample performance of portmanteau tests of serial correlation.

A natural extension of the portmanteau framework is through the residuals of a regression model. In the linear regression setting, the most well-known test for serial correlation is the  $d$ -test of Durbin and Watson (1950, 1951, 1971).<sup>2</sup> The  $d$  test has several limitations however. In particular, it uses a first-order autocorrelation as the alternative model and cannot be used when the regressors include lagged values of the dependent variable. Moreover, its usage is subject to standard tables that are cumbersome and often lead to inconclusive results, and are designed for a one-sided test against positive serial correlation when often negative serial correlation is of interest as well. The inability to carry out a two-sided test is a serious limitation that we will address in our tests. Alternative tests proposed by Breusch (1978) and Godfrey (1978) are based on the Lagrange multiplier principle, but although they allow for higher order serial correlation and lagged dependent variables, their finite sample performance can be poor.

Unlike time-domain tests, spectral tests may offer attractive frequency localization features with potential small sample improvements. Hong (1996) uses the kernel estimator of the spectral density for testing serial correlation of arbitrary form. His procedure relies on a distance measure between two spectral densities of the data and the one under the null hypothesis of no serial correlation. Paparoditis (2000) proposes a test statistic based on the distance between a kernel estimator of the ratio between the true and the hypothesized spectral density and the expected value of the estimator under the null. However, estimation methods, like the kernel method, cannot easily detect spatially varying local features, such as jumps. Hence, it is important to design test procedures with the

---

<sup>1</sup>See for instance Yule (1926).

<sup>2</sup>There are several important papers in this area, such as Robinson (1991) which allows for the presence of conditional heteroskedasticity and long memory alternatives. Andrews and Ploberger (1996) facilitate the testing for white noise against ARMA(1,1) alternatives and Godfrey (2007) enables comparisons of the Lagrange multiplier tests with bootstrap-based tests. This list is by no means exhaustive.

ability to have high power against such alternatives. This paper aims to accomplish this goal.

Wavelet methods are particularly suitable in such situations where the data has jumps, kinks, seasonality and nonstationary features. The framework established by Lee and Hong (2001) is a wavelet-based test for serial correlation of unknown form that effectively takes into account local features, such as peaks and spikes in a spectral density.<sup>3</sup> Duchesne (2006) extends the Lee and Hong (2001) framework to a multivariate time series setting. Hong and Kao (2004) extend the wavelet spectral framework to the panel regression. The simulation results of Lee and Hong (2001) and Duchesne (2006) indicate size over-rejections and modest power in small samples. Reliance on the estimation of the nonparametric spectral density together with the choice of the smoothing/resolution parameter intimately affects their small sample performance. Recently, Duchesne *et al.* (2010) have made use of wavelet shrinkage (noise suppression) estimators to alleviate the sensitivity of the wavelet spectral tests to the choice of the resolution parameter. This framework requires a data-driven threshold choice and the empirical size may remain relatively far from the nominal size. Therefore, although a shrinkage framework provides some refinement, the reliance on the estimation of the nonparametric spectral density slows down the rate of convergence of the wavelet-based tests, and consequently leads to poor small sample performance.

Our approach builds on the wavelet methodology, but is directly based on the variance-ratio principle, rather than the estimation of the spectral density, often associated with poor small sample performance. By decomposing the variance (energy) of the underlying process into the variance of its low and high frequency components via wavelet transformation, we propose to design variance-ratio type serial correlation tests that have substantial power relative to existing tests.<sup>4</sup>

The *originality* of our approach resides in the fact that we directly utilize the wavelet coefficients of the observed time series to construct the wavelet-based test statistics in the spirit of Von Neumann variance ratio tests. Since the proposed test statistic is not based on a quadratic norm of the distance between empirical spectral density and the spectral density under the null hypothesis, a nonparametric spectral density estimator is not needed and the rate of convergence issues relating to the nonparametric spectral density are not of first order of importance. In addition, this framework is *important* and *innovative*, as the design we propose leads to serial correlation tests with desirable empirical size and power in small samples and does not suffer from the aforementioned small sample limitations of the existing tests. Because the construction of the tests is based on the additive decomposition of the wavelet and scaling coefficients, the ratio of the sum of the squared wavelet to scaling coefficients converges to the normal distribution at the parametric rate under the null hypothesis. Equally importantly, the proposed tests are easy to implement as their asymptotic null

---

<sup>3</sup>Such features can arise from the strong autocorrelation or seasonal or business cycle periodicities in economic and financial time series.

<sup>4</sup>Recently, Fan and Gençay (2010) propose a unified wavelet spectral approach to unit root testing by providing a spectral interpretation of existing Von Neumann unit root tests. Xue and Gençay (2010) propose wavelet-based jump tests to detect jump arrival times in high frequency financial time series data. These wavelet-based unit root, cointegration and jump tests have desirable empirical size and higher power relative to the existing tests.

distributions are nuisance parameter free.

In Section 2, we illustrate our tests and generalize it to the linear regression setting subsequently. In Section 4, we present the Monte Carlo simulations. Conclusions follow afterwards.

## 2 Portmanteau Tests

Let  $\{y_t\}_{t=1}^T$  be a univariate weakly stationary time series process with  $E(y_t) = \mu = 0$ ,  $Var(y_t) = \sigma^2$ ,  $Cov(y_t, y_{t-j}) = E[(y_t - \mu)(y_{t-j} - \mu)] = \gamma_j$  for all  $j \geq 0$ . The  $j$ th order autocorrelation is  $\rho_j = \gamma_j/\gamma_0$ . We consider tests for  $H_0 : \rho_j = 0$  for all  $j \geq 1$  against  $H_1 : \rho_j \neq 0$  and  $0 < |\rho_j| < 1$  for some  $j \geq 1$ . Our starting point will be the unit-scale discrete wavelet transformation (DWT) with Haar filter.<sup>5</sup> This will be a good benchmark to compare against with Daubechies (1992) compactly supported wavelet filters. A further extension will be the unit-scale maximum overlap discrete wavelet transformation (MODWT) to gain further efficiency. We will show that the test based on MODWT with the Haar filter resembles the Durbin-Watson  $d$  test in the linear regression setting.

### 2.1 DWT - Haar Filter Case

Let  $\{h_l\} = (h_0, \dots, h_{L-1})$  be a finite length discrete wavelet filter such that it integrates (sums) to zero  $\sum_{l=0}^{L-1} h_l = 0$  and has unit variance  $\sum_{l=0}^{L-1} h_l^2 = 1$ . In addition, the wavelet (or high-pass) filter  $h_l$  is orthogonal to its even shifts; that is,  $\sum_{l=0}^{L-1} h_l h_{l+2n} = 0$  for all nonzero integers  $n$ . For all the wavelets considered here, the scaling (low-pass) filter coefficients are determined by the *quadrature mirror relationship*

$$g_l = (-1)^{l+1} h_{L-1-l} \quad \text{for } l = 0, \dots, L-1. \quad (1)$$

The inverse relationship is given by  $h_l = (-1)^l g_{L-1-l}$ . The scaling filter coefficients integrates (sums) to  $\sum_{l=0}^{L-1} g_l = \sqrt{2}$  and has unit variance  $\sum_{l=0}^{L-1} g_l^2 = 1$ , orthogonal to its even shifts; that is,  $\sum_{l=0}^{L-1} g_l g_{l+2n} = 0$  for all nonzero integers  $n$ .<sup>6</sup>

Consider the unit scale Haar DWT,  $\{h_l\}_0^1 = (h_0 = 1/\sqrt{2}, h_1 = -1/\sqrt{2})$  of  $\{y_t\}_{t=1}^T$  where  $T$  is assumed to be even.<sup>7</sup> The wavelet and scaling coefficients are given by

$$W_{t,1} = \frac{1}{\sqrt{2}}(y_{2t} - y_{2t-1}), \quad t = 1, 2, \dots, T/2, \quad (2)$$

$$V_{t,1} = \frac{1}{\sqrt{2}}(y_{2t} + y_{2t-1}), \quad t = 1, 2, \dots, T/2. \quad (3)$$

The wavelet coefficients  $\{W_{t,1}\}$  capture the behavior of  $\{y_t\}$  in the high frequency band  $(1/2, 1)$ , while the scaling coefficients  $\{V_{t,1}\}$  capture the behavior of  $\{y_t\}$  in the low frequency band  $(0, 1/2)$ .

<sup>5</sup>A technical introduction to wavelet transformations is presented in Appendix A1.

<sup>6</sup>Note that  $\sum_{l=0}^{L-1} g_l h_{l+2n} = 0$  for all integers  $n$ , see Percival and Walden (2000), page 77.

<sup>7</sup>This assumption can easily be relaxed under several boundary treatment conditions.

The total variance of  $\{y_t\}_{t=1}^T$  is given by the sum of the variances of  $\{W_{t,1}\}$  and  $\{V_{t,1}\}$ . Since for a white noise process, the variance of the scaling coefficients  $\{V_{t,1}\}$  and the wavelet coefficients  $\{W_{t,1}\}$  are distributed evenly, the following test statistic is proposed:

$$\widehat{G}_{T,1} = \frac{\sum_{t=1}^{T/2} W_{t,1}^2}{\sum_{t=1}^{T/2} V_{t,1}^2 + \sum_{t=1}^{T/2} W_{t,1}^2}. \quad (4)$$

Heuristically,  $\widehat{G}_{T,1}$  should be close to  $1/2$  under  $H_0$ , since the numerator is half of the denominator, while under  $H_1$ ,  $0 < \widehat{G}_{T,1} < 1$ . These statements are formalized in the following lemma.

**Lemma 2.1** *Under  $H_0$ ,  $\widehat{G}_{T,1} = \frac{1}{2} + o_p(1)$ , while under  $H_1$ ,  $\widehat{G}_{T,1} = \frac{E(y_{2t} - y_{2t-1})^2}{E(y_{2t} + y_{2t-1})^2 + E(y_{2t} - y_{2t-1})^2} + o_p(1)$ .*

Equations (2) and (3) imply:

$$W_{t,1}^2 = \frac{1}{2}(y_{2t}^2 + y_{2t-1}^2 - 2y_{2t}y_{2t-1}) \text{ and } V_{t,1}^2 = \frac{1}{2}(y_{2t}^2 + y_{2t-1}^2 + 2y_{2t}y_{2t-1}) \quad (5)$$

Using Equation (5), together with Equation (4), we obtain the following under  $H_0$

$$\widehat{G}_{T,1} = \frac{\sum_{t=1}^{T/2} W_{t,1}^2}{\sum_{t=1}^{T/2} V_{t,1}^2 + \sum_{t=1}^{T/2} W_{t,1}^2} \quad (6)$$

$$= \frac{\frac{1}{2} \sum_{t=1}^{T/2} (y_{2t}^2 + y_{2t-1}^2) - \sum_{t=1}^{T/2} y_{2t}y_{2t-1}}{\sum_{t=1}^{T/2} (y_{2t}^2 + y_{2t-1}^2)} \quad (7)$$

$$= \frac{1}{2} - \frac{\sum_{t=1}^{T/2} y_{2t}y_{2t-1}}{\sum_{t=1}^{T/2} (y_{2t}^2 + y_{2t-1}^2)} = \frac{1}{2} - \frac{o_p(T)}{O_p(T)} = \frac{1}{2} + o_p(1) \quad (8)$$

Note that:

$$0 < \frac{E(y_{2t} - y_{2t-1})^2}{E(y_{2t} + y_{2t-1})^2 + E(y_{2t} - y_{2t-1})^2} = \frac{E(W_{t,1}^2)}{E(V_{t,1}^2) + E(W_{t,1}^2)} < 1.$$

We conclude that it is the relative magnitude of the variance of the wavelet coefficients to that of the scaling coefficients that determines the power of the test based on  $\widehat{G}_{T,1}$  and we expect test based on  $\widehat{G}_{T,1}$  to have substantive power against  $H_1$ . Under  $H_1$ , the ratio approaches to zero for a stationary long memory process and approaches to one for a short-memory process. The asymptotic null distribution of  $\widehat{G}_{T,1}$  under  $H_0$  is summarized in the following theorem.

**Theorem 2.2** *Under  $H_0$ ,  $\sqrt{2T}(\widehat{G}_{T,1} - 1/2) \implies N(0, 1)$  where  $N(0, 1)$  is the standard normal distribution.*

**Proof:** Noting that

$$\widehat{G}_{T,1} - \frac{1}{2} = -\frac{\sum_{t=1}^{T/2} y_{2t}y_{2t-1}}{\sum_{t=1}^{T/2} (y_{2t}^2 + y_{2t-1}^2)} \quad (9)$$

$$= -\frac{N(0, \sigma^4 T/2)}{2\sigma^2 T/2} + o_p(1) = -\frac{\sigma^2 (T/2)^{1/2} N(0, 1)}{2\sigma^2 T/2} + o_p(1) \quad (10)$$

$$= -\frac{N(0, 1)}{\sqrt{2T}} + o_p(1) \quad (11)$$

$$\sqrt{2T} (\widehat{G}_{T,1} - \frac{1}{2}) = N(0, 1) + o_p(1) \text{ under the } H_0. \quad (12)$$

because of the symmetry of the normal distribution around its mean.

## 2.2 DWT - Length-2 Filter Case

Consider the unit scale  $\{h_l\}_0^1 = (h_0, h_1)$  filter DWT of  $\{y_t\}_{t=1}^T$  where  $T$  is assumed to be even.<sup>8</sup> The wavelet and scaling coefficients are given by

$$W_{t,1} = h_0 y_{2t} + h_1 y_{2t-1}, \quad t = 1, 2, \dots, T/2, \quad (13)$$

$$V_{t,1} = g_0 y_{2t} + g_1 y_{2t-1}, \quad t = 1, 2, \dots, T/2. \quad (14)$$

These statements are formalized in the following lemma.

**Lemma 2.3** *Under  $H_0$ ,  $\widehat{G}_{T,1} = \frac{1}{2} + o_p(1)$ , while under  $H_1$ ,  $\widehat{G}_{T,1} = \frac{E(h_0 y_{2t} + h_1 y_{2t-1})^2}{E(h_0 y_{2t} + h_1 y_{2t-1})^2 + E(g_0 y_{2t} + g_1 y_{2t-1})^2} + o_p(1)$ .*

Equations (13) and (14) imply:

$$W_{t,1}^2 = (h_0 y_{2t}^2 + h_1 y_{2t-1}^2 + 2h_0 h_1 y_{2t} y_{2t-1}) \text{ and } V_{t,1}^2 = (g_0 y_{2t}^2 + g_1 y_{2t-1}^2 + 2g_0 g_1 y_{2t} y_{2t-1}) \quad (15)$$

Using Equation (15), together with Equation (4), we obtain the following under  $H_0$

$$\widehat{G}_{T,1} = \frac{\sum_{t=1}^{T/2} W_{t,1}^2}{\sum_{t=1}^{T/2} V_{t,1}^2 + \sum_{t=1}^{T/2} W_{t,1}^2} \quad (16)$$

$$= \frac{h_0^2 \sum_{t=1}^{T/2} y_{2t}^2 + h_1^2 \sum_{t=1}^{T/2} y_{2t-1}^2 + 2h_0 h_1 \sum_{t=1}^{T/2} y_{2t} y_{2t-1}}{(h_0^2 + g_0^2) \sum_{t=1}^{T/2} y_{2t}^2 + (h_1^2 + g_1^2) \sum_{t=1}^{T/2} y_{2t-1}^2 + 2(h_0 h_1 + g_0 g_1) \sum_{t=1}^{T/2} y_{2t} y_{2t-1}} \quad (17)$$

---

<sup>8</sup>This assumption can easily be relaxed under several boundary treatment conditions.

$(h_0h_1 + g_0g_1) = (h_0h_1 - h_0h_1) = 0$ ,  $h_0^2 + h_1^2 = 1$ ,  $g_0^2 + g_1^2 = 1$ ,  $h_1 = -h_0$ ,  $h_1^2 = h_0^2$  and  $g_0 = -h_1$ ,  $g_1 = h_0$ <sup>9</sup> so that

$$\widehat{G}_{T,1} = \frac{h_0^2(\sum_{t=1}^{T/2} y_{2t}^2 + \sum_{t=1}^{T/2} y_{2t-1}^2) - 2h_0^2 \sum_{t=1}^{T/2} y_{2t}y_{2t-1}}{\sum_{t=1}^{T/2} y_{2t} + \sum_{t=1}^{T/2} y_{2t-1}} \quad (18)$$

$$= h_0^2 - \frac{2h_0^2 \sum_{t=1}^{T/2} y_{2t}y_{2t-1}}{\sum_{t=1}^{T/2} y_{2t}^2 + \sum_{t=1}^{T/2} y_{2t-1}^2} = h_0^2 - \frac{o_p(T)}{O_p(T)} = h_0^2 + o_p(1) \quad (19)$$

The asymptotic null distribution of  $\widehat{G}_{T,1}$  under  $H_0$  is summarized in the following theorem.

**Theorem 2.4** *Under  $H_0$ ,  $\sqrt{T/2} \frac{(\widehat{G}_{T,1} - h_0^2)}{h_0^2} \implies N(0, 1)$  where  $N(0, 1)$  is the standard normal distribution.*

**Proof:** Noting that

$$\widehat{G}_{T,1} - h_0^2 = -2h_0^2 \frac{\sum_{t=1}^{T/2} y_{2t}y_{2t-1}}{\sum_{t=1}^{T/2} (y_{2t}^2 + y_{2t-1}^2)} \quad (20)$$

$$= -2h_0^2 \frac{N(0, \sigma^4 T/2)}{2\sigma^2 T/2} + o_p(1) = -\frac{2h_0^2 \sigma^2 (T/2)^{1/2} N(0, 1)}{2\sigma^2 T/2} + o_p(1) \quad (21)$$

$$= -h_0^2 \frac{N(0, 1)}{\sqrt{T/2}} + o_p(1) \quad (22)$$

$$\sqrt{T/2} \frac{(\widehat{G}_{T,1} - h_0^2)}{h_0^2} = N(0, 1) + o_p(1) \text{ under the } H_0. \quad (23)$$

For Haar filter,  $h_0^2 = 1/2$  so that we obtain the same result in Equation (12).

### 2.3 DWT - General Filter Case

Since as the length of the filter  $L$  increases, the approximation of the Daubechies wavelet filter to the ideal high-pass filter improves<sup>10</sup>, we expect tests based on  $\widehat{G}_{T,1}^L$  to gain power as  $L$  increases. Our goal here is to capitalize on such power gains through more general filters. For a general wavelet filter  $\{h_l\}_{l=0}^{L-1}$ , the unit scale wavelet and scaling coefficients are<sup>11</sup> given by

$$W_{t,1} = \sum_{l=0}^{L-1} h_l y_{2t-l} \text{ mod } T, \quad V_{t,1} = \sum_{l=0}^{L-1} g_l y_{2t-l} \text{ mod } T, \quad (24)$$

<sup>9</sup>This is from the quadrature mirror filter property,  $g_l = (-1)^{l+1} h_{L-1-l}$ .

<sup>10</sup>Percival and Walden (2000) provide an excellent discussion on this matter.

<sup>11</sup> $a - b \text{ mod } T$  stands for “ $a - b$  modulo  $T$ ”. If  $j$  is an integer such that  $1 \leq j \leq T$ , then  $j \text{ mod } T \equiv j$ . If  $j$  is another integer, then  $j \text{ mod } T \equiv j + nT$  where  $nT$  is the unique integer multiple of  $T$  such that  $1 \leq j + nT \leq T$ .

where  $t = 1, \dots, T/2$  and  $T$  is assumed to be even. Again the wavelet coefficients  $\{W_{t,1}\}$  extract the high frequency information in  $\{y_t\}$ , whereas scaling coefficients  $\{V_{t,1}\}$  extract the low frequency information in  $\{y_t\}$ . This implies that the variance of the wavelet and scaling coefficients should be evenly distributed under  $H_0$ , which forms the basis for serial correlation tests. The following definition for  $\widehat{G}_{T,1}^L$

$$\widehat{G}_{T,1}^L = \frac{\sum_{t=1}^{T/2} W_{t,1}^2}{\sum_{t=1}^{T/2} V_{t,1}^2 + \sum_{t=1}^{T/2} W_{t,1}^2} \quad (25)$$

forms the basis of the serial correlation test. Heuristically,  $\widehat{G}_{T,1}^L$  should be close to  $1/2$  under  $H_0$ , since the numerator is the half of the denominator, while under  $H_1$ ,  $\widehat{G}_{T,1}^L$  is bounded in interval  $(0, 1)$ . It is the relative magnitude of the variance of the wavelet coefficients to that of the scaling coefficients, together with filters with better frequency localization features, which will determine the power.

Equations 24 imply:

$$\begin{aligned} \sum_{t=1}^{T/2} W_{t,1}^2 &= \sum_{t=1}^{T/2} \left( \sum_{l=0}^{L-1} h_l y_{2t-l \bmod T} \right)^2 \\ &= \sum_{l=0}^{L-1} h_l^2 \sum_{t=1}^{T/2} y_{2t-l \bmod T}^2 + 2 \sum_{j=0}^{L-2} h_j \sum_{l=j}^{L-2} h_{l+1} \sum_{t=1}^{T/2} y_{2t-j \bmod T} y_{2t-1-l \bmod T} \end{aligned} \quad (26)$$

$$\begin{aligned} \sum_{t=1}^{T/2} V_{t,1}^2 &= \sum_{t=1}^{T/2} \left( \sum_{l=0}^{L-1} g_l y_{2t-l \bmod T} \right)^2 \\ &= \sum_{l=0}^{L-1} g_l^2 \sum_{t=1}^{T/2} y_{2t-l \bmod T}^2 + 2 \sum_{j=0}^{L-2} g_j \sum_{l=j}^{L-2} h_{l+1} \sum_{t=1}^{T/2} y_{2t-j \bmod T} y_{2t-1-l \bmod T} \end{aligned} \quad (27)$$

$$\sum_{t=1}^{T/2} (W_{t,1}^2 + V_{t,1}^2) = \quad (28)$$

$$\begin{aligned} &= \sum_{l=0}^{L-1} (h_l^2 + g_l^2) \sum_{t=1}^{T/2} y_{2t-l \bmod T}^2 + \\ &\quad 2 \sum_{j=0}^{L-2} (h_j + g_j) \sum_{l=j}^{L-2} (h_{l+1} + g_{l+1}) \sum_{t=1}^{T/2} y_{2t-j \bmod T} y_{2t-1-l \bmod T} \end{aligned} \quad (29)$$

The reduced form of the denominator for  $\widehat{G}_{T,1}^L$  in Equation (25) is stated in the following lemma.

**Lemma 2.5**  $\sum_{t=1}^{T/2} (W_{t,1}^2 + V_{t,1}^2) = \sum_{t=1}^{T/2} y_{2t}^2 + \sum_{t=2}^{T/2} y_{2t-1}^2 = \sum_{t=1}^T y_t^2$ .

**Proof:** See Appendix B.

The asymptotic null distribution of  $\widehat{G}_{T,1}$  under  $H_0$  is summarized in the following theorem.

**Theorem 2.6** Under  $H_0$ ,  $\sqrt{T/2} \frac{(\widehat{G}_{T,1} - 1/2)}{\left(\sum_{l=1}^{L-(L/2)} h_{2l-1}^2\right)^2} = \sqrt{2T}(\widehat{G}_{T,1} - 1/2) \implies N(0, 1) + o_p(1)$  where  $N(0, 1)$  is the standard normal distribution.

**Proof:** See Appendix B.

Since  $\left(\sum_{l=1}^{L-(L/2)} h_{2l-1}^2\right)^2 = 1/2$ , the limiting distribution of the test statistic is same as in Equation (12).

## 2.4 MODWT - General Filter Case

For a general wavelet filter  $\{\tilde{h}_l\}_{l=0}^{L-1}$ , the unit scale wavelet and scaling coefficients are<sup>12</sup> given by

$$\widetilde{W}_{t,1} = \sum_{l=0}^{L-1} \tilde{h}_l y_{t-l} \bmod T, \quad \widetilde{V}_{t,1} = \sum_{l=0}^{L-1} \tilde{g}_l y_{t-l} \bmod T, \quad (30)$$

where  $t = 1, \dots, T$ . Again the wavelet coefficients  $\{\widetilde{W}_{t,1}\}$  extract the high frequency information in  $\{y_t\}$ , whereas scaling coefficients  $\{\widetilde{V}_{t,1}\}$  extract the low frequency information in  $\{y_t\}$ . This implies that the variance of the wavelet and scaling coefficients should be evenly distributed under  $H_0$ , which forms the basis for serial correlation tests. The following definition for  $\widetilde{G}_{T,1}^L$

$$\widetilde{G}_{T,1}^L = \frac{\sum_{t=1}^T \widetilde{W}_{t,1}^2}{\sum_{t=1}^T \widetilde{V}_{t,1}^2 + \sum_{t=1}^T \widetilde{W}_{t,1}^2} \quad (31)$$

forms the basis of the serial correlation test. Heuristically,  $\widetilde{G}_{T,1}^L$  should be close to 1/2 under  $H_0$ , since the numerator is the half of the denominator, while under  $H_1$ ,  $\widetilde{G}_{T,1}^L$  is bounded in interval  $(0, 1)$ . It is the relative magnitude of the variance of the wavelet coefficients to that of the scaling coefficients, together with filters with better frequency localization features, which will determine the power.

Equation (30) imply:

$$\sum_{t=1}^T \widetilde{W}_{t,1}^2 = \sum_{t=1}^T \left( \sum_{l=0}^{L-1} \tilde{h}_l y_{t-l} \bmod T \right)^2 \quad (32)$$

---

<sup>12</sup> $a - b \bmod T$  stands for “ $a - b$  modulo  $T$ ”. If  $j$  is an integer such that  $1 \leq j \leq T$ , then  $j \bmod T \equiv j$ . If  $j$  is another integer, then  $j \bmod T \equiv j + nT$  where  $nT$  is the unique integer multiple of  $T$  such that  $1 \leq j + nT \leq T$ .

$$\begin{aligned}
&= \sum_{l=0}^{L-1} \tilde{h}_l^2 \sum_{t=1}^T y_{t-l \bmod T}^2 + 2 \sum_{j=0}^{L-2} \tilde{h}_j \sum_{l=j}^{L-2} \tilde{h}_{l+1} \sum_{t=1}^T y_{t-j \bmod T} y_{t-1-l \bmod T} \\
\sum_{t=1}^T \tilde{V}_{t,1}^2 &= \sum_{t=1}^T \left( \sum_{l=0}^{L-1} \tilde{g}_l y_{t-l \bmod T} \right)^2 \tag{33} \\
&= \sum_{l=0}^{L-1} \tilde{g}_l^2 \sum_{t=1}^T y_{t-l \bmod T}^2 + 2 \sum_{j=0}^{L-2} \tilde{g}_j \sum_{l=j}^{L-2} \tilde{g}_{l+1} \sum_{t=1}^T y_{t-j \bmod T} y_{t-1-l \bmod T}
\end{aligned}$$

$$\sum_{t=1}^T \left( \tilde{W}_{t,1}^2 + \tilde{V}_{t,1}^2 \right) = \tag{34}$$

$$\begin{aligned}
&= \sum_{l=0}^{L-1} (\tilde{h}_l^2 + \tilde{g}_l^2) \sum_{t=1}^T y_{t-l \bmod T}^2 + \\
&\quad 2 \sum_{j=0}^{L-2} (\tilde{h}_j + \tilde{g}_j) \sum_{l=j}^{L-2} (\tilde{h}_{l+1} + \tilde{g}_{l+1}) \sum_{t=1}^T y_{t-j \bmod T} y_{t-1-l \bmod T} \tag{35}
\end{aligned}$$

The reduced form of the denominator for  $\hat{G}_{T,1}^L$  in Equation (31) is stated in the following lemma.

**Lemma 2.7**  $\sum_{t=1}^T \left( \tilde{W}_{t,1}^2 + \tilde{V}_{t,1}^2 \right) = \sum_{t=1}^T y_t^2$ .

**Proof:** See Appendix B.

The asymptotic null distribution of  $\tilde{G}_{T,1}$  under  $H_0$  is summarized in the following theorem.

**Theorem 2.8** Under  $H_0$ ,  $\sqrt{T} \frac{(\tilde{G}_{T,1} - 1/2)}{\left(2 \sum_{l=1}^{L-(L/2)} \tilde{h}_{2l-1}\right)^2} = \sqrt{4T}(\tilde{G}_{T,1} - 1/2) \implies N(0, 1) + o_p(1)$  where

$N(0, 1)$  is the standard normal distribution and  $\left( \sum_{l=1}^{L-(L/2)} \tilde{h}_{2l-1}^2 \right)^2 = 1/4$ .

**Proof:** See Appendix B.

### 3 Residual-based Tests

Let  $y_t = \mathbf{x}_t' \beta + u_t$  where  $\mathbf{x}_t$  is a vector of exogenous regressors.  $\{u_t\}$  is a weakly stationary process with  $E(u_t) = 0$ ,  $Var(u_t) = \sigma^2$ ,  $Cov(u_t, u_{t-j}) = E[u_t u_{t-j}] = \gamma_j$  for all  $j \geq 0$ . The  $j$ th order autocorrelation is  $\rho_j = \gamma_j / \gamma_0$ . Let  $\hat{\beta}$  be any consistent estimator of  $\beta$  obtained from the observed

sample and let  $\hat{u}_t = y_t - \mathbf{x}'_t \hat{\beta}$ . We consider tests for  $H_0 : \rho_j = 0$  for all  $j \geq 1$  against  $H_1 : \rho_j \neq 0$  and  $0 < |\rho_j| < 1$  for some  $j \geq 1$ .

We illustrate the test with level-one MODWT decomposition with Haar filter below,

$$\tilde{G}_{T,1}^r = \frac{\sum_{t=1}^T \tilde{W}_{t,1}^2}{\sum_{t=1}^T \tilde{V}_{t,1}^2 + \sum_{t=1}^T \tilde{W}_{t,1}^2} = \frac{\frac{1}{2} \sum_{t=1}^T \hat{u}_t^2 - \frac{1}{2} \sum_{t=2}^T \hat{u}_t \hat{u}_{t-1}}{\sum_{t=1}^T \hat{u}_t^2} = \frac{1}{2} - \frac{\frac{1}{2} \sum_{t=2}^T \hat{u}_t \hat{u}_{t-1}}{\sum_{t=1}^T \hat{u}_t^2}$$

The null distribution in this particular case will be  $\sqrt{4T}(\tilde{G}_{T,1}^r - 1/2) \implies N(0, 1)$ .

It is interesting to note that under  $H_0$ ,  $\tilde{G}_{T,1}^r$  can be expressed as

$$\tilde{G}_{T,1}^r = \frac{\sum_{t=1}^T \tilde{W}_{t,1}^2}{\sum_{t=1}^T \tilde{W}_{t,1}^2 + \sum_{t=1}^T \tilde{V}_{t,1}^2} = \frac{\frac{1}{4} \sum_{t=2}^T (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^T \hat{u}_t^2} \quad (36)$$

since the denominator is equal to the overall variance of the data. Equation (36) differs from the Durbin-Watson test only by the factor 1/4 in the numerator. The value of the Durbin-Watson test lies between 0 and 4 and the wavelet test lies between 0 and 1. The wavelet test has a simple null distribution which is standard normal.

## 4 Monte Carlo Simulations

In this section, we investigate the finite sample performance of the new wavelet tests.<sup>13</sup> Figures 1 and 2 illustrate that empirical distribution of the wavelet-based tests closely approximates the standard normal distribution for sample sizes as small as 50. Tables 1 and 2 report the results of the portmanteau tests where the wavelet test ( $\hat{G}_{T,1}$ ) is compared to the Ljung-Box (LB) and Box-Pierce (BP) tests. Comparisons are carried out at the 1% and 5% levels. The data is simulated from an AR(1) process,  $y_t = \phi y_{t-1} + u_t$ , where  $u_t \sim iidN(0, 1)$  and MA(1) process,  $y_t = u_t + \theta u_{t-1}$ , where  $u_t \sim iidN(0, 1)$ . All simulations are with 200 observations and 5,000 replications.

We provide two sets of wavelet tests results, one with discrete wavelet transformation (DWT) in Table 1, and the other is maximum overlap discrete wavelet transformation (MODWT) in Table 2. The DWT portmanteau test in Table 1,  $\hat{G}_{T,1}$ , has good empirical size relative to LB and BP tests. The empirical size of  $\hat{G}_{T,1}$  is 0.011 and 0.050 at the 1% and 5% levels. The empirical size of LB and BP tests are 0.019, 0.063 and 0.011, 0.041 for 1% and 5% levels, respectively. The LB test over rejects at both nominal levels. The BP tests under rejects at the 5% level. Furthermore, empirical size of LB and BP tests are sensitive to the lag length selection where the degree of over or under rejection magnifies significantly at different lag lengths. The DWT portmanteau test possesses significant power advantage over its competitors. The power of  $\hat{G}_{T,1}$  can be as large as 91% higher than its competitors.

---

<sup>13</sup>In the following tables, we report empirical size and power and do not adjust the empirical power for variations in empirical size.

In Table 2, we study the MODWT-based portmanteau test. Similar to Table 1, the  $\tilde{G}_{T,1}$  test has almost exact size whereas its competitors suffer from size distortions. The size distortions of the LB and BP tests vary across different lag lengths which is difficult to choose optimally. With MODWT-based wavelet test  $\tilde{G}_{T,1}$ , the power can be as large of 354% relative to the powers of LB and BP tests.

In Table 3, we study the residual-based tests. The MODWT test,  $\tilde{G}_{T,1}^r$ , has good empirical size relative to DW- $d$  and BG tests. BG test has serious size distortions and this distortion gets worse at higher lags. Given such desirable empirical size and better power, the  $\tilde{G}_{T,1}^r$  test is a reliable, practical residual-based test statistic.

## 5 Conclusions

Our tests provide a novel approach in separating the variance of the data by constructing test statistics from its lower and higher frequency dynamics. Our results provide a unifying framework where Durbin-Watson  $d$  test is a special case of a wavelet-based test. The intuitive construction and simplicity are worth emphasizing. The simulation studies demonstrate the significant power improvement of our tests with desirable empirical sizes.

		$\widehat{G}_{T,1}$		LB		BP	
	AR(1)/ $\phi$	1%	5%	1%	5%	1%	5%
	-0.30	0.675	0.866	0.596	0.782	0.549	0.746
	-0.20	0.272	0.516	0.207	0.371	0.172	0.322
Size	0.00	0.011	0.050	0.019	0.063	0.011	0.041
	0.20	0.275	0.517	0.179	0.353	0.146	0.302
	0.30	0.682	0.866	0.553	0.740	0.501	0.699
		MA(1)/ $\theta$		LB		BP	
	MA(1)/ $\theta$	1%	5%	1%	5%	1%	5%
	-0.30	0.575	0.794	0.421	0.644	0.369	0.592
	-0.20	0.237	0.477	0.156	0.317	0.124	0.262
Size	0.00	0.011	0.050	0.019	0.063	0.011	0.041
	0.20	0.242	0.480	0.176	0.339	0.139	0.289
	0.30	0.580	0.800	0.441	0.667	0.385	0.611

Table 1: SIZE AND POWER OF THE  $\widehat{G}_{T,1}$  TEST

The wavelet test statistic is calculated with a unit scale DWT and with the Haar filter. The AR(1) data is simulated from  $y_t = \phi y_{t-1} + u_t$ , where  $u_t \sim iidN(0, 1)$ . The MA(1) data is simulated from  $y_t = u_t + \theta u_{t-1}$ , where  $u_t \sim iidN(0, 1)$ . All simulations are with 5,000 replications and 200 observations.  $\widehat{G}_{T,1}$  is the wavelet test which is based on standard normal critical values of a two-sided test. LB and BP are Ljung-Box and Box-Pierce tests which are based on chi-squared distribution with 20 degrees of freedom.

		$\tilde{G}_{T,1}$		LB		BP	
	AR(1)/ $\phi$	1%	5%	1%	5%	1%	5%
	-0.30	0.949	0.987	0.596	0.782	0.549	0.746
	-0.20	0.593	0.805	0.207	0.371	0.172	0.322
Size	0.00	0.009	0.051	0.019	0.063	0.011	0.041
	0.20	0.582	0.796	0.179	0.353	0.146	0.302
	0.30	0.950	0.984	0.553	0.740	0.501	0.699
		$\tilde{G}_{T,1}$		LB		BP	
	MA(1)/ $\theta$	1%	5%	1%	5%	1%	5%
	-0.30	0.921	0.979	0.421	0.644	0.369	0.592
	-0.20	0.552	0.786	0.156	0.317	0.124	0.262
Size	0.00	0.009	0.051	0.019	0.063	0.011	0.041
	0.20	0.553	0.776	0.176	0.339	0.139	0.289
	0.30	0.920	0.982	0.441	0.667	0.385	0.611

Table 2: SIZE AND POWER OF THE  $\tilde{G}_{T,1}$  TEST

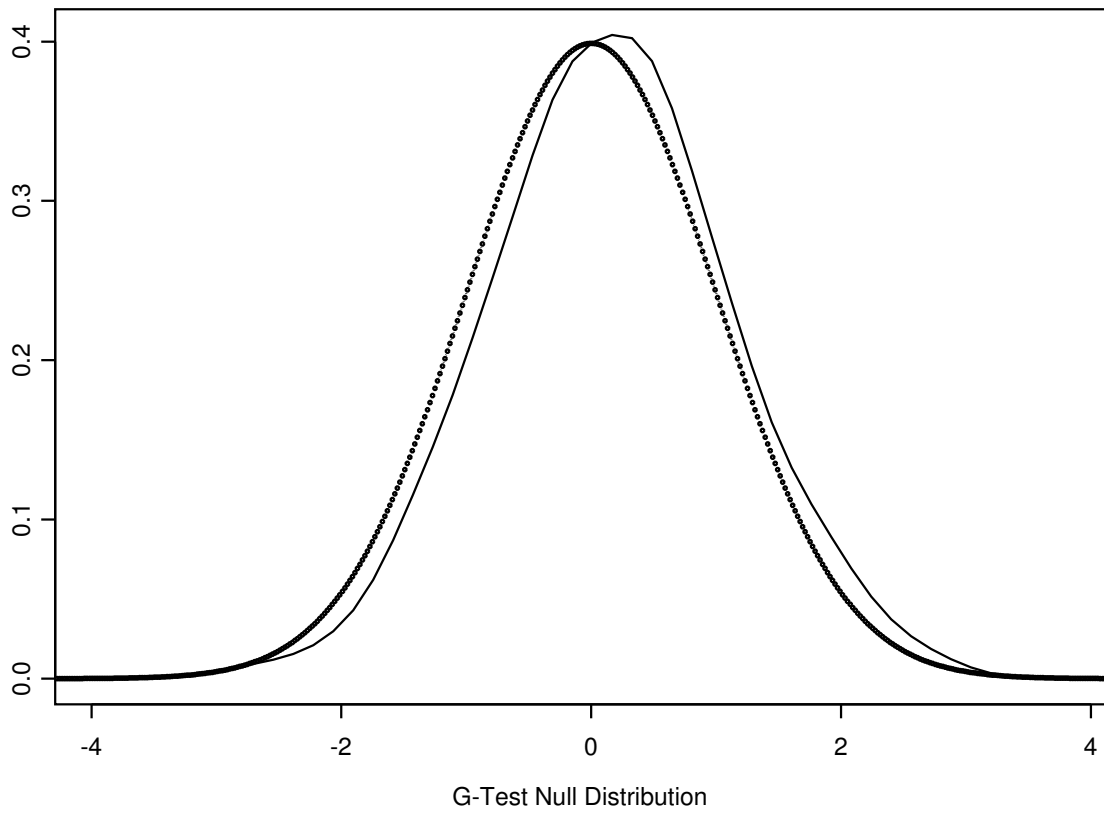
The wavelet test statistic is calculated with a unit scale MODWT and with the Haar filter. The AR(1) data is simulated from  $y_t = \rho y_{t-1} + u_t$ , where  $u_t \sim iidN(0, 1)$ . The MA(1) data is simulated from  $y_t = u_t + \theta u_{t-1}$ , where  $u_t \sim iidN(0, 1)$ . All simulations are with 5,000 replications and 200 observations.  $\tilde{G}_{T,1}$  is the wavelet test which is based on standard normal critical values of a two-sided test. LB and BP are Ljung-Box and Box-Pierce tests which are based on chi-squared distribution with 20 degrees of freedom.

	$\tilde{G}_{T,1}^r$			DW- $d$		BG	
AR(1)/ $\phi$	2%	10%		2%	10%	2%	10%
-0.30	0.345	0.633		0.267	0.519	0.341	0.621
Size 0.00	0.019	0.097		0.012	0.084	0.025	0.120
0.30	0.256	0.526		0.249	0.502	0.215	0.511
MA(1)/ $\theta$	2%	10%		2%	10%	2%	10%
-0.30	0.168	0.427		0.166	0.402	0.161	0.418
Size 0.00	0.019	0.097		0.012	0.084	0.025	0.120
0.30	0.282	0.567		0.195	0.445	0.279	0.562

Table 3: SIZE AND POWER OF THE  $\tilde{G}_{T,1}^r$  TEST

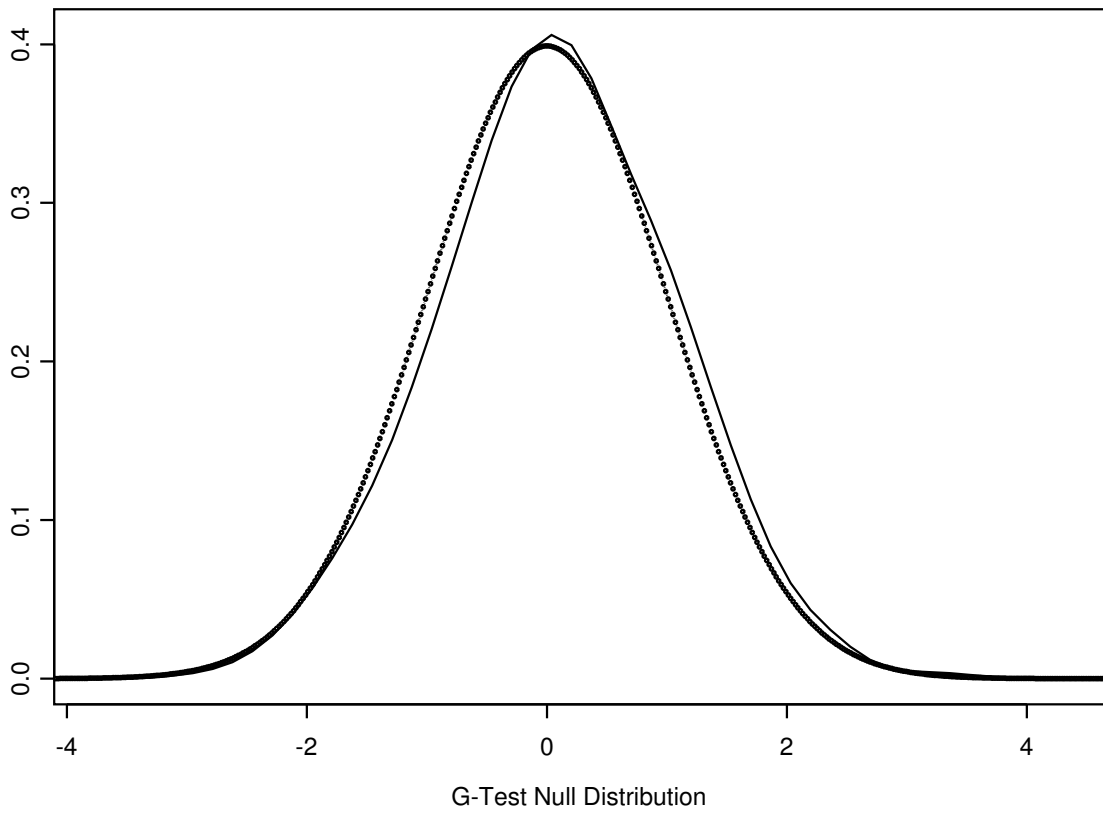
The wavelet test statistic is calculated with a unit scale MODWT and with the Haar filter. The data is simulated from  $y_t = 1 + 2x_{1t} + 3x_{2t} + 4x_{3t} - 5x_{4t} - 6x_{5t} + u_t$ ,  $u_t = \rho u_{t-1} + \epsilon_t$  where  $\epsilon_t \sim iidN(0, 1)$  and  $|\rho| < 1$ . Under the null hypothesis,  $\rho = 0$  and under the alternative  $\rho \neq 0$ .  $\{x_{it}\}_{i=1}^5$  are generated from multivariate normal distribution with a correlation coefficient of 0.1. All simulations are with 50 observations and 5,000 replications.  $\tilde{G}_{T,1}^r$  is the wavelet test, DW- $d$  is the Durbin-Watson test and  $BG$  is the Breusch-Godfrey test. Durbin-Watson significance levels are calculated for a two-sided alternative with critical vales of  $d > 4 - 1.32$ , or  $d < 1.32$  for 2 percent level and  $d > 4 - 1.50$ , or  $d < 1.50$  for 10 percent level for  $T = 50$ . Breusch-Godfrey test critical values are calculated with  $\chi^2(1)$ .

Figure 1: The null distribution of  $\tilde{G}_{T,1}^r$



Circles: The null distribution of  $\tilde{G}_{T,1}^r$  for  $T = 50$  with 5,000 simulations. Solid Line:  $N(0, 1)$ .

Figure 2: The null distribution of  $\tilde{G}_{T,1}^r$



Circles: The null distribution of  $\tilde{G}_{T,1}^r$  for  $T = 100$  with 5,000 simulations. Solid Line:  $N(0, 1)$ .

## Appendix A - Wavelet Transformations<sup>14</sup>

A wavelet is a small wave which grows and decays in a limited time period.<sup>15</sup> To formalize the notion of a wavelet, let  $\psi(\cdot)$  be a real valued function such that its integral is zero,  $\int_{-\infty}^{\infty} \psi(t) dt = 0$ , and its square integrates to unity,  $\int_{-\infty}^{\infty} \psi(t)^2 dt = 1$ . Thus, although  $\psi(\cdot)$  has to make some excursions away from zero, any excursions it makes above zero must cancel out excursions below zero, i.e.,  $\psi(\cdot)$  is a small wave, or a wavelet.

Fundamental properties of the continuous wavelet functions (filters), such as integration to zero and unit variance, have discrete counterparts. Let  $h = (h_0, \dots, h_{L-1})$  be a finite length discrete wavelet (or high pass) filter such that it integrates (sums) to zero,  $\sum_{l=0}^{L-1} h_l = 0$ , and has unit variance,  $\sum_{l=0}^{L-1} h_l^2 = 1$ . In addition, the wavelet filter  $h$  is orthogonal to its even shifts; that is,

$$\sum_{l=0}^{L-1} h_l h_{l+2n} = \sum_{l=-\infty}^{\infty} h_l h_{l+2n} = 0, \quad \text{for all nonzero integers } n. \quad (37)$$

The natural object to complement a high-pass filter is a low-pass (scaling) filter  $g$ . We will denote a low-pass filter as  $g = (g_0, \dots, g_{L-1})$ . The low-pass filter coefficients are determined by the *quadrature mirror relationship*<sup>16</sup>

$$g_l = (-1)^{l+1} h_{L-1-l} \quad \text{for } l = 0, \dots, L-1 \quad (38)$$

and the inverse relationship is given by  $h_l = (-1)^l g_{L-1-l}$ . The basic properties of the scaling filter are:  $\sum_{l=0}^{L-1} g_l = \sqrt{2}$ ,  $\sum_{l=0}^{L-1} g_l^2 = 1$ ,

$$\sum_{l=0}^{L-1} g_l g_{l+2n} = \sum_{l=-\infty}^{\infty} g_l g_{l+2n} = 0, \quad (39)$$

for all nonzero integers  $n$ , and

$$\sum_{l=0}^{L-1} g_l h_{l+2n} = \sum_{l=-\infty}^{\infty} g_l h_{l+2n} = 0 \quad (40)$$

for all integers  $n$ . Thus, scaling filters are average filters and their coefficients satisfy the orthonormality property that they possess unit variance and are orthogonal to even shifts. By applying both

---

<sup>14</sup>This appendix offers a brief introduction to Wavelet transformations. Interested readers can consult Gençay *et al.* (2001) or Percival and Walden (2000) for more details.

<sup>15</sup>This section closely follows Gençay *et al.* (2001), see also Percival and Walden (2000). The contrasting notion is a big wave such as the sine function which keeps oscillating indefinitely.

<sup>16</sup>Quadrature mirror filters (QMFs) are often used in the engineering literature because of their ability for perfect reconstruction of a signal without aliasing effects. Aliasing occurs when a continuous signal is sampled to obtain a discrete time series.

$h$  and  $g$  to an observed time series, we can separate high-frequency oscillations from low-frequency ones. In the following sections, we will briefly describe discrete wavelet transformation (DWT) and maximum overlap discrete wavelet transformation (MODWT).

### A.1 Discrete Wavelet Transformation

With both wavelet filter coefficients and scaling filter coefficients, we can decompose the data using the (discrete) wavelet transformation (DWT). Formally, let us introduce the DWT through a simple matrix operation. Let  $\mathbf{y}$  to be the dyadic length vector ( $T = 2^J$ ) of observations. The length  $T$  vector of discrete wavelet coefficients  $\mathbf{w}$  is obtained via

$$\mathbf{w} = \mathcal{W}\mathbf{y}$$

where  $\mathcal{W}$  is an  $T \times T$  orthonormal matrix defining the DWT. The vector of wavelet coefficients can be organized into  $J + 1$  vectors,  $\mathbf{w} = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_J, \mathbf{v}_J]'$ , where  $\mathbf{w}_j$  is a length  $T/2^j$  vector of wavelet coefficients associated with changes on a scale of length  $\lambda_j = 2^{j-1}$ , and  $\mathbf{v}_J$  is a length  $T/2^J$  vector of scaling coefficients associated with averages on a scale of length  $2^J = 2\lambda_J$ .

The matrix  $\mathcal{W}$  is composed of the wavelet and scaling filter coefficients arranged on a row-by-row basis. Let

$$\mathbf{h}_1 = [h_{1,N-1}, h_{1,N-2}, \dots, h_{1,1}, h_{1,0}]'$$

be the vector of zero-padded unit scale wavelet filter coefficients in reverse order. Thus, the coefficients  $h_{1,0}, \dots, h_{1,L-1}$  are taken from an appropriate ortho-normal wavelet family of length  $L$ , and all values  $L < t < T$  are defined to be zero. Now circularly shift  $\mathbf{h}_1$  by factors of two so that

$$\begin{aligned} \mathbf{h}_1^{(2)} &= [h_{1,1}, h_{1,0}, h_{1,N-1}, h_{1,N-2} \dots, h_{1,3}, h_{1,2}]' \\ \mathbf{h}_1^{(4)} &= [h_{1,3}, h_{1,2}, h_{1,1}, h_{1,0} \dots, h_{1,5}, h_{1,4}]' \end{aligned}$$

and so on. Define the  $T/2 \times T$  dimensional matrix  $\mathcal{W}_1$  to be the collection of  $T/2$  circularly shifted versions of  $\mathbf{h}_1$ . Hence,

$$\mathcal{W}_1 = [\mathbf{h}_1^{(2)}, \mathbf{h}_1^{(4)}, \dots, \mathbf{h}_1^{(T/2-1)}, \mathbf{h}_1]'$$

Let  $\mathbf{h}_2$  be the vector of zero-padded scale 2 wavelet filter coefficients defined similarly to  $\mathbf{h}_1$ .  $\mathcal{W}_2$  is constructed by circularly shifting the vector  $\mathbf{h}_2$  by factor of four. Repeat this to construct  $\mathcal{W}_j$  by circularly shifting the vector  $\mathbf{h}_j$  (the vector of zero-padded scale  $j$  wavelet filter coefficients) by  $2^j$ . The matrix  $\mathcal{V}_J$  is simply a column vector whose elements are all equal to  $1/\sqrt{T}$ . Then, the  $T \times T$  dimensional matrix  $\mathcal{W}$  is  $\mathcal{W} = [\mathcal{W}_1, \mathcal{W}_2, \dots, \mathcal{W}_J, \mathcal{V}_J]'$ .

When we are provided with a dyadic length time series, it is not necessary to implement the DWT down to level  $J = \log_2(T)$ . A *partial* DWT may be performed instead that terminates at level

$J_p < J$ . The resulting vector of wavelet coefficients will now contain  $T - T/2^{J_p}$  wavelet coefficients and  $T/2^{J_p}$  scaling coefficients.

The orthonormality of the matrix  $\mathcal{W}$  implies that the DWT is a variance preserving transformation:

$$\|\mathbf{w}\|^2 = \sum_{t=1}^{T/2^J} v_{t,J}^2 + \sum_{j=1}^J \left( \sum_{t=1}^{T/2^j} w_{t,j}^2 \right) = \sum_{t=1}^T y_t^2 = \|\mathbf{y}\|^2.$$

This can be easily proven through basic matrix manipulation via

$$\|\mathbf{y}\|^2 = \mathbf{y}'\mathbf{y} = (\mathcal{W}\mathbf{w})'\mathcal{W}\mathbf{w} = \mathbf{w}'\mathcal{W}'\mathcal{W}\mathbf{w} = \mathbf{w}'\mathbf{w} = \|\mathbf{w}\|^2.$$

Given the structure of the wavelet coefficients,  $\|\mathbf{y}\|^2$  is decomposed on a scale-by-scale basis via

$$\|\mathbf{y}\|^2 = \sum_{j=1}^J \|w_j\|^2 + \|v_J\|^2 \quad (41)$$

where  $\|w_j\|^2 = \sum_{t=1}^{T/2^j} w_{t,j}^2$  is the sum of squared variation of  $y$  due to changes at scale  $\lambda_j$  and  $\|v_J\|^2 = \sum_{t=1}^{T/2^J} v_{t,J}^2$  is the information due to changes at scales  $\lambda_J$  and higher.

## A.2 Maximum Overlap Discrete Wavelet Transformation

An alternative wavelet transform is maximum overlap discrete wavelet transformation (MODWT) which is computed by *not* sub-sampling the filtered output. Let  $\mathbf{y}$  be an arbitrary length  $T$  vector of observations. The length  $(J+1)T$  vector of MODWT coefficients  $\tilde{\mathbf{w}}$  is obtained via

$$\tilde{\mathbf{w}} = \tilde{\mathcal{W}}\mathbf{y},$$

where  $\tilde{\mathcal{W}}$  is a  $(J+1)T \times T$  matrix defining the MODWT. The vector of MODWT coefficients may be organized into  $J+1$  vectors via

$$\tilde{\mathbf{w}} = [\tilde{\mathbf{w}}_1, \tilde{\mathbf{w}}_2, \dots, \tilde{\mathbf{w}}_J, \tilde{\mathbf{v}}_J]^T, \quad (42)$$

where  $\tilde{\mathbf{w}}_j$  is a length  $T$  vector of wavelet coefficients associated with changes on a scale of length  $\lambda_j = 2^{j-1}$  and  $\tilde{\mathbf{v}}_J$  is a length  $T$  vector of scaling coefficients associated with averages on a scale of length  $2^J = 2\lambda_J$ , just as with the DWT.

Similar to the orthonormal matrix defining the DWT, the matrix  $\tilde{\mathcal{W}}$  is also made up of  $J+1$  sub-matrices, each of them  $T \times T$ , and may be expressed as

$$\tilde{\mathcal{W}} = \begin{bmatrix} \tilde{\mathcal{W}}_1 \\ \tilde{\mathcal{W}}_2 \\ \vdots \\ \tilde{\mathcal{W}}_J \\ \tilde{\mathbf{v}}_J \end{bmatrix}.$$

The MODWT utilizes the rescaled filters ( $j = 1, \dots, J$ )

$$\tilde{\mathbf{h}}_j = \mathbf{h}_j/2^{j/2} \quad \text{and} \quad \tilde{\mathbf{g}}_J = \mathbf{g}_J/2^{J/2}.$$

To construct the  $T \times T$  dimensional sub matrix  $\widetilde{\mathcal{W}}_1$ , we circularly shift the rescaled wavelet filter vector  $\tilde{\mathbf{h}}_1$  by integer units to the right so that

$$\widetilde{\mathcal{W}}_1 = \left[ \tilde{\mathbf{h}}_1^{(1)}, \tilde{\mathbf{h}}_1^{(2)}, \tilde{\mathbf{h}}_1^{(3)}, \dots, \tilde{\mathbf{h}}_1^{(N-2)}, \tilde{\mathbf{h}}_1^{(N-1)}, \tilde{\mathbf{h}}_1 \right]^T. \quad (43)$$

This matrix may be interpreted as the interweaving of the DWT sub matrix  $\mathcal{W}_1$  with a circularly shifted (to the right by one unit) version of itself. The remaining sub matrices  $\widetilde{\mathcal{W}}_2, \dots, \widetilde{\mathcal{W}}_J$  are formed similarly to Equation 43, only replace  $\tilde{\mathbf{h}}_1$  by  $\tilde{\mathbf{h}}_j$ .

In practice, a pyramid algorithm is utilized similar to that of the DWT to compute the MODWT. Starting with the data  $x_t$  (no longer restricted to be a dyadic length), filter it using  $\tilde{\mathbf{h}}_1$  and  $\tilde{\mathbf{g}}_1$  to obtain the length  $T$  vectors of wavelet and scaling coefficients  $\tilde{\mathbf{w}}_1$  and  $\tilde{\mathbf{v}}_1$ , respectively.

For each iteration of the MODWT pyramid algorithm, we require three objects: the data vector  $\mathbf{x}$ , the wavelet filter  $\tilde{h}_l$  and the scaling filter  $\tilde{g}_l$ . The first iteration of the pyramid algorithm begins by filtering (convolving) the data with each filter to obtain the following wavelet and scaling coefficients:

$$\tilde{w}_{1,t} = \sum_{l=0}^{L-1} \tilde{h}_l y_{t-l \bmod T} \quad \text{and} \quad \tilde{v}_{1,t} = \sum_{l=0}^{L-1} \tilde{g}_l y_{t-l \bmod T},$$

where  $t = 1, \dots, T$ . The length  $T$  vector of observations has been high- and low-pass filtered to obtain  $T$  coefficients associated with this information. The second step of the MODWT pyramid algorithm starts by defining the data to be the scaling coefficients  $\tilde{\mathbf{v}}_1$  from the first iteration and apply the filtering operations as above to obtain the second level of wavelet and scaling coefficients

$$\tilde{w}_{2,t} = \sum_{l=0}^{L-1} \tilde{h}_l \tilde{v}_{1,t-2l \bmod T} \quad \text{and} \quad \tilde{v}_{2,t} = \sum_{l=0}^{L-1} \tilde{g}_l \tilde{v}_{1,t-2l \bmod T},$$

$t = 1, \dots, T$ . Keeping all vectors of wavelet coefficients, and the final level of scaling coefficients, we have the following length  $T$  decomposition:  $\tilde{\mathbf{w}} = [\tilde{\mathbf{w}}_1 \tilde{\mathbf{w}}_2 \tilde{\mathbf{v}}_2]'$ . After the third iteration of the pyramid algorithm, where we apply filtering operations to  $\mathbf{v}_2$ , the decomposition now looks like  $\tilde{\mathbf{w}} = [\tilde{\mathbf{w}}_1 \tilde{\mathbf{w}}_2 \tilde{\mathbf{w}}_3 \tilde{\mathbf{v}}_3]'$ . This procedure may be repeated up to  $J$  times where  $J = \log_2(T)$  and gives the vector of MODWT coefficients in Equation 42.

Similar to DWT, MODWT wavelet and scaling coefficients are variance preserving

$$\|\tilde{\mathbf{w}}\|^2 = \sum_{t=1}^T \tilde{v}_{t,J}^2 + \sum_{j=1}^J \left( \sum_{t=1}^T \tilde{w}_{t,j}^2 \right) = \sum_{t=1}^T y_t^2 = \|\mathbf{y}\|^2.$$

and a partial decomposition  $J_p < J$  may be performed when it deems necessary.

The following properties are important for distinguishing the MODWT from the DWT. The MODWT can accommodate any sample size  $T$ , while the  $J_p$ th order partial DWT restricts the sample size to a multiple of  $2^{J_p}$ . The detail and smooth coefficients of a MODWT are associated with zero phase filters. Thus, events that feature in the original time series can be properly aligned with features in the MODWT multi resolution analysis. The MODWT is invariant to circular shifts in the original time series. This property does not hold for the DWT. The MODWT wavelet variance estimator is asymptotically more efficient than the same estimator based on the DWT. For both MODWT and DWT, the scaling coefficients contain the lowest frequency information. But each level's wavelet coefficients contain progressively lower frequency information.

## Appendix B – Proofs

**Proof of Lemma 2.5:** Let  $\gamma_j = (2/T) \sum_{t=1}^{T/2} y_{2t \bmod T} y_{2t-j \bmod T}$  so that

$$\begin{aligned}
 (2/T) \sum_{t=1}^{T/2} (W_{t,1}^2 + V_{t,1}^2) &= \tag{44} \\
 & \sum_{l=0}^{L-1} (h_l^2 + g_l^2) \gamma_0 + 2(h_0 + g_0) \sum_{l=0}^{L-2} (h_{l+1} + g_{l+1}) \gamma_{l+1} + \\
 & 2(h_1 + g_1) \sum_{l=1}^{L-2} (h_{l+1} + g_{l+1}) \gamma_{l+1} + 2(h_2 + g_2) \sum_{l=2}^{L-2} (h_{l+1} + g_{l+1}) \gamma_{l+1} + \dots + \\
 & 2(h_{L-2} + g_{L-2})(h_{L-1} + g_{L-1}) \gamma_1.
 \end{aligned}$$

Alternatively,

$$\begin{aligned}
 (2/T) \sum_{t=1}^{T/2} (W_{t,1}^2 + V_{t,1}^2) &= \tag{45} \\
 & \gamma_0 \sum_{l=0}^{L-1} (h_l^2 + g_l^2) + 2\gamma_1 \sum_{l=0}^{L-2} (h_l h_{l+1} + g_l g_{l+1}) + 2\gamma_2 \sum_{l=0}^{L-3} (h_l h_{l+2} + g_l g_{l+2}) + \\
 & 2\gamma_3 \sum_{l=0}^{L-4} (h_l h_{l+3} + g_l g_{l+3}) + \dots + 2\gamma_{L-1} (h_0 h_{L-1} + g_0 g_{L-1}).
 \end{aligned}$$

Noting that  $\{h_l\}_{l=0}^{L-1}$  and  $\{g_l\}_{l=0}^{L-1}$  are  $\sum_{l=0}^{L-1} h_l^2 = 1$ ,  $\sum_{l=0}^{L-1} g_l^2 = 1$ , orthogonal to their even shifts

$$\begin{aligned}
 h_0 h_2 + h_1 h_3 + \dots + h_{L-3} h_{L-1} &= 0 \tag{46} \\
 h_0 h_4 + h_1 h_5 + \dots + h_{L-5} h_{L-1} &= 0 \\
 &\vdots \\
 h_0 h_{L-2} + h_1 h_{L-1} &= 0
 \end{aligned}$$

$$\begin{aligned}
 g_0 g_2 + g_1 g_3 + \dots + g_{L-3} g_{L-1} &= 0 \tag{47} \\
 g_0 g_4 + g_1 g_5 + \dots + g_{L-5} g_{L-1} &= 0 \\
 &\vdots \\
 g_0 g_{L-2} + g_1 g_{L-1} &= 0
 \end{aligned}$$

and because of the quadrature mirror filter  $g_l = (-1)^{l+1}h_{L-1-l}$ ,

$$\begin{aligned}
(h_0h_1 + g_0g_1) + (h_1h_2 + g_1g_2) + \dots + (h_{L-2}h_{L-1} + g_{L-2}g_{L-1}) &= 0 \\
(h_0h_3 + g_0g_3) + (h_1h_4 + g_1g_4) + \dots + (h_{L-4}h_{L-1} + g_{L-4}g_{L-1}) &= 0 \\
&\vdots \\
(h_0h_{L-1} + g_0g_{L-1}) &= 0
\end{aligned} \tag{48}$$

Placing the restrictions implied by Equations (46 - 48), together with  $\sum_{l=0}^{L-1} h_l^2 = 1$ ,  $\sum_{l=0}^{L-1} g_l^2 = 1$ , into Equation (45) leads to

$$\begin{aligned}
\sum_{t=1}^{T/2} (W_{t,1}^2 + V_{t,1}^2) &= 2(T/2)\gamma_0 \\
&= \sum_{t=1}^T y_t^2 = \sum_{t=1}^{T/2} y_{2t}^2 + \sum_{t=2}^{T/2} y_{2t-1}^2
\end{aligned} \tag{49}$$

**Proof of Theorem 2.6:** Let  $\gamma_j = (2/T) \sum_{t=1}^{T/2} y_{2t \bmod T} y_{2t-j \bmod T}$ .

$$(2/T) \sum_{t=1}^{T/2} (W_{t,1}^2) = \gamma_0 \sum_{l=0}^{L-1} h_l^2 + 2\gamma_1 \sum_{l=0}^{L-2} (h_l h_{l+1}) + 2\gamma_2 \sum_{l=0}^{L-3} (h_l h_{l+2}) + 2\gamma_3 \sum_{l=0}^{L-4} (h_l h_{l+3}) + \dots + 2\gamma_{L-1} (h_0 h_{L-1}).$$

because  $\{h_l\}_{l=0}^{L-1}$  is orthogonal to its even shifts and  $\sum_{l=0}^{L-1} h_l^2 = 1$

$$\begin{aligned} h_0 h_2 + h_1 h_3 + \dots + h_{L-3} h_{L-1} &= 0 \\ h_0 h_4 + h_1 h_5 + \dots + h_{L-5} h_{L-1} &= 0 \\ &\vdots \\ h_0 h_{L-2} + h_1 h_{L-1} &= 0 \end{aligned}$$

so that

$$(2/T) \sum_{t=1}^{T/2} (W_{t,1}^2) = \gamma_0 + 2\gamma_1 \sum_{l=0}^{L-2} (h_l h_{l+1}) + 2\gamma_3 \sum_{l=0}^{L-4} (h_l h_{l+3}) + 2\gamma_5 \sum_{l=0}^{L-6} (h_l h_{l+5}) + \dots + 2\gamma_{L-1} (h_0 h_{L-1})$$

By using Lemma (2.5),  $\widehat{G}_{T,1}^L$  is written as

$$\begin{aligned} \widehat{G}_{T,1}^L &= \frac{\sum_{t=1}^{T/2} W_{t,1}^2}{\sum_{t=1}^{T/2} V_{t,1}^2 + \sum_{t=1}^{T/2} W_{t,1}^2} \\ &= \frac{\sum_{t=1}^{T/2} y_{2t}^2 + 2(T/2) \left( \gamma_1 \sum_{l=0}^{L-2} (h_l h_{l+1}) + \gamma_3 \sum_{l=0}^{L-4} (h_l h_{l+3}) + \gamma_5 \sum_{l=0}^{L-6} (h_l h_{l+5}) + \dots + \gamma_{L-1} (h_0 h_{L-1}) \right)}{\sum_{t=1}^{T/2} y_{2t}^2 + \sum_{t=1}^{T/2} y_{2t-1}^2} \\ \widehat{G}_{T,1}^L - \frac{1}{2} &= \frac{2(T/2) \left( \gamma_1 \sum_{l=0}^{L-2} (h_l h_{l+1}) + \gamma_3 \sum_{l=0}^{L-4} (h_l h_{l+3}) + \gamma_5 \sum_{l=0}^{L-6} (h_l h_{l+5}) + \dots + \gamma_{L-1} (h_0 h_{L-1}) \right)}{\sum_{t=1}^{T/2} y_{2t}^2 + \sum_{t=1}^{T/2} y_{2t-1}^2} \\ &= \frac{2N(0, \sigma^4 T/2) \left( \sum_{l=0}^{L-2} (h_l h_{l+1}) + \sum_{l=0}^{L-4} (h_l h_{l+3}) + \sum_{l=0}^{L-6} (h_l h_{l+5}) + \dots + (h_0 h_{L-1}) \right)}{\sum_{t=1}^{T/2} y_{2t}^2 + \sum_{t=1}^{T/2} y_{2t-1}^2} \\ &= -\frac{2 \left( \sum_{l=1}^{L-(L/2)} h_{2l-1} \right)^2 N(0, \sigma^4 T/2)}{2\sigma^2 T/2} + o_p(1) \\ &= -\frac{N(0, 1) \left( \sum_{l=1}^{L-(L/2)} h_{2l-1} \right)^2}{\sqrt{T/2}} + o_p(1) \end{aligned}$$

Since  $\left(\sum_{l=1}^{L-(L/2)} h_{2l-1}\right)^2 = 1/2$ ,  $\sqrt{2T} \left(\widehat{G}_{T,1}^L - \frac{1}{2}\right) = N(0, 1) + o_p(1)$ .

**Proof of Lemma 2.7:** Let  $\gamma_j = (1/T) \sum_{t=1}^T y_{t \bmod T} y_{t-j \bmod T}$  so that

$$\begin{aligned}
T^{-1} \sum_{t=1}^T \left( \widetilde{W}_{t,1}^2 + \widetilde{V}_{t,1}^2 \right) &= \tag{50} \\
&\sum_{l=0}^{L-1} (\tilde{h}_l^2 + \tilde{g}_l^2) \gamma_0 + 2(\tilde{h}_0 + \tilde{g}_0) \sum_{l=0}^{L-2} (\tilde{h}_{l+1} + \tilde{g}_{l+1}) \gamma_{l+1} + \\
&2(\tilde{h}_1 + \tilde{g}_1) \sum_{l=1}^{L-2} (\tilde{h}_{l+1} + \tilde{g}_{l+1}) \gamma_{l+1} + 2(\tilde{h}_2 + \tilde{g}_2) \sum_{l=2}^{L-2} (\tilde{h}_{l+1} + \tilde{g}_{l+1}) \gamma_{l+1} + \dots + \\
&2(\tilde{h}_{L-2} + \tilde{g}_{L-2}) (\tilde{h}_{L-1} + \tilde{g}_{L-1}) \gamma_1.
\end{aligned}$$

Alternatively,

$$\begin{aligned}
T^{-1} \sum_{t=1}^T \left( \widetilde{W}_{t,1}^2 + \widetilde{V}_{t,1}^2 \right) &= \tag{51} \\
&\gamma_0 \sum_{l=0}^{L-1} (\tilde{h}_l^2 + \tilde{g}_l^2) + 2\gamma_1 \sum_{l=0}^{L-2} (\tilde{h}_l \tilde{h}_{l+1} + \tilde{g}_l \tilde{g}_{l+1}) + 2\gamma_2 \sum_{l=0}^{L-3} (\tilde{h}_l \tilde{h}_{l+2} + \tilde{g}_l \tilde{g}_{l+2}) + \\
&2\gamma_3 \sum_{l=0}^{L-4} (\tilde{h}_l \tilde{h}_{l+3} + \tilde{g}_l \tilde{g}_{l+3}) + \dots + 2\gamma_{L-1} (\tilde{h}_0 \tilde{h}_{L-1} + \tilde{g}_0 \tilde{g}_{L-1}).
\end{aligned}$$

Noting that  $\{\tilde{h}_l\}_{l=0}^{L-1}$  and  $\{\tilde{g}_l\}_{l=0}^{L-1}$  are  $\sum_{l=0}^{L-1} \tilde{h}_l^2 = 1/2$ ,  $\sum_{l=0}^{L-1} \tilde{g}_l^2 = 1/2$ , orthogonal to their even shifts

$$\begin{aligned}
\tilde{h}_0 \tilde{h}_2 + \tilde{h}_1 \tilde{h}_3 + \dots + \tilde{h}_{L-3} \tilde{h}_{L-1} &= 0 \tag{52} \\
\tilde{h}_0 \tilde{h}_4 + \tilde{h}_1 \tilde{h}_5 + \dots + \tilde{h}_{L-5} \tilde{h}_{L-1} &= 0 \\
&\vdots \\
\tilde{h}_0 \tilde{h}_{L-2} + \tilde{h}_1 \tilde{h}_{L-1} &= 0
\end{aligned}$$

$$\begin{aligned}
\tilde{g}_0 \tilde{g}_2 + \tilde{g}_1 \tilde{g}_3 + \dots + \tilde{g}_{L-3} \tilde{g}_{L-1} &= 0 \tag{53} \\
\tilde{g}_0 \tilde{g}_4 + \tilde{g}_1 \tilde{g}_5 + \dots + \tilde{g}_{L-5} \tilde{g}_{L-1} &= 0 \\
&\vdots \\
\tilde{g}_0 \tilde{g}_{L-2} + \tilde{g}_1 \tilde{g}_{L-1} &= 0
\end{aligned}$$

and because of the quadrature mirror filter  $\tilde{g}_l = (-1)^{l+1}\tilde{h}_{L-1-l}$ ,

$$\begin{aligned}
(\tilde{h}_0\tilde{h}_1 + \tilde{g}_0\tilde{g}_1) + (\tilde{h}_1\tilde{h}_2 + \tilde{g}_1\tilde{g}_2) + \dots + (\tilde{h}_{L-2}\tilde{h}_{L-1} + \tilde{g}_{L-2}\tilde{g}_{L-1}) &= 0 \\
(\tilde{h}_0\tilde{h}_3 + \tilde{g}_0\tilde{g}_3) + (\tilde{h}_1\tilde{h}_4 + \tilde{g}_1\tilde{g}_4) + \dots + (\tilde{h}_{L-4}\tilde{h}_{L-1} + \tilde{g}_{L-4}\tilde{g}_{L-1}) &= 0 \\
&\vdots \\
(\tilde{h}_0\tilde{h}_{L-1} + \tilde{g}_0\tilde{g}_{L-1}) &= 0
\end{aligned} \tag{54}$$

Placing the restrictions implied by Equations (52 - 54), together with  $\sum_{l=0}^{L-1} \tilde{h}_l^2 = 1/2$ ,  $\sum_{l=0}^{L-1} \tilde{g}_l^2 = 1/2$ , into Equation (51) leads to

$$\begin{aligned}
\sum_{t=1}^T (\tilde{W}_{t,1}^2 + \tilde{V}_{t,1}^2) &= T\gamma_0 \\
&= \sum_{t=1}^T y_t^2
\end{aligned} \tag{55}$$

**Proof of Theorem 2.8:** Let  $\gamma_j = T^{-1} \sum_{t=1}^T y_{t \bmod T} y_{t-j \bmod T}$ .

$$T^{-1} \sum_{t=1}^T \left( \widetilde{W}_{t,1}^2 \right) = \gamma_0 \sum_{l=0}^{L-1} \tilde{h}_l^2 + 2\gamma_1 \sum_{l=0}^{L-2} (\tilde{h}_l \tilde{h}_{l+1}) + 2\gamma_2 \sum_{l=0}^{L-3} (\tilde{h}_l \tilde{h}_{l+2}) + 2\gamma_3 \sum_{l=0}^{L-4} (\tilde{h}_l \tilde{h}_{l+3}) + \dots + 2\gamma_{L-1} (\tilde{h}_0 \tilde{h}_{L-1}).$$

because  $\{\tilde{h}_l\}_{l=0}^{L-1}$  is orthogonal to its even shifts and  $\sum_{l=0}^{L-1} \tilde{h}_l^2 = 1/2$

$$\begin{aligned} \tilde{h}_0 \tilde{h}_2 + \tilde{h}_1 \tilde{h}_3 + \dots + \tilde{h}_{L-3} \tilde{h}_{L-1} &= 0 \\ \tilde{h}_0 \tilde{h}_4 + \tilde{h}_1 \tilde{h}_5 + \dots + \tilde{h}_{L-5} \tilde{h}_{L-1} &= 0 \\ &\vdots \\ \tilde{h}_0 \tilde{h}_{L-2} + \tilde{h}_1 \tilde{h}_{L-1} &= 0 \end{aligned}$$

so that

$$T^{-1} \sum_{t=1}^{T/2} \left( \widetilde{W}_{t,1}^2 \right) = \gamma_0 + 2\gamma_1 \sum_{l=0}^{L-2} (\tilde{h}_l \tilde{h}_{l+1}) + 2\gamma_3 \sum_{l=0}^{L-4} (\tilde{h}_l \tilde{h}_{l+3}) + 2\gamma_5 \sum_{l=0}^{L-6} (\tilde{h}_l \tilde{h}_{l+5}) + \dots + 2\gamma_{L-1} (\tilde{h}_0 \tilde{h}_{L-1})$$

By using Lemma (2.7),  $\tilde{G}_{T,1}^L$  is written as

$$\begin{aligned} \tilde{G}_{T,1}^L &= \frac{\sum_{t=1}^T \widetilde{W}_{t,1}^2}{\sum_{t=1}^T \widetilde{V}_{t,1}^2 + \sum_{t=1}^T \widetilde{W}_{t,1}^2} \\ &= \frac{\sum_{t=1}^T y_t^2 + 2T \left( \gamma_1 \sum_{l=0}^{L-2} (\tilde{h}_l \tilde{h}_{l+1}) + \gamma_3 \sum_{l=0}^{L-4} (\tilde{h}_l \tilde{h}_{l+3}) + \gamma_5 \sum_{l=0}^{L-6} (\tilde{h}_l \tilde{h}_{l+5}) + \dots + \gamma_{L-1} (\tilde{h}_0 \tilde{h}_{L-1}) \right)}{(1/2) \sum_{t=1}^T y_t^2 + (1/2) \sum_{t=1}^T y_{t-1}^2} \\ \tilde{G}_{T,1}^L - \frac{1}{2} &= \frac{2T \left( \gamma_1 \sum_{l=0}^{L-2} (\tilde{h}_l \tilde{h}_{l+1}) + \gamma_3 \sum_{l=0}^{L-4} (\tilde{h}_l \tilde{h}_{l+3}) + \gamma_5 \sum_{l=0}^{L-6} (\tilde{h}_l \tilde{h}_{l+5}) + \dots + \gamma_{L-1} (\tilde{h}_0 \tilde{h}_{L-1}) \right)}{(1/2) \sum_{t=1}^T y_t^2 + (1/2) \sum_{t=1}^T y_{t-1}^2} \\ &= \frac{2N(0, \sigma^4 T) \left( \sum_{l=0}^{L-2} (\tilde{h}_l \tilde{h}_{l+1}) + \sum_{l=0}^{L-4} (\tilde{h}_l \tilde{h}_{l+3}) + \sum_{l=0}^{L-6} (\tilde{h}_l \tilde{h}_{l+5}) + \dots + (\tilde{h}_0 \tilde{h}_{L-1}) \right)}{(1/2) \sum_{t=1}^T y_t^2 + (1/2) \sum_{t=1}^T y_{t-1}^2} \\ &= -\frac{2 \left( \sum_{l=1}^{L-(L/2)} \tilde{h}_{2l-1} \right)^2 N(0, \sigma^4 T)}{\sigma^2 T} + o_p(1) \\ &= -\frac{2N(0, 1) \left( \sum_{l=1}^{L-(L/2)} \tilde{h}_{2l-1} \right)^2}{\sqrt{T}} + o_p(1) \end{aligned}$$

Since  $\left(\sum_{l=1}^{L-(L/2)} \tilde{h}_{2l-1}\right)^2 = 1/4$ ,  $\sqrt{4T} \left(\tilde{G}_{T,1}^L - \frac{1}{2}\right) = N(0, 1) + o_p(1)$ .

## References

- Andrews, D. W. K. and Ploberger, W. (1996). Testing for serial correlation against an ARMA(1,1) process. *Journal of the American Statistical Association*, **91**, 1331–1342.
- Box, G. and Pierce, D. A. (1970). Distribution of residual autocorrelations in autoregressive integrated moving average time series models. *Journal of the American Statistical Association*, **65**, 1509–1526.
- Breusch, T. (1978). Testing for autocorrelation in dynamic linear models. *Australian Economic Papers*, **17**, 334–355.
- Daubechies, I. (1992). *Ten Lectures on Wavelets*, volume 61 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. SIAM, Philadelphia.
- Duchesne, P. (2006). On testing for serial correlation with a wavelet-based spectral density estimator in multivariate time series. *Econometric Theory*, **22**, 633–676.
- Duchesne, P., Li, L., and Vandermeerschen, J. (2010). On testing for serial correlation of unknown form using wavelet thresholding. *Computational Statistics & Data Analysis*, **54**, 2512–2531.
- Durbin, J. and Watson, G. S. (1950). Testing serial correlation in least squares regression: I. *Biometrika*, **37**, 409–428.
- Durbin, J. and Watson, G. S. (1951). Testing serial correlation in least squares regression: II. *Biometrika*, **38**, 159–177.
- Durbin, J. and Watson, G. S. (1971). Testing serial correlation in least squares regression: III. *Biometrika*, **58**, 1–19.
- Escanciano, J. C. and Lobato, I. N. (2009). An automatic portmanteau test for serial correlation. *Journal of Econometrics*, **151**(2), 140–149.
- Fan, Y. and Gençay, R. (2010). Unit root tests with wavelets. *Econometric Theory*, **26**, 1305–1331.
- Gençay, R., Selçuk, F., and Whitcher, B. (2001). *An introduction to Wavelets and Other Filtering Methods in Finance and Economics*. Academic Press, San Diego.
- Godfrey, L. G. (1978). Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables. *Econometrica*, **46**, 1293–1302.
- Godfrey, L. G. (2007). Alternative approaches to implementing lagrange multiplier tests for serial correlation in dynamic regression models. *Computational Statistics & Data Analysis*, **51**, 3282–3295.

- Hong, Y. (1996). Consistent testing for serial correlation of unknown form. *Econometrica*, **64**, 837–864.
- Hong, Y. and Kao, C. (2004). Wavelet-based testing for serial correlation of unknown form in panel models. *Econometrica*, **72**, 1519–1563.
- Lee, J. and Hong, Y. (2001). Testing for serial correlation of unknown form using wavelet methods. *Econometric Theory*, **17**, 386–423.
- Ljung, G. M. and Box, G. (1978). On a measure of lack of fit in time series. *Biometrika*, **65**, 297–303.
- Lobato, I., Nankervis, J. C., and Savin, N. (2002). Testing for zero autocorrelation in the presence of statistical dependence. *Econometric Theory*, **18**(03), 730–743.
- Papadoditis, E. (2000). Spectral density based goodness-of-fit tests for time series analysis. *Scandinavian Journal of Statistics*, **27**, 143–176.
- Percival, D. B. and Walden, A. T. (2000). *Wavelet Methods for Time Series Analysis*. Cambridge Press, Cambridge.
- Robinson, P. M. (1991). Testing strong serial correlation and dynamic conditional heteroskedasticity in multiple regression. *Journal of Econometrics*, **47**, 67–84.
- Xue, Y. and Gençay, R. (2010). Testing for jump arrivals in financial time series. Technical report, Department of Economics, Simon Fraser University.
- Yule, U. (1926). Why do we sometimes get nonsense-correlations between time-series? *Journal of the Royal Statistical Society*, **89**, 1–63.