MULTI-SCALE TESTS FOR SERIAL CORRELATION

RAMAZAN GENÇAY AND DANIELE SIGNORI

November 5, 2013.

Abstract

This paper introduces a new family of portmanteau tests for serial correlation. Using the wavelet transform, we decompose the variance of the underlying process into the variance of its low frequency and of its high frequency components and we design a variance ratio test of no serial correlation in the presence of dependence. Such decomposition can be carried out iteratively, each wavelet filter leading to a rich family of tests whose joint limiting null distribution is a multivariate normal. We illustrate the size and power properties of the proposed tests through Monte Carlo simulations.

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

JEL Classification Numbers: C1, C2, C12, C22, C58, F31, G0, G1.

We are thankful to Joon Park, Don Andrews, Mirza Trokić, Bertille Antoine, Chris Muris, Marco Rossi, Ken Kasa, Stephen Fagan, Marco Rossi, the participants of Canadian Econometric Study Group (2012), and the SFU Brown Bag Seminar for helpful comments. Both authors are the 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 (R. Gençay), dsignori@sfu.ca (D. Signori).

Multi-scale tests for serial correlation

Abstract. This paper introduces a new family of portmanteau tests for serial correlation. Using the wavelet transform, we decompose the variance of the underlying process into the variance of its low frequency and of its high frequency components and we design a variance ratio test of no serial correlation in the presence of dependence. Such decomposition can be carried out iteratively, each wavelet filter leading to a rich family of tests whose joint limiting null distribution is a multivariate normal. We illustrate the size and power properties of the proposed tests through Monte Carlo simulations.

I. Introduction

This paper proposes a new family of frequency-domain tests for the white noise hypothesis, the assumption that a process is uncorrelated. Frequency-domain tests take as their starting point the result that, under stationarity conditions, the linear dependence structure of a process $\{y_t\}$ is fully captured by its spectral density function $S_y(f)$. We focus our attention on the relation between the spectral density function and the variance,

$$\mathrm{var}(y) = 2\int_0^{1/2} S_y(f)\, df \ ,$$

which, paraphrasing, says that the contribution of the frequencies in a small interval Δf containing f is approximately $S_y(f)\Delta f$. It is an elementary result that—when defined—the spectral density function of an uncorrelated process is constant or, in other words, that each frequency contributes equally to the variance of a white noise process; instead, when a process is serially correlated, each frequency generally contributes in different amounts and the spectral density function is non-constant.

Such contrast is the basis for the tests developed in this paper. Imagine that $\{y_t\}$ is a gaussian white noise process (Fig. 1, left panel). Then high frequencies, say those in the band $[1/4, 1/2]$, will contribute exactly half of the total variance of $\{y_t\}$. On the other hand, if $\{y_t\}$ is an autoregressive process of order 1 with a positive coefficient (right panel), high frequencies will account for less than half of the total variance. This example motivates the introduction of the variance ratio $\mathcal{E}(a, b)$, defined as the ratio of the total variance contributed by the frequency band (a, b). Under the null of no serial correlation, $\mathcal{E}(a, b)$ is equal to the length of the interval (a, b) and any departure from this benchmark provides the means to detect serial correlation.

Although the variance ratio can be defined for an arbitrary frequency domain, the need to estimate the corresponding integral of the spectral density function—the numerator of $\mathcal{E}$—imposes practical limitations. We resort to wavelet analysis to address this need. For frequency bands of a particular form, the numerator of the statistic $\mathcal{E}$ is a well known quantity, the wavelet variance,[1] which can be estimated efficiently using the maximum-overlap discrete wavelet transformation estimator. In this light, given the temporal resolution properties of the wavelet transform, it is

[1] The wavelet variance was studied, among others, by Allan (1966), Percival (1983), Percival and Guttorp (1994), Percival (1995), and Howe and Percival (1995).

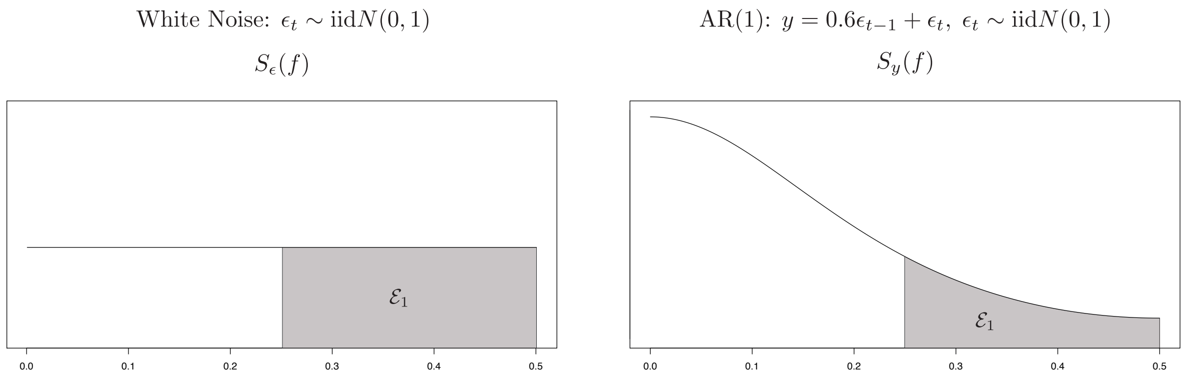

FIGURE 1. High frequency contribution (in grey) to the total variance of a white noise process (left) and an AR(1) process (right).

appropriate to refer to $\mathcal{E}(a, b)$ as a multiscale variance ratio. The recursive application of this procedure generates a family of tests whose joint limit distribution is multivariate normal under mild restrictions.

While the main intuition behind multiscale variance ratios originates under covariance stationarity assumptions, the corresponding test statistics is informative in more general scenarios. Indeed, the null hypothesis can be relaxed to allow for a degree of non-stationarity, specifically, for heteroskedastic white noise. Heteroskedastic white noise is an uncorrelated process with varying variance. We develop the asymptotic theory of multiscale variance ratios for uncorrelated but possibly dependent processes within the framework of *near-epoch dependence* (NED). Besides accommodating heterogeneity, there are three further benefits of this approach. Firstly, the asymptotic results originate from one of the most general gaussian central limit theorems for dependent processes (De Jong, 1997). Secondly, it permits trending higher moments (see Assumption A and Assumption B1). Finally, it leverages a rich literature devoted to the derivation of the NED property for many nonlinear time series models and, thus, parametric restrictions for the validity of our test can be obtained in several typical cases.[2]

We contribute to the literature on tests for serial correlations in several ways. First, the design we propose leads to serial correlation tests with desirable empirical size and power in small samples. Second, as argued in the previous paragraph, our test is robust to the presence of higher order dependence, heteroskedasticity, and trending moments, while at the same time the asymptotic theory is developed in great generality. Third, our is the first test of serial correlation that utilized directly the wavelet coefficients of the observed time series to construct the wavelet-based test statistics.[3] The tests we design generalize, on one hand, variance ratios tests (Lo and MacKinlay, 1988), on the other, they are related to ratios of quadratic forms and Von Neumann ratios (1941). In addition, since the proposed test statistic does not rely on a point estimate of the spectral

[2]These results include GARCH, IGARCH, FIGARCH, ARCH(∞) (Davidson, 2004), ARMA, Bilinear models, switching and threshold autoregressive models, and smooth nonlinear autoregressions. (Davidson, 2002).

[3]This approach was originally put forth by Fan and Gençay (2010) in unit root testing. Within a similar framework, Xue et al. (2010) propose discrete wavelet-based jump tests to detect jump arrival times in high frequency financial time series data.

density, the rate of convergence issues relating to the nonparametric spectral density are not of first order of importance.

One of the well-known time-domain portmanteau tests for serial correlation is the Box and Pierce's test Q_K (BP). Given independent and identically distributed observations, Box and Pierce (1970) show that the sum of K sample autocovariances times the number of observation is approximately distributed as a Chi-squared distribution with K degrees of freedom; statistically large values of Q_K indicate a likely serial correlation among the data. In practice, the strict restriction of independence and homogeneity are violated, leading to possibly very inaccurate inference. There is a long streak of papers that address these limitations, starting from the small sample improvements of Ljung and Box (1978), to the more recent robustification program of Lobato (2001) and Lobato, Nankervis, and Savin (2002). Robust inference can also be achieved using bootstrapping methods. Building on the block bootstrap inference for autocorrelations of Romano and Thombs (1996), Horowitz, Lobato, Nankervis, and Savin (2006) develop a blocks-of-blocks bootstrap that reduces the error rejection probability to nearly zero for samples with at least 500 observations. Finally, Escanciano and Lobato (2009) (EL) combine robustification techniques with a data-driven approach for automatic lag selection. The resulting adaptive test has particularly high empirical power in finite samples.

Frequency-domain tests provide an alternative framework for test of serial correlation. Hong (1996) uses a 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. 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. 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 parameter 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.

In Section **II**, we fix the notation, describe the discrete wavelet transform, and present the concept of near-epoch dependence together with the law of large numbers and the central limit theorem from which our main results will obtain. In Section **III**, we introduce and motivate our tests. In Section **IV** we study its large sample distribution. In Section **V**, we analyze the small sample properties through several Monte Carlo simulations. A brief conclusion follows afterwards.

II. Preliminaries

Let y_t be a stochastic sequence with $E(y_t) = 0$ and $\mathrm{var}(y_t) = \sigma_t^2$. If y_t is homoskedastic, that is $\sigma_t^2 = \sigma^2$ for all t, and uncorrelated, that is $\mathrm{cov}(y_t, y_s) = 0$ for all $s \neq t$, then y_t is called *white noise*. If homoskedasticity is violated, we refer to y_t as *heteroskedastic white noise*. We consider tests of the null hypothesis of no correlation, $H_0 : \mathrm{cov}(y_t, y_s) = 0$ for all $s \neq t$, against correlated alternatives, $H_1 : \mathrm{cov}(y_t, y_s) \neq 0$ for some $s \neq t$. A finite sample realization of y_t with T observation is denoted with $\{y_t\}$ and, when viewed as a vector in $\mathbb{R}^T$, we use the notation $\boldsymbol{y}^T$, or simply $\boldsymbol{y}$, leaving T understood when there is no chance for confusion. Throughout the paper we impose periodic boundary conditions on $\{y_t\}$, that is

$$y_t \equiv y_{t \bmod T} \ ,$$ [4]

and we define $s_n^2(y)$ as

$$s_n^2(y) = \sum_{t=1}^{n} \mathrm{var}(y_t) + 2 \sum_{t=2}^{n} \sum_{k=1}^{n-1} \mathrm{cov}(y_t, y_{t-k}) \ . \tag{1}$$

A stochastic sequence y_t gives rise to a filtration of sigma fields

$$\mathcal{F}_{t-m}^{t+m}(x) \equiv \sigma(x_{t-m}, \ldots, x_{t+m}) \ ,$$

where $\mathcal{F}_{t-m}^{t+m}(x)$ is the smallest sigma field on which $\{x_{t-m}, \ldots, x_{t+m}\}$ are measurable, that is the collection of sets of the form $x_i^{-1}(B)$ where B is a measurable set in the codomain of x_i and the index i ranges from $t-m$ to $t+m$. Either bounds can be let go to infinity, yielding the sigma fields $\mathcal{F}_{-\infty}^{t}$—containing the information from the remote past up to now—and $\mathcal{F}_{t}^{\infty}$—containing the information from the present to the remote future. When there is no risk of confusion, we will write $\mathcal{F}_{t-m}^{t+m}$ for $\mathcal{F}_{t-m}^{t+m}(x)$. All proofs can be found in the Appendix.

In developing the statistical properties of our test for serial correlation, we consider a very general null hypothesis, namely that the data generating process is heteroskedastic white noise, thus restricting only the correlation properties of the process while leaving higher order dependence completely unconstrained. In order to remain close to the intention of a very general null hypothesis, we develop the asymptotic theory for our serial correlation test in terms of concept of *near-epoch dependence* (NED). For a stochastic sequence x_t define

$$\alpha_m \equiv \sup_{i\in\mathbb{Z}} \sup_{\{A\in\mathcal{F}_{-\infty}^{t}, B\in\mathcal{F}_{t}^{\infty}\}} |P(A \cup B) - P(A)P(B)|$$

$$\phi_m \equiv \sup_{i\in\mathbb{Z}} \sup_{\{A\in\mathcal{F}_{-\infty}^{t}, B\in\mathcal{F}_{t}^{\infty}, P(A)>0\}} |P(B|A) - P(B)| \ .$$

Then, if $\phi_m = o(m^{-a-\varepsilon})$ for $\varepsilon > 0$, then x_t is *ϕ-mixing of size $-a$*. If $\alpha_m = o(m^{-a-\varepsilon})$ for $\varepsilon > 0$, then x_t is *α-mixing of size $-a$*.

[4]The notation $a-b \bmod T$ 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 $1 \leq j+nT \leq T$.

Definition 1 (Adapted from Davidson (1995), Definition 17.1, page 261)**.** *A stochastic sequence x_t is said to be near-epoch dependent on ϵ_t in L_p-norm for $p > 0$ if*

$$\|x_t - \mathbb{E}[x_t|\mathcal{F}_{t-m}^{t+m}(\epsilon)]\|_p \leq d_t \nu_m \tag{2}$$

where $\nu_m \to 0$ as $m \to \infty$ and d_n is a sequence of positive real numbers such that $d_t = O(\|x_t\|_p)$.[5]

Any process x_t satisfies Definition 1 will be referred to as "L_p-NED on ϵ_t" for short. The concept of near-epoch dependence was popularized in the econometrics literature by Gallant and White (1988), but its inception can be traced back to the work of Ibragimov (1962). As pointed out by Davidson (1995), near-epoch dependence is not an alternative to mixing assumptions, instead it allows to establish useful memory properties of x_t in terms of those of ϵ_t.

When the innovation process ϵ_t is mixing, powerful laws of large numbers and central limit theorems can be established for NED processes.[6] In order to apply these results, the following proposition will be useful (a generalization of Theorem 17.9 in Davidson, 1995, from L_2 to L_p processes).

Proposition 2. *If x_t and y_t be L_p-NED on $\{\epsilon_t\}$ of size $-\phi_x$ and $-\phi_y$ respectively, then $x_t y_y$ is $L_{p/2}$-NED of size $-\min(\phi_x, \phi_y)$ on $\{\epsilon_t\}$.*

A. Wavelet Transformations

In this section we introduce the Maximum Overlap Discrete Wavelet Transform (MODWT).[7] A vector $\{h_l\} = (h_0, \dots, h_{L-1})$ in $\mathbb{R}^L$ gives rise to a linear time invariant filter by means of the convolution operation: Given a sequence to be filtered $\{y_t\}$, the convolution of $\{h_l\}$, and $\{y_t\}$ is the sequence

$$h * y_t = \sum_{l=-\infty}^{l=\infty} h_l y_{t-l} \ , \quad \forall t \ ,$$

where we define $h_l = 0$ for all $l < 0$ and $l \leq L$.

A *wavelet filter* is a linear time invariant filter $\{h_l\}$ of length L, such that for all $n \neq 0$:

$$\sum_{l=0}^{L-1} h_l = 0 \ , \quad \sum_{l=0}^{L-1} h_l^2 = 1/2 \ , \quad \sum_{l=-\infty}^{\infty} h_l h_{l+2n} = 0 \ . \tag{3}$$

In words, h sums to zero, has norm $1/2$, and is orthogonal to its even shifts. The natural complement to the wavelet filter $\{h_l\}$ is the *scaling filter* $\{g_l\}$ determined by the *quadrature mirror relationship*

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

[5] The sequence d_t is a technical device used to accommodate trending moments. For all the data generating processes encountered in the examples, it can be set equal to 1.

[6] See, among others, Davidson (1992, 1993, 1995); De Jong (1997).

[7] This section closely follows Gençay et al. (2001), see also (Percival and Walden, 2000, Chap. 5).It is common in the literature distinguish the objects related Discrete Wavelet Transform from those related to the Maximum Overlap Discrete Wavelet Transform by placing a tilde ($\sim$) in the latter case. Since all quantities in the main part of the paper refer to the MODWT and we believe there is little scope for confusion, we warn the reader that in this paper we do not follow this convention.

The scaling filter satisfies the following basic properties, analogous to Equations 3:

$$\sum_{l=0}^{L-1} g_l = 1 \ , \quad \sum_{l=0}^{L-1} g_l^2 = 1/2 \ , \quad \sum_{l=-\infty}^{\infty} g_l g_{l+2n} = 0 \ , \quad \sum_{l=-\infty}^{\infty} g_l h_{l+2n} = 0 \ , \tag{4}$$

for all nonzero integers n.

In general, the definitions of wavelet and scaling filter do not imply any specific band-pass properties (see Percival and Walden, 2000, Chap. 4, Pag. 105, for an in-depth discussion). Further conditions must be imposed to recover the domain frequency interpretation associated with the continuous wavelet transform and to guarantee that $\{h_l\}$ is a high-pass filter (which, as a consequence of the QMF relationship, implies that $\{g_l\}$ is a low-pass filter). An example of such additional constrains, sometimes referred to as *regularity conditions*, are the vanishing moment conditions introduced by Daubechies (1993). Nevertheless, all the results in the paper hold without any regularity conditions on the filters and hence to any arbitrary dyadic band-pass decomposition. In particular, when the filters $\{h_l\}$ and $\{g_l\}$ applied to an observed time series are from a wavelet filter-bank, we can separate high-frequency oscillations from low-frequency ones.

Formally, the MODWT of level M is a linear operator and can be represented in terms of matrix operations:

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

where $\mathcal{W}$ is a $(M+1)T \times T$ matrix. The matrix $\mathcal{W}$ is constructed by assembling $M+1$ sub-matrices of dimensions $T \times T$:

$$\mathcal{W} = [\mathcal{W}_1, \mathcal{W}_2, \cdots, \mathcal{W}_M, \mathcal{V}_M]' \ ,$$

whose action is defined in terms of wavelet filter $\{h_l\}$ and scaling filter $\{g_l\}$. Specifically,

$$(\mathcal{W}_m \boldsymbol{y})_t = \sum_{l=0}^{L_m} h_{m,l} v_{m,t-l \bmod T}$$

where $L_m := (2^m - 1)(L-1) + 1$. The m-th level filter $\{h_{m,l}\}$ can be written as a filter cascade

$$h_m = h * \underbrace{g * \ldots * g}_{m-1} \ ,$$

where g is the scaling filter and $*$ denotes a convolution.[8]

The MODWT of the observed time series $\boldsymbol{y}^T$ can be organized into $M+1$ vectors of length T

$$\boldsymbol{w} = (\boldsymbol{w}'_1, \ldots, \boldsymbol{w}'_M, \boldsymbol{v}'_M)' \ , \tag{5}$$

[8]A general explicit formula for h_m requires working with transfer functions in Fourier space

$$h_m(l) = \frac{1}{L} \sum_{f=0}^{L-1} H\left(\frac{2^{m-1} f}{N}\right) \prod_{k=1}^{m-2} G\left(\frac{2^k f}{N}\right) \mathrm{e}^{2ifl\pi/L}$$

where H and G are the Discrete Fourier Transforms of h and g, respectively:

$$H(f) = \sum_{l=0}^{L-1} h_l \mathrm{e}^{2ifl\pi/L} \ , \quad G(f) = \sum_{l=0}^{L-1} g_l \mathrm{e}^{2ifl\pi/L} \ .$$

where $M \leq \log_2 T$ be the decomposition level of the MODWT. In practice, $\boldsymbol{w}$ is computed recursively via a so-called pyramid algorithm. Each iteration of the MODWT pyramid algorithm, requires three objects: the data vector to be filtered, the wavelet filter $\{h_l\}$ and the scaling filter $\{g_l\}$. The initial step consists of applying the wavelet and scaling filter to the data with each filter to obtain the first level wavelet and scaling coefficients:

$$w_{1,t} = (\boldsymbol{w}_1)_t = \sum_{l=0}^{L-1} h_l y_{t-l \bmod T} \quad \text{and} \quad v_{1,t} = (\boldsymbol{v}_1)_t = \sum_{l=0}^{L-1} g_l y_{t-l \bmod T} \quad \text{for all} \quad t = 1, \ldots, T \ .$$

The length T vector of observations has been high- and low-pass filtered to obtain T coefficients associated with this information. The m-th step consists of applying the filtering operations as above to obtain the $(m+1)$-st level of wavelet and scaling coefficients

(6)

$$w_{m+1,t} = (\boldsymbol{w}_1)_t = \sum_{l=0}^{L-1} h_l v_{m,t-l \bmod T} \quad \text{and} \quad v_{m+1,t} = (\boldsymbol{v}_1)_t = \sum_{l=0}^{L-1} g_l v_{m,t-l \bmod T} \quad \text{for all} \quad t = 1, \ldots, T \ .$$

Keeping all vectors of wavelet coefficients, and the level M scaling coefficients, we obtain the decomposition of Equation 5.

III. Multi-scale Variance Ratios

Consider the general variance ratio

$$\mathcal{E}(a,b) = 2 \int_a^b S_y(f)\, df \Big/ \mathrm{var}(y) \ .$$

The numerator of $\mathcal{E}(a,b)$ can, for specific intervals, be expressed in terms of the wavelet variance. Indeed, neglecting the leakage of the wavelet filter, the following approximation holds[9]

$$\mathrm{wvar}_m(y) \approx 2 \int_{1/2^{j+1}}^{1/2^j} S_y(f)\, df \ . \tag{7}$$

For $m = 1$, the integral in Equation (7) corresponds to the area $\mathcal{E}_1$ in Figure 1. Formally, the wavelet variance for a stationary process y is defined as

$$\mathrm{wvar}_m(y) \equiv \mathrm{var}(w_{m,t}) \ . \tag{8}$$

From equation (6), we see that $w_{m,t}$ is a linear process, obtained by applying the time invariant filter h_m to a zero mean process y. If y is stationary, then the spectrum of $w_{m,t}$ is $S_m(f) = |H_m(f)|^2 S_y(f)$, where $H_m(f)$ is the discrete Fourier transform of the filter $\{h_i\}$ (see Brockwell and Davis, 2009, Page 121, Eq. 4.4.3.). If follows that

$$\mathrm{wvar}_m(y) = \int_{-1/2}^{1/2} S_m(f)\, df = \int_{-1/2}^{1/2} |H_m(f)|^2 S_y(f)\, df \tag{9}$$

[9] See Percival and Walden (2000), Equation (297a), page 297.

In particular, if $\{y_t\}$ is a covariance stationary white noise, then $S_y(f) = \sigma_y^2$ and

$$\begin{aligned}\mathrm{wvar}_m(y) &= \sigma_y^2 \int_{-1/2}^{1/2} |H_m(f)|^2\, df = \sigma_y^2 \|h_m\|_2 \\ &= \sigma_y^2 \|g\|_2 \prod_{i=1}^{m-1} \|h\|_2 = \sigma_y^2 2^{-m}\end{aligned}$$

The second equality uses Parseval's identity, the third equality holds because the norm of a convolution is the product of the norms, and the last equality follows from the normalization Equation (3). In conclusion, we proved the following

Theorem 3. *The wavelet variance ratio for a stationary white noise process is*

$$\mathcal{E}_m(y) \equiv \frac{\mathrm{wvar}_m(y)}{\mathrm{var}(y)} = \frac{1}{2^m}\ .$$

When there is no risk of confusion, we will write $\mathcal{E}_m$ for $\mathcal{E}_m(y)$. In the reminder of this section we introduce a family of statistics that detect serial correlation by testing the implications of Theorem 3.

A. Sample Multiscale Variance Ratios: Scale One

The Maximum Overlap Discrete Wavelet Transform (MODWT) consists of a set of linear filters that given a time series generates a collection of vectors. The design of the MODWT filters are such that each of the resulting vectors contains the characteristics of the original time series corresponding to a specific time-scale.[10]

We illustrate the workings of the MODWT and the intuition behind our test with the simple case of a first level decomposition using the Haar filter. Consider the Haar wavelet filter $\{h_l\}_0^1 = (1/2, -1/2)$ and the corresponding scaling filter $\{g_l\}_0^1 = (1/2, 1/2)$. The wavelet and scaling coefficients of a time series $\{y_t\}_{t=1}^T$ are given by

$$w_{t,1} = \frac{1}{2}(y_t - y_{t-1}),\ \ t = 1, 2, \ldots, T, \tag{10}$$

$$v_{t,1} = \frac{1}{2}(y_t + y_{t-1}),\ \ t = 1, 2, \ldots, T. \tag{11}$$

The wavelet coefficients $\{w_{t,1}\}$ capture the behavior of $\{y_t\}$ in the high frequency band $[1/4, 1/2]$, while the scaling coefficients $\{v_{t,1}\}$ capture the behavior of $\{y_t\}$ in the low frequency band $[0, 1/4]$. A sample analogue of $\mathcal{E}_1$ is readily constructed following the analogy principle

$$\hat{\mathcal{E}}_{1,T} = \frac{\widehat{\mathrm{wvar}_1\, y}}{\widehat{\mathrm{var}\, y}} = \frac{\sum_{t=1}^T w_{1,t}^2}{\sum_{t=1}^T y_{1,t}^2}\ . \tag{12}$$

We show (see Theorem 4) that under H_0, $\hat{\mathcal{E}}_{1,T}$ is close to $1/2$, since the numerator is the half of the denominator, while under H_1 the variance ratio $\hat{\mathcal{E}}_{1,T}$, in general, deviates from $1/2$.

[10]The MODWT goes by several names in the literature, such as the stationary DWT by Nason and Silverman (1995) and the translation-invariant DWT by Coifman and Donoho (1995). A detailed treatment of MODWT can be found in Percival and Mofjeld (1997), Percival and Walden (2000) and Gençay et al. (2001).

The definition of the variance ratio $\hat{\mathcal{E}}_{1,T}$ can be applied to the wavelet decomposition obtained from a generic filter wavelet $\{h_i\}$. As before, we expect $\hat{\mathcal{E}}_{1,T}$ to be close to $1/2$ under H_0.

B. Sample Multiscale Variance Ratios: Scale m

The intuitive results that we discussed above can be generalized to arbitrary scales. For a white noise process, variance is asymptotically equi-partitioned in Fourier space: each frequency contributes an equal share to the total variance of the process. An analogous result holds in "wavelet space": the variance at scale m contributes a ratio of 2^{-m} to the total variance. The variance ratio corresponding to the resolution scale m is defined as

$$\hat{\mathcal{E}}_{m,T} = \frac{\widehat{\mathrm{wvar}_m}\, y}{\widehat{\mathrm{var}\, y}} = \frac{\sum_{t=1}^{T} w_{m,t}^2}{\sum_{t=1}^{T} y_{m,t}^2} \,.$$

where $\boldsymbol{w}_m$ are the m-th level wavelet coefficients of $\boldsymbol{y}$.

To formalize the above discussion, we need to prove that $\hat{\mathcal{E}}_{m,T}$ is a consistent estimator of the wavelet variance ratio. Indeed, the next result goes a step further: as the sample multiscale variance ratio well is defined for nonstationary processes, we show that $\widehat{\mathcal{E}}$ converges in probability to 2^{-m} even for (unconditionally) *heteroskedastic white noise* processes, that is uncorrelated processes that may fail to be covariance stationary.

Assumption A. *$\{y_t\}$ is stochastic sequence that is L_r bounded for $r > 2$ and L_p-NED on an α-mixing process for $p \geq 2$.*

Theorem 4. *Let $\{y_t\}$ be a heteroskedastic white noise process with zero mean. Under Assumption A*

$$\hat{\mathcal{E}}_{m,T} \xrightarrow{p} \frac{1}{2^m}$$

Example 5 (GARCH(1,1) with α-mixing innovations, Hansen (1991))**.** Let $\{\epsilon_t\}$ be a α-mixing process and define

$$x_t = \sigma_t \epsilon_t, \quad \sigma_t^2 = \omega + \beta \sigma_{t-1}^2 + \alpha x_{t-1}^2$$

for some real numbers ω, β, and α. Hansen (1991) shows that if

$$\left(\mathbb{E}\left[\left(\beta + \alpha \epsilon_t^2\right)^p |\mathcal{F}_{-\infty}^{t-1}(\epsilon)\right]\right)^{1/r} \leq c < 1 \quad \text{a.s. for all } t, \tag{13}$$

then $\{x_t, \sigma_t\}$ is L_r-NED on $\{\epsilon_t\}$ with an exponential decay of NED coefficients. With $p = 2$, the condition (13) is equivalent to

$$\beta^2 + 2\alpha\beta\mu_t^2 + \alpha^2 \mu_t^4 < 1 \quad \text{a.s. for all } t,$$

in which $\mu_t^4 = \mathbb{E}(\epsilon_t^4 | \mathcal{F}_{-\infty}^t)$ is the conditional kurtosis.

Example 6 (ARCH(∞) with i.i.d. innovations, Davidson (2004))**.** Let $\{\epsilon_t\}$ be a i.i.d. process, with zero mean and unit variance, and define:

$$x_t = \sigma_t \epsilon_t, \quad \sigma_t^2 = \omega + \sum_{i=1}^{\infty} \alpha_i x_{t-i}^2 \,.$$

This specification is called ARCH(∞) model. It encompasses several nonlinear time series, including GARCH (Bollerslev, 1986), IGARCH (Engle and Bollerslev, 1986), FIGARCH Baillie et al. (1996). Assume that $\mathbb{E}\epsilon^4$ exists and $\sum_{i=1}^{\infty}\alpha_i < (\mathbb{E}\epsilon^4)^2$. Davidson (2004) shows that if $0 \leq \alpha_i \leq Ci^{-1-\lambda}$ for some $\lambda > \lambda_0$, then x_t is L_2-NED on ϵ_t of size $-\lambda_0$.

Example 7 (Bilinear Model with i.i.d. innovations, Davidson (2002))**.** Consider the following bilinear models

$$x_t = \sum_{j=1}^{p}\alpha_j x_{t-j} + \sum_{j=1}^{m}\beta_j x_{t-j}\epsilon_{t-1} + \sum_{j=1}^{r}\gamma_j\epsilon_{t-j}\ ,$$

This parametric family is referred to as BL$(p, r, m, 1)$ and it is discussed in detail in (Priestley, 1988, Chapter 4). Under the assumption of Davidson (2002) concludes that the covariance stationary $BL(p, r, m, 1)$ is L_2-NED on $\{\epsilon_t\}$ with an exponential decay of NED coefficients. A simple example of bilinear white noise is the process

$$x_t = \beta x_{t-2}\epsilon_{t-1} + \epsilon, \quad \epsilon_t \sim \text{i.i.d(0,1)}\ .$$

It is covariance stationary if $0 < \beta < 1/\sqrt{2}$ (see Granger and Newbold, 1986).

In the next section we study the asymptotic distribution of the wavelet ratio $\hat{\mathcal{E}}_{m,T}$.

IV. Asymptotic Analysis

In the reminder of the paper, the process $\{z_{m,t}\}$ is defined as the cross-product component of the square of each wavelet coefficient

$$z_{m,t} := \sum_{i=0}^{L-1}\sum_{j>i}^{L} h_{m,i}h_{m,j}y_{t-i}y_{t-j}\ .$$

When there is no risk of confusion, we omit the index m. Our next result establishes the asymptotic distribution of the wavelet variance ratio $\hat{\mathcal{E}}_{m,T}$.

Assumption B. *Fix a wavelet filter h_m.*

B1. for $r > 1$ and for all i, j, k, l such that $0 \leq i < j \leq L_m$ and $0 \leq k < l \leq L_m$, $\{y_{t-i}y_{t-j}y_{t-k}y_{t-l}/M_{4,t}\}$ is uniformly L_r-bounded for $r > 1$, where

$$M_{4,t} = \sum_{i=0}^{L_m}\sum_{j>1}^{L_m}\sum_{k=0}^{L_m}\sum_{l>1}^{L_m} h_i h_j h_k h_l \mathbb{E}(y_{t-i}y_{t-j}y_{t-k}y_{t-l});$$

B2. For all positive $i \leq L_m$, $\{y_t y_{t-i}\}$ is a stochastic sequence that is L_r-bounded for $r > 2$ and L_p-NED of size $-1/2$ on a ϕ-mixing process for $p \geq 2$.

B3. $var(z_t) \sim t^{\beta}$ and $s_n^2(z) \sim n^{1+\gamma}$, $\beta \leq \gamma$.

Assumption B imposes very mild restrictions on $\{y_t\}$ and allows for substantial deviation from stationarity. Condition B3 can alternatively be expressed in terms of rate of growth the fourth order cumulants of $\{y_t\}$, we omit the resulting expression as it is not particularly revealing. Condition B1 is infinitesimally stricter than allowing for trending *joint* fourth moments in $\{y_t\}$. Notice that neither B1 nor B2 require finite *joint* fourth moments for $\{y_t\}$ but place no explicit restrictions on the fourth moments $\mathbb{E}y_t^4$. For instance, our asymptotic results are valid under the null of independently (but possibly heterogenously) distributed Student's t shocks with $\nu \geq 3$ degrees of freedom. We discuss GARCH(1,1) processes in detail below (Example 9).

Proposition 8. *Let $\{y_t\}$ be a a heteroskedastic white noise process with zero mean and let*

$$T^{-1}\sum_{t=1}^{T}\mathbb{E}y_t^2 \xrightarrow{p} \sigma^2 < \infty \; .$$

Under Assumption B

$$\sqrt{\frac{T\sigma^4}{4s_T^2(z)}}\left(\hat{\mathcal{E}}_{m,T} - \frac{1}{2^m}\right) \xrightarrow{d} N(0,1) \; ,$$

where $s_T(z)$ is defined in Equation (1).

Proposition 8 suggests the following definition for a test statistics

$$GS_m = \sqrt{\frac{T\sigma^4}{4\,\mathrm{avar}(z)}}\left(\hat{\mathcal{E}}_{m,T} - \frac{1}{2^m}\right) \; ,$$

where $\mathrm{avar}(z)$ is the probability limit of $s_T^2(z)$. To implement the test, generally the asymptotic variance of $\{z_t\}$ needs to be estimated. The asymptotic results considered here extend seamlessly to the case of estimated normalizations (Davidson, 1995, Chapter 25). Generally any estimator from the class of kernel estimators is appropriate.[11]

Example 9 (GARCH(1,1) with α-mixing innovations)**.** Consider again Example 5. A straightforward generalization of of Hansen's computation (1991, *Proof of Theorem 1*, page 185) shows that $\{y_t y_{t-1}\}$ is L_2-NED if and only if condition (13) with $p = 4$ is satisfied. Specifically, $\{y_t y_{t-1}\}$ is L_2-NED whenever

$$\beta^4 + 4\alpha\beta^3\mu_t^2 + 6\alpha^2\beta^2\mu_t^4 + 4\alpha^3\beta\mu_t^6 + \alpha^4\mu_t^8 \leq 1 \quad \text{a.s. for all } t \; ,$$

in which $\mu_t^k = \mathbb{E}[\epsilon^k|\mathcal{F}_{-\infty}^t]$. If $\epsilon_t \sim N(0,1)$ are i.i.d., the condition reads

$$\beta^4 + 4\beta^3\alpha + 18\beta^2\alpha^2 + 60\beta\alpha^3 + 105\alpha^4 \leq 1 \quad \text{a.s. for all } t \; .$$

The solution set of this inequality is depicted in Figure 2.

Estimating the asymptotic variance is not always necessary. If y_t is a white noise whose cross-joint cumulants of order four are zero, the asymptotic variance of test can be computed exactly. More specifically, let $X_t^{ijkl} = (X_{t-i}, X_{t-j}, X_{t-k}, X_{t-l})$ and ξ a vector in $\mathbb{R}^4$ and $M(\xi)$. be the moment generating function X_t^{ijkl}

$$M_t^{ijkl}(\xi) = \mathbb{E}\exp(\xi' X_t^{ijkl})$$

has as coefficients of its Taylor expansion

$$M(\xi) = \sum_a \xi_a\kappa^a + \frac{1}{2!}\sum_{a,b}\xi_a\xi_b\kappa^{ab} + \frac{1}{3!}\sum_{a,b,c}\xi_a\xi_b\xi_c\kappa^{abc} + \cdots$$

[11] See Andrews (1991) for a general theory of kernel estimators. Among several approaches and kernel choices we did not find significance differences pointing to a strong preference for one method over the others.

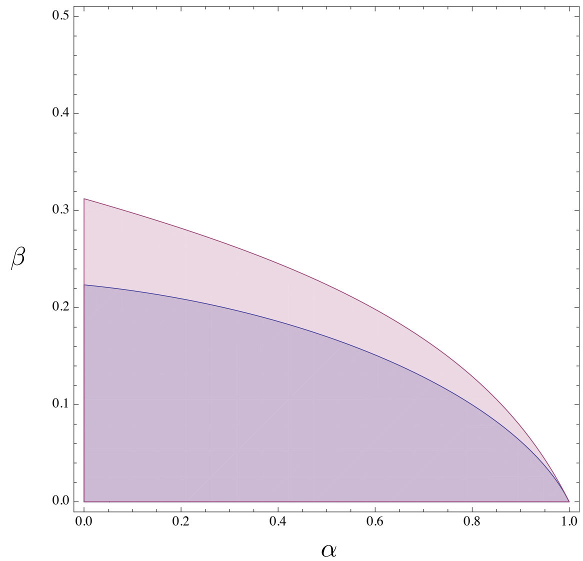

FIGURE 2. Let $\{\epsilon_t\}$ be a identically and independently normally distributed. Let $x_t = \sigma_t\epsilon_t$ and $\sigma_t^2 = \omega + \beta\sigma_{t-1}^2 + \alpha x_{t-1}^2$ for some real numbers ω, β, and α. The pink region depicts the solution to the inequality $\beta^2 + 2\alpha\beta + \alpha^2\mu_t^4 < 1$. In this case x_t satisfies Assumption A. The purple region depicts the solution to the inequality $\beta^4 + 4\beta^3\alpha + 18\beta^2\alpha^2 + 60\beta\alpha^3 + 105\alpha^4 \leq 1$. In this case x_t satisfies Assumption B.

The cumulants of X_t^{ijkl} are defined as the coefficients $\kappa^{(\bullet)}$ in the Taylor expansion

$$\log M(\xi) = \sum_a \xi_a\kappa^a + \frac{1}{2!}\sum_{a,b}\xi_a\xi_b\kappa^{a,b} + \frac{1}{3!}\sum_{a,b,c}\xi_a\xi_b\xi_c\kappa^{a,b,c} + \cdots$$

Notice how commas separating indexes serve to distinguish cumulants from moments when necessary.

Corollary 10. *Let $\{y_t\}$ be white noise process with zero four order cumulants. Then*

$$\sqrt{\frac{T}{a_m}}\left(\hat{\mathcal{E}}_{m,T} - \frac{1}{2^m}\right) \to \mathcal{N}(0,1)$$

with

$$a_m = \sum_{s\in\mathbb{Z}}\sum_{i=i_{min}}^{i_{max}}\sum_{j>i}^{j_{max}} h_{m,i}h_{m,j}h_{m,i-s}h_{m,j-s}\ ,$$

where h_m is the wavelet filter used in the construction of $\hat{\mathcal{E}}_m$ and

$$i_{min} = \max(0,s)\ ,\quad i_{max} = \min(L_m, L_n + s) - 2\ ,\quad j_{max} = \min(L_m, L_n + s) - 1\ .$$

The computation of a_m is trivial but tedious.[12] The following Corollary contains several asymptotic results for the Haar filter.

[12]We implement a routine in a symbolic algebra program to compute both exact and approximate values of a_m for different filters and different resolution scales. The source code is available upon request.

Corollary 11 (Asymptotics for the Haar filter)**.** *Let* $h_1 = \left(\frac{1}{2}, -\frac{1}{2}\right)$ *(the Haar filter). The* GS_m *test statistics for the scales 1 to 4 are*

$$\sqrt{4T}\left(\hat{\mathcal{E}}_{1,T} - \frac{1}{2}\right) , \quad \sqrt{\frac{32T}{3}}\left(\hat{\mathcal{E}}_{2,T} - \frac{1}{4}\right) , \quad \sqrt{\frac{256T}{15}}\left(\hat{\mathcal{E}}_{3,T} - \frac{1}{8}\right) , \quad \sqrt{\frac{2048T}{71}}\left(\hat{\mathcal{E}}_{4,T} - \frac{1}{16}\right) ,$$

respectively. Their asymptotic distribution is the standard normal.

A. Multivariate multiscale tests

Each test in the GS family has a particularly strong power against specific alternatives. For example, for $m = 1$, the test is particularly powerful against AR(1) and MA(1) alternatives, while for $m = 2$, the test has significant power against AR(2) and MA(2) alternatives. In the reminder of this section we derive the asymptotic joint distribution of these tests. These results will allow us to combine these tests to gain power against a wide range of alternatives.

Theorem 12. *Let* $\{y_t\}$ *be a heteroskedastic white noise process with zero mean. Under Assumption B, the vector* $(GS_1, \ldots, GS_N)$ *has asymptotic distribution* $\mathcal{N}(0, \Sigma)$*, where*

$$\Sigma_{i,j} = \frac{\mathrm{acov}(z_i z_j)}{\mathrm{avar}(z_i)\,\mathrm{avar}(z_j)} .$$

Moreover, Large sample inference can be implemented using the test statistics

$$GSM_N = (GS_1, \ldots, GS_N)\Sigma^{-1}(GS_1, \ldots, GS_N)^T ,$$

which is asymptotically distributed as a χ^2_N *distribution.*

The proof of this results follows closely the proof of Proporition 8, we omit it in the interest of space. Large sample inference on the values of the vector $(GS_1, \ldots, GS_N)$ can be handily implemented using the χ^2 distribution. Indeed, it is a standard result (see Bierens, 2004, Theorem 5.9, page 118) that for a multivariate normal n-dimensional vector X and a non-singular $n \times n$ matrix Σ, $X^T \Sigma^{-1} X$ is distributed as a χ^2_n. Accordingly, we define the test statistics

$$GSM_N = (GS_1, \ldots, GS_N)\Sigma^{-1}(GS_1, \ldots, GS_N)^T ,$$

whose asymptotic distribution is a χ^2_N.

As before, if the fourth cumulants of y_t vanish, the asymptotic variance can be computed explicitly as a function of the filters $\{h_m\}$. Let

$$\gamma_{m,n}(s) = \sigma^4 \sum_{i=i_{\min}}^{i_{\max}} \sum_{j \geq i}^{j_{\max}} h_{m,i} h_{m,j} h_{n,l-s} h_{n,k-s}$$

with

$$i_{\min} = \max(0, s) , \quad i_{\max} = \min(L_m, L_n + s) - 2 , \quad j_{\max} = \min(L_m, L_n + s) - 1 .$$

Define, furthermore,

$$a_{m,n} = \frac{1}{\sigma^4} \sum_{s \in \mathbb{Z}} \gamma_{m,n}(s)$$

and let A be an $N \times N$ matrix with ones on the main diagonal and off-diagonal entries

$$A_{mn} = \frac{a_{m,n}}{\sqrt{a_m a_n}}$$

Corollary 13. *The vector* $(GS_1, \ldots, GS_N)$ *has asymptotic distribution* $\mathcal{N}(0, A)$.

In the case of the Haar filter we have:

Corollary 14 (Multi-scale asymptotics for the Haar filter)**.**

$$\begin{pmatrix} GS_1 \\ GS_2 \\ GS_3 \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, A) \ , \quad \textit{with} \quad A = \begin{pmatrix} 1 & -1/\sqrt{6} & -5/\sqrt{60} \\ -1/\sqrt{6} & 1 & 2/\sqrt{360} \\ -5/\sqrt{60} & 2/\sqrt{360} & 1 \end{pmatrix} .$$

V. Asymptotic Local Power and Finite Sample Performance

In this section, we evaluate of the GSM test family generated by the Haar filter using two criteria, namely asymptotic local power and finite sample performance.[13]

First, we illustrate, through an example, the inconsistency of the family GSM_N. Consider the spectrum S_y of the stochastic process y:

$$S_y(f) = \begin{cases} \frac{1}{2} + \frac{1}{4} sin(8\pi f) & \text{if } f \in \left(\frac{1}{4}, \frac{1}{2}\right] \\ \frac{1}{2} + \frac{1}{8} sin(16\pi f) & \text{if } f \in \left(\frac{1}{8}, \frac{1}{4}\right] \\ \ldots \end{cases} \tag{14}$$

The spectrum $S_f(y)$ is shown in Figure 3 and it is non-flat and, hence, the corresponding time series is correlated. At the same time the area underneath S_y within any of the blocks considered by the dyadic decomposition of the frequency space is consistent with the equipartition of variance result valid for white noise processes (Theorems 3 and 4).

For a feasible wavelet filter whose Fourier transform is H, a process x for which the test is inconsistent is one whose spectrum S_x is a solution to the integral equation $H * S_x = S_y$, where $*$ denotes convolution and S_y is given by (14).

At the same time, for any finite ARMA model there is a test in the $\{GSM_N\}$ family which is consistent against it. Recall that the spectrum of a finite ARMA process is a trigonometric rational function in the frequency domain (Theorem 4.4.2 Brockwell and Davis, 2009, page 121):

$$S_y(f) = \frac{P(f)}{Q(f)} \tag{15}$$

where $P(f)$ and $Q(f)$ are trigonometric polynomials. With no loss of generality, assume that $\text{var}(y) = 1$. Let $\mathcal{F}$ be the set of solutions to the equation

$$\frac{P(f)}{Q(f)} - \frac{1}{2} = 0, \tag{16}$$

and let $f_{\min}$ be

$$f_{\min} = \min_{f>0} \{f \in \mathcal{F}\} .$$

[13]Results for other wavelet filters are similar and available from the authors upon request.

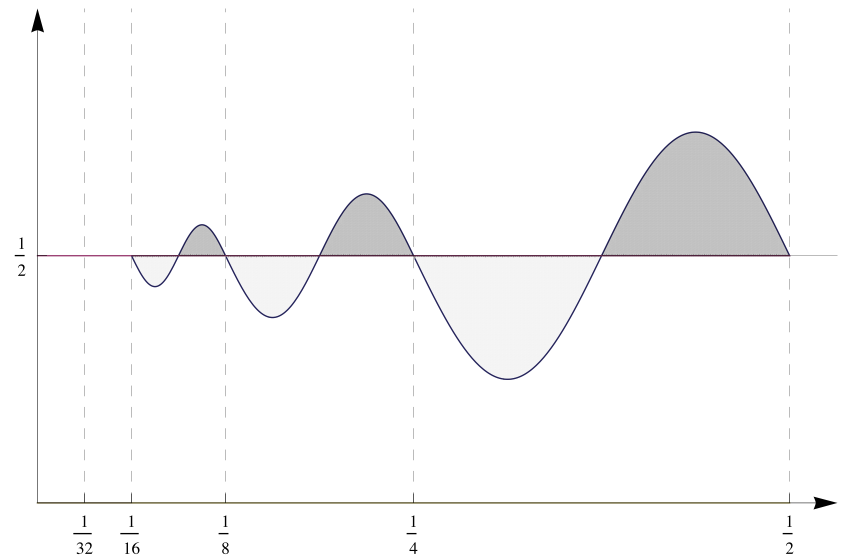

FIGURE 3. Spectrum of an ARMA model of infinite degree. No test of the GSM family is consistent against this alternative.

Since Equation (16) has only a finite number of solutions on a compact set (see Powell, 1981), $f_{\min}$ is well defined and positive. Choose k such that

$$2^{-k-1} < f_{\min} \ ,$$

then the test GS_k is consistent against $H_1 : S_y(f) = \frac{P(f)}{Q(f)}$. Indeed, $S_y(f) > 1/2$ or $S_y(f) < 1/2$ for all f in $(2^{-k-1}, 2^{-k-2})$ and therefore the expected value of GS_k on the process y with spectrum S_y is $\mathbb{E}[GS_k(X_f)] \neq 0$.

A. Asymptotic Local Power

Let $\chi^2_\ell(c)$ denote the non-central χ^2 distribution with non-centrality parameter c and ℓ degrees of freedom. Consider the family of alternative hypothesis

$$H_{1,T}: \ S_T(f) = T^{-1/2}\left(S(f) - \frac{1}{2}\right) + \frac{1}{2}, \tag{17}$$

where $S_(f)$ is a non-constant spectrum. Recall that

$$\mathcal{E}_k = \int_{2^{-k-1}}^{2^k} |H_k(f)|^2 S(f) df$$

and that, in probability, $\widehat{\mathcal{E}}_k \to \mathcal{E}_k$ and, therefore, $GS_k(X) \to \mathcal{E}_k/\mathcal{E}_0 - 1/2^k$. Let

$$\begin{aligned} TGS_N &= \sqrt{T}\mathbb{E}(GS_1(X), \ldots, GS_N(X)) \\ &= (\mathcal{E}_1/\mathcal{E}_0 - 1/2^1, \ldots, \mathcal{E}_N/\mathcal{E}_0 - 1/2^N) \ . \end{aligned}$$

Since the estimator of the covariance matrix of $(GS_1(X), \ldots, GS_N(X))$ is consistent under $H_{1,T}$, it follows that the distribution of the test GS_N is the non-central $\chi^2_N(c)$, where

$$\begin{aligned} c &= T * GSM_N(X) \\ &= TGS'_{1,N} \text{avar}(TGS_{1,N})^{-1} TGS_{1,N} \ . \end{aligned}$$

Therefore the asymptotic local power of GSM_N is given by

$$\Pr(\chi^2_N(c) > \chi^2_{N,1-\alpha}),$$

where $\chi^2_{N,1-\alpha}$ denotes the $(1-\alpha)$-quantile of a χ^2_N distribution.

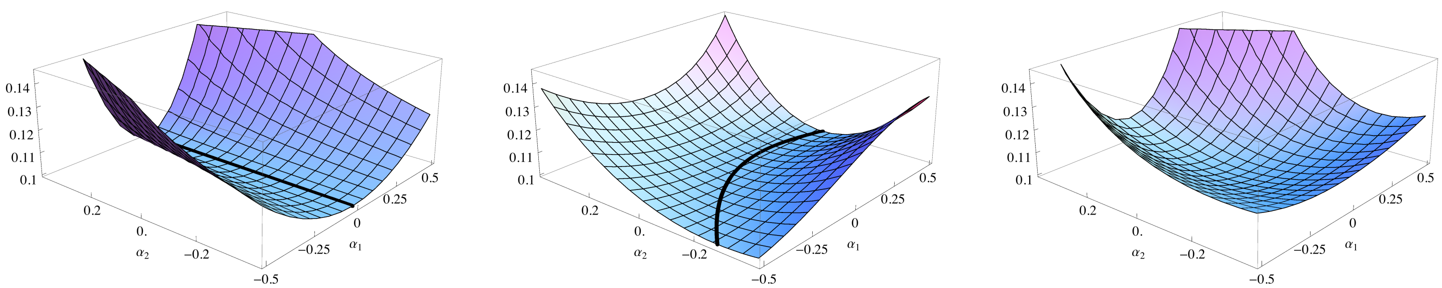

Figure 4. Asymptotic rejection rates at the nominal level $\alpha = 0.10$ against a two-dimensional AR family. The first and second plot (left and center, respectively) depict the asymptotic rejection rate of the one dimensional tests GS_1 and GS_2 together with their 0.10 level (in black). The third plot (right) shows the asymptotic power of the bivariate test GSM_2: in this case the 0.10 level is only one point, corresponding to $\alpha_1 = \alpha_2 = 0$.

Figure 4 plots the asymptotic rejection rate for the nominal level $\alpha = 0.05$ against the two dimensional family of alternatives

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \epsilon_t$$

where ϵ_t is Gaussian white noise. The first and second plot (left and center) depict the asymptotic power of the univariate tests GS_1 and GS_2 for the Haar wavelet. The black lines correspond to the 0.10 levels and highlight the subset of the parameter space for which the tests are inconsistent. The third plot (right) shows the asymptotic power of the bivariate test GSM_2: its 0.10 level is the intersection of the 0.10 levels for the univariate tests and it consists of only one point, the origin $(\alpha_1, \alpha_2) = (0, 0)$.

B. *Monte Carlo Simulations*

Feasible tests are obtained from Theorem 12 replacing the matrix Σ with a known matrix. A natural choice is to replace all the asymptotic quantities with consistent estimators, for example using the Newey and West (1987) estimator. We denote the corresponding statistic with GSM, and also consider two additional feasible statistics:

1. First, the test statistics can be computed under the assumption that the fourth order cumulants vanish, combining Corollary 11 and 14. We denote these statistics GS^g and GSM^g in the univariate case and multivariate case, respectively.
2. Second, each level GS_i can be computed using an estimator of the long run variance (again, we use the Newey and West's estimator) while using the asymptotic covariance matrix implied by vanishing fourth order cumulants. This feasible statistic is denoted with GSM^{Δ}.

The GS^g and GSM^g tests display accurate empirical size in small samples. With 100 observations and 50,000 replications, the rejection rates at the 1% level against $y_t \sim N(0,1)$ are 0.78%, 1.07%, and 0.82% for the tests ${GS^g}_1$, ${GS^g}_2$, and GSM_2^g, respectively. At the 5% nominal level, the rejection rates are 4.72%, 4.52%, 4.77%. Tables 1 and 2 contain a systematic comparison of the rejection rates of GSM_2^g, GSM_2^{Δ}, GSM_2, the Q_k test of Box and Pierce (1970), and the Esconciano-Lobato test (EL, see Escanciano and Lobato, 2009). We consider sample sizes of 100, 300, 1000, and 5,000 observations and compute the empirical rejection rates form 50,000 replications of the following five different data generating processes under the null hypothesis:

(1) A standard normal process y_t, such that $y_t \sim N(0,1)$;

(2) A GARCH(1,1) process with i.i.d. standard normal innovations,

$$y_t = \sigma_t \epsilon_t \ , \quad \epsilon_t \sim N(0,1) \ , \quad \sigma_t^2 = 0.001 + 0.05 y_{t-1}^2 + 0.90 \sigma_{t-1}^2 \ ;$$

(3) A GARCH(1,1) process with i.i.d innovations following a Student's t with 5 degrees of freedom (and an otherwise identical specification as above)

(4) An EGARCH(1,1) process with i.i.d standard normal innovations

$$y_t = \sigma_t \epsilon_t \ , \quad \epsilon_t \sim N(0,1) \ , \quad \log \sigma_t^2 = 0.001 + 0.5|\epsilon_t| - 0.2\epsilon_t + 0.95\sigma_{t-1}^2 \ ;$$

(5) An mixture of two normals $N(0,1/2)$ and $N(0,1)$ with mixing probability 1/2.

(6) An heterogeneous normal with trending mean: $y_t \sim N(0,t)$.

Insert Table 1 here.

Insert Table 2 here.

For a small sample size (100 observations), the GSM_2^g test has an accurate rejection rate across several of the models analyzed, both at the 1% level and the 5% level, with the exception of the EGARCH model and model (6) (trending variance). With larger sample sizes (1000 and above) and in the presence of a marked deviation from normality, the gains from estimating the asymptotic covariance matrix are significant. Indeed, under these circumstances, the size of the $GSM_{1,2}$ is accurate across all models (in particular at the 5% level). In general, the test GSM^{Δ} performs satisfactorily across all models: at the 1% level GSM^{Δ} dominates EL in all cases but against EGARCH, while at the 5% level with T $\geq$ 300 the two test perform very similarly (although, EL maintains a significant edge against EGARCH).

Figure 5 illustrates the empirical power functions of the tests GS_1^g, GS_2^g, and GSM_2^g against two one-dimensional families of alternatives, an AR(1) model (AR1: $y_t = \alpha y_{t-1} + \epsilon_t$) and a restricted AR(2) model (RAR2: $y_t = \alpha y_{t-2} + \epsilon_t$) with standard normal innovations. The rejection rates are computed with respect to a 1% nominal size for sample sizes of 100, 300, and 1000 observations. From the first row, it is apparent that the test GS_1^g has strong power against an AR1 alternative while at the same time its power is practically orthogonal to an RAR2 deviation from the null. The second row shows that the test GS_2^g has a complementary behavior: its power against AR1 deviations from the null is uneven, while it displays strong power against RAR2 deviations form the null. Finally, the last row illustrates how the joint test GSM_2^g incorporates the best properties of the single scale tests. The power of GSM_2^g is consistently high against AR1 and RAR2 alternatives. The panels in Figure 5 also show that the power of the various tests increases steadily as the sample size increases.

Insert Figure 5 here.

To further understand how the power of the GS test family varies against the two-parameter family

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \epsilon_t \ , \quad \epsilon_t \sim N(0,1) \ , \tag{18}$$

we plot in Figure 6 the contours of the power surface obtained varying α_1 in the interval $(-0.50, 0.50)$ and α_2 in $(-0.45, 0.45)$. Simulations are carried out for a grid of values of the parameters spaced by 0.05 ans intermediate values are interpolated. The black lines correspond to 25%, 50%, 75%, and 100% percent power (starting form the center), while the grey lines correspond to 5% increments. Approximately, contour lines of the power function of GS_1^g test (first panel) run vertically, an indication that the first scale test is not very sensitive to variations in the parameter α_2. This picture is approximately reversed in the second panel: the contour lines for the GS_2^g test run horizontally. In the third panel we see that the contour lines of the multi-scale test GSM_2^g are, even in small samples, close to ellipses, the shape predicted by our asymptotic results.

Insert Figure 6 here.

In the reminder of this section we restrict our analysis to a size of 1% (results are similar at the 5% level) and a sample size of 100 observations.

An accurate analysis is contained in Table 3, where we compare the size adjusted power of the three tests against the two-dimensional Gaussian AR(2) alternative defined in Equation (18). The first column contains the size adjusted power of each test for various alternatives.[14] In the second column we report the relative power gains of the multi-scale test GSM_2^g with respect to the LB tests, the BP test and the EL test. Against the great majority of the alternatives the GSM_2^g test

[14]Size adjusted power is computed using, for a given sample size, the empirical critical values obtained from Monte Carlo simulations with 100,000 replications.

outperforms the BP and LB tests.[15] The GSM_2^g test clearly outperforms the EL test when the first order parameter is negative ($\alpha_1 < 0$) with a power improvement of up to 125%. When α_1 is positive, neither test has a clear edge, with variations in power against various alternatives between +44% and −49%. [16]

In Table 4 we repeat the previous power analysis for AR(2) models with GARCH(1,1) innovations (with the same parameters as in model (5)). Qualitatively the results are unchanged: the GSM_2^g outperforms the BP and LB tests across a wide variety of alternatives (by up to 283% and 311%, respectively); the GSM_2^g also outperforms the EL test when the first order autoregressive coefficient is negative (by up to 134%), while when $\alpha_1 > 0$, neither test has a clear advantage.

Insert Table 3 here.

Insert Table 4 here.

In econometric practice, it necessary to choose a value for N. Ultimately, this choice is dictated by the amount of data available, as deeper wavelet decompositions consume more degrees of freedom. According to Percival and Walden (2000), the properties of the wavelet variance estimator are well approximated by its asymptotic distribution whenever $T - L_{h,m} > 128$, where $L_{h,m}$ is the length of the $m - th$ level filter. Recall that $L_{h,m} = (2^m - 1)(L_h - 1) + 1$, where L_h is the length of wavelet filter. We report some size and power simulations comparing various of N up to 6. Table 5 shows that in general there is a trade off between the depth of the wavelet decomposition and the sample size: for small sample size, a shallower wavelet decomposition has better size properties.

Insert Table 5 here.

To investigate power as N is allowed to vary, we consider the restricted autoregressive model $rar(p)$ as $y_t = 0.1y_{t-p} + \epsilon_t$ for $p = 1, 2, 4$. Table 6 illustrates another tradeoff: lower values of N correspond to higher power but only against ARMA models of lower order.

Insert Table 6 here.

Finally, in our simulations the choice of the wavelet family was generally influential, with small idiosyncratic differences across various nulls and alternative models.

VI. Application to High Frequency Finance

In this section we apply our test and the AQ test to high frequency market data, specifically to returns from transactions of Apple Inc. (AAPL). We use intraday data from January 2, 2012 to

[15]Analogous results hold for Gaussian MA(2) and Gaussian ARMA(2,2) alternatives. The results are very close to those of Table 3. These results are available upon request.

[16]Despite our adjustments, sized-distortions remain because of the random nature of the Monte Carlo simulations.

December 28, 2012 and restrict our sample to the 10 minutes time interval form 11:50 to 12:00. Using data from TAQ we construct 1-second returns from transactions for the entire period and test each day for serial correlation, so that for each test the sample consists of 600 observations. Serial correlation at high frequency is on hand to liquidity measures (as an indirect estimate of the bid-ask spread, see Roll, 1984) and on the other to market efficiency (see, for example, Jegadeesh and Titman, 2001).

The average p-values over the 251 testing days for the tests GM_4^{Δ}, GSM_4 and AQ are 0.0077, 0.0109, and 0.0130, respectively (we do not report the other test because of the large size distortions). On average, our wavelet based tests reject the null of no serial correlation slightly more strongly than the AQ test. This example shows that our test can be useful in econometric practice.

VII. Conclusions

We use the wavelet coefficients of the observed time series to construct a test statistics in the spirit of Von Neumann variance ratio tests. In our approach, there is no intermediate step such as the estimation of the spectral density for the null and alternative hypotheses. Therefore, we are not constrained with the rate of convergence of nonparametric estimators.

Our analysis of consistency and power does not apply to more general local alternatives, such as

$$H_{1,t} : S_y(f) = T^{-1/2}\left(S(f;T) - \frac{1}{2}\right) + \frac{1}{2}\ ,$$

where the lag order is allowed to grow with T. On one hand, we have already established that all tests are inconsistent against certain carefully designed alternatives. On the other, we expect that, much like variance tests in the spirit of Lo and MacKinlay (1989), there is an optimal choice of N that will maximize power(see for example Deo and Richardson, 2003; Perron and Vodounou, 2005). A related, and more general, issue is that of choosing optimally the wavelet decomposition to be used. Intuitively, it is clear that, for a given alternative, there is a choice of frequency bands that will maximize power, namely those bands that deviate the most from the white noise baseline. The development of an adaptive version of the current test could resolve the problem of inconsistency while providing better all round power properties.

Another 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). 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. Our current framework can be generalized to residual-based tests and it embeds Durbin-Watson's d-test as a special case. These extensions are currently under investigation by the authors.

References

Allan, D. W. (1966). Statistics of atomic frequency standards. *Proceedings of the IEEE 31*, 221–230.

Andrews, D. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica 59*(3), 817–858.

Baillie, R., T. Bollerslev, and H. Mikkelsen (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics 74*(1), 3–30.

Bierens, H. (2004). *Introduction to the Mathematical and Statistical Foundations of Econometrics.* Cambridge Univiversity Press.

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

Box, G. and D. Pierce (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association 65*(332), 1509–1526.

Breusch, T. (1978). Testing for autocorrelation in dynamic linear models. *Australian Economic Papers 17*, 334–355.

Brockwell, P. and R. Davis (2009). *Time Series: Theory and Methods*. Springer, New York.

Coifman, R. and D. Donoho (1995). Translation invariant de noising. In A. A. Antoniadis and G. Oppenheim (Eds.), *Wavelets and Statistics*, Volume 103, pp. 125–150. Springer, New York.

Daubechies, I. (1993). Orthonormal bases of compactly supported wavelets ii. variations on a theme. *SIAM Journal on Mathematical Analysis 24*, 499.

Davidson, J. (1992). A central limit theorem for globally nonstationary near-epoch dependent functions of mixing processes. *Econometric Theory 8*, 313–329.

Davidson, J. (1993). The central limit theorem for globally nonstationary near-epoch dependent functions of mixing processes: the asymptotically degenerate case. *Econometric Theory 9*, 402–412.

Davidson, J. (1995). *Stochastic Limit Theory*. Oxford University Press, Oxford.

Davidson, J. (2002). Establishing conditions for the functional central limit theorem in nonlinear and semiparametric time series processes. *Journal of Econometrics 106*(2), 243–269.

Davidson, J. (2004). Moment and memory properties of linear conditional heteroscedasticity models, and a new model. *Journal of Business and Economic Statistics 22*(1), 16–29.

De Jong, R. (1997). Central limit theorems for dependent heterogeneous random variables. *Econometric Theory 13*(03), 353–367.

Deo, R. and M. Richardson (2003). On the asymptotic power of the variance ratio test. *Econometric Theory 19*(2), 231–239.

Duchesne, P. (2006). On testing for serial correlation with a wavelet-based spectral density estimator in multivariate time series. *Econometric Theory 22*(4), 633.

Duchesne, P., L. Li, and J. Vandermeerschen (2010). On testing for serial correlation of unknown form using wavelet thresholding. *Computational Statistics & Data Analysis 54*(11), 2512–2531.

Durbin, J. and G. Watson (1950). Testing for serial correlation in least squares regression. i. *Biometrika 37*(3/4), 409–428.

Engle, R. and T. Bollerslev (1986). Modelling the persistence of conditional variances. *Econometric reviews 5*(1), 1–50.

Escanciano, J. and I. Lobato (2009). An automatic portmanteau test for serial correlation. *Journal of Econometrics 151*(2), 140–149.

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

Gallant, A. and H. White (1988). *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*. Basil Blackwell New York.

Gençay, R., F. Selçuk, and B. Whitcher (2001). *An Introduction to Wavelets and Other Filtering Methods in Finance and Economics.* Academic Press, San Diego.

Godfrey, L. (1978). Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables. *Econometrica 46*(6), 1293–1301.

Granger, C. and P. Newbold (1986). *Forecasting time series*. Academic Press, New York.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press.

Hansen, B. (1991). GARCH(1, 1) processes are near epoch dependent. *Economics Letters 36*(2), 181–186.

Hong, Y. (1996). Consistent testing for serial correlation of unknown form. *Econometrica 64*, 837–864.

Hong, Y. and C. Kao (2004). Wavelet-based testing for serial correlation of unknown form in panel models. *Econometrica 72*, 1519–1563.

Horowitz, J., I. Lobato, J. Nankervis, and N. Savin (2006). Bootstrapping the Box–Pierce Q-test: A robust test of uncorrelatedness. *Journal of Econometrics 133*(2), 841–862.

Howe, D. A. and D. B. Percival (1995). Wavelet variance, Allan variance, and leakage. *IEEE Transactions on Intrumentation and Measurement 44*, 94–97.

Ibragimov, I. (1962). Some limit theorems for stationary processes. *Theory of Probability & Its Applications 7*(4), 349–382.

Jegadeesh, N. and S. Titman (2001). Profitability of momentum strategies: An evaluation of alternative explanations. *The Journal of Finance 56*(2), 699–720.

Lee, J. and Y. Hong (2001). Testing for serial correlation of unknown form using wavelet methods. *Econometric Theory 17*, 386–423.

Ljung, G. and G. Box (1978). On a measure of lack of fit in time series models. *Biometrika 65*(2), 297–303.

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

Lo, A. W. and A. C. MacKinlay (1989). The size and power of the variance ratio test in finite samples: A monte carlo investigation. *Journal of econometrics 40*(2), 203–238.

Lobato, I. (2001). Testing that a dependent process is uncorrelated. *Journal of the American Statistical Association 96*(455), 1066–1076.

Lobato, I., J. Nankervis, and N. Savin (2002). Testing for zero autocorrelation in the presence of statistical dependence. *Econometric Theory 18*(3), 730–743.

McCullagh, P. (1987). *Tensor Methods in Statistics*. Chapman and Hall London.

Nason, G. P. and B. W. Silverman (1995). The stationary wavelet transform and some statistical applications. In A. Antoniadis and G. Oppenheim (Eds.), *Wavelets and Statistics*, Volume 103 of *Lecture Notes in Statistics*, pp. 281–300. Spinger, New York.

Newey, W. and K. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*(3), 703–708.

Paparoditis, E. (2000). Spectral density based goodness-of-fit tests for time series analysis. *Scandinavian Journal of Statistics 27*, 143–176.

Percival, D. B. (1983). *The Statistics of Long Memory Processes*. Ph. D. thesis, Department of Statistics, University of Washington.

Percival, D. B. (1995). On estimation of the wavelet variance. *Biometrica 82*, 619–631.

Percival, D. B. and P. Guttorp (1994). Long-memory processes, the Allan variance and wavelets. In E. Foufoula-Georgiou and P. Kumar (Eds.), *Wavelets in Geophysics*, Volume 4 of *Wavelet Analysis and its Applications*, pp. 325–344. Academic Press.

Percival, D. B. and H. O. Mofjeld (1997). Analysis of subtidal coastal sea level fluctuations using wavelets. *Journal of the American Statistical Association 92*, 868–880.

Percival, D. B. and A. T. Walden (2000). *Wavelet Methods for Time Series Analysis*. Cambridge: Cambridge Press.

Perron, P. and C. Vodounou (2005). The variance ratio test: An analysis of size and power based on a continuous-time asymptotic framework. *Econometric Theory 21*(03), 562–592.

Powell, M. J. D. (1981). *Approximation theory and methods.* Cambridge University Press.

Priestley, M. (1988). *Non-linear and non-stationary time series analysis.* Academic Press.

Reed, M. and B. Simon (1972). *Methods of Modern Mathematical Physics I: Functional Analysis.* New York, San Fransisco, London: Academic Press.

Roll, R. (1984). A simple implicit measure of the effective bid-ask spread in an efficient market. *The Journal of Finance 39*(4), 1127–1139.

Romano, J. and L. Thombs (1996). Inference for autocorrelations under weak assumptions. *Journal of the American Statistical Association 91*(434), 590–600.

Von Neumann, J. (1941). Distribution of the ratio of the mean square successive difference to the variance. *The Annals of Mathematical Statistics 12*(4), 367–395.

Xue, Y., R. Gençay, and S. Fagan (2010). Testing for jump arrivals in financial time series. *Simon Fraser University*.

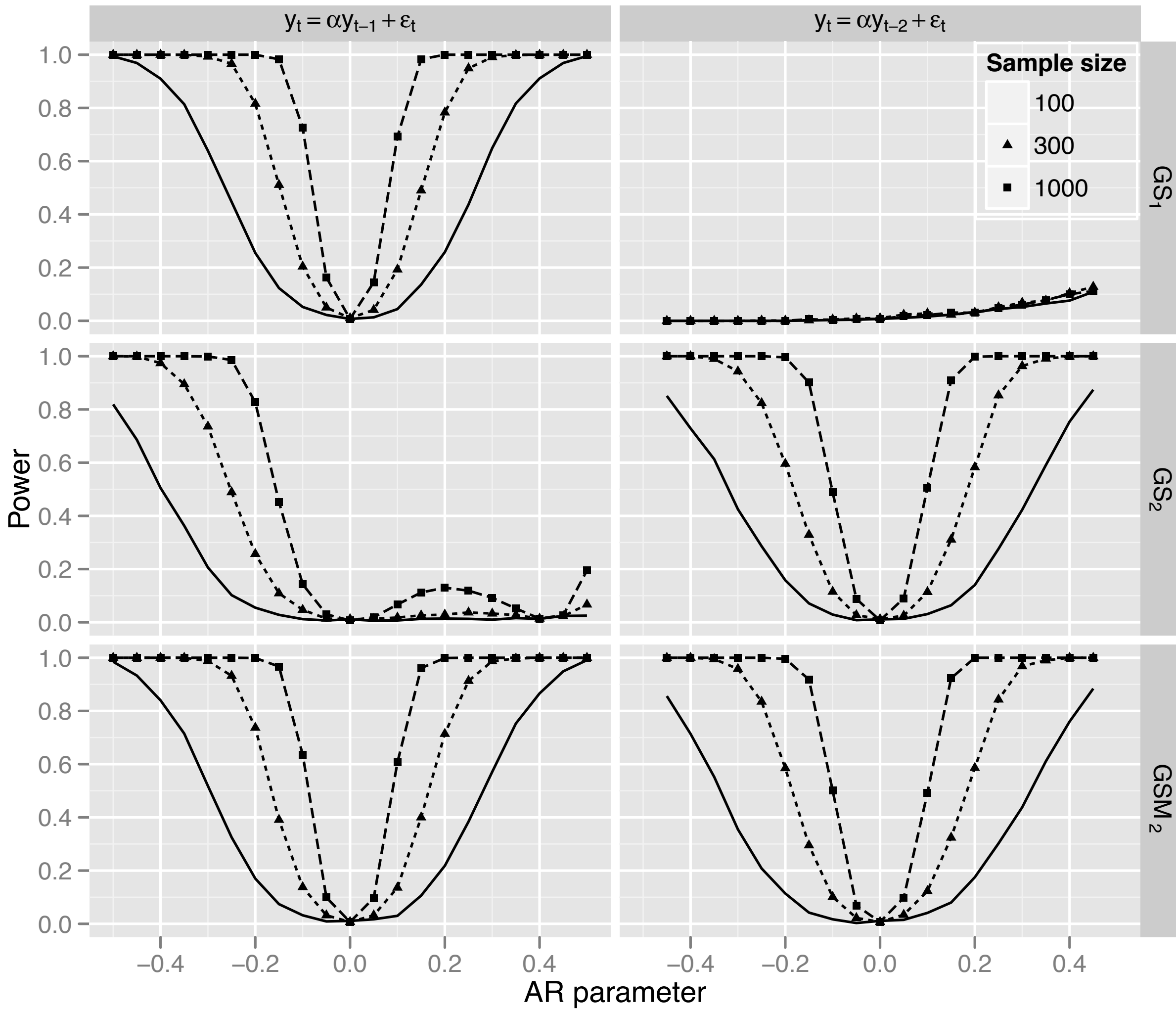

FIGURE 5. Empirical power functions of the tests GS_1^g, GS_2^g, and GSM_2^g (first, second, and third row, respectively) against AR(1) and AR(2) alternatives (first and second columns, respectively). The rejection rates are based on 5,000 replications with 1% nominal size for sample sizes of 100 (circle), 300 (triangle), and 1000 (square) observations.

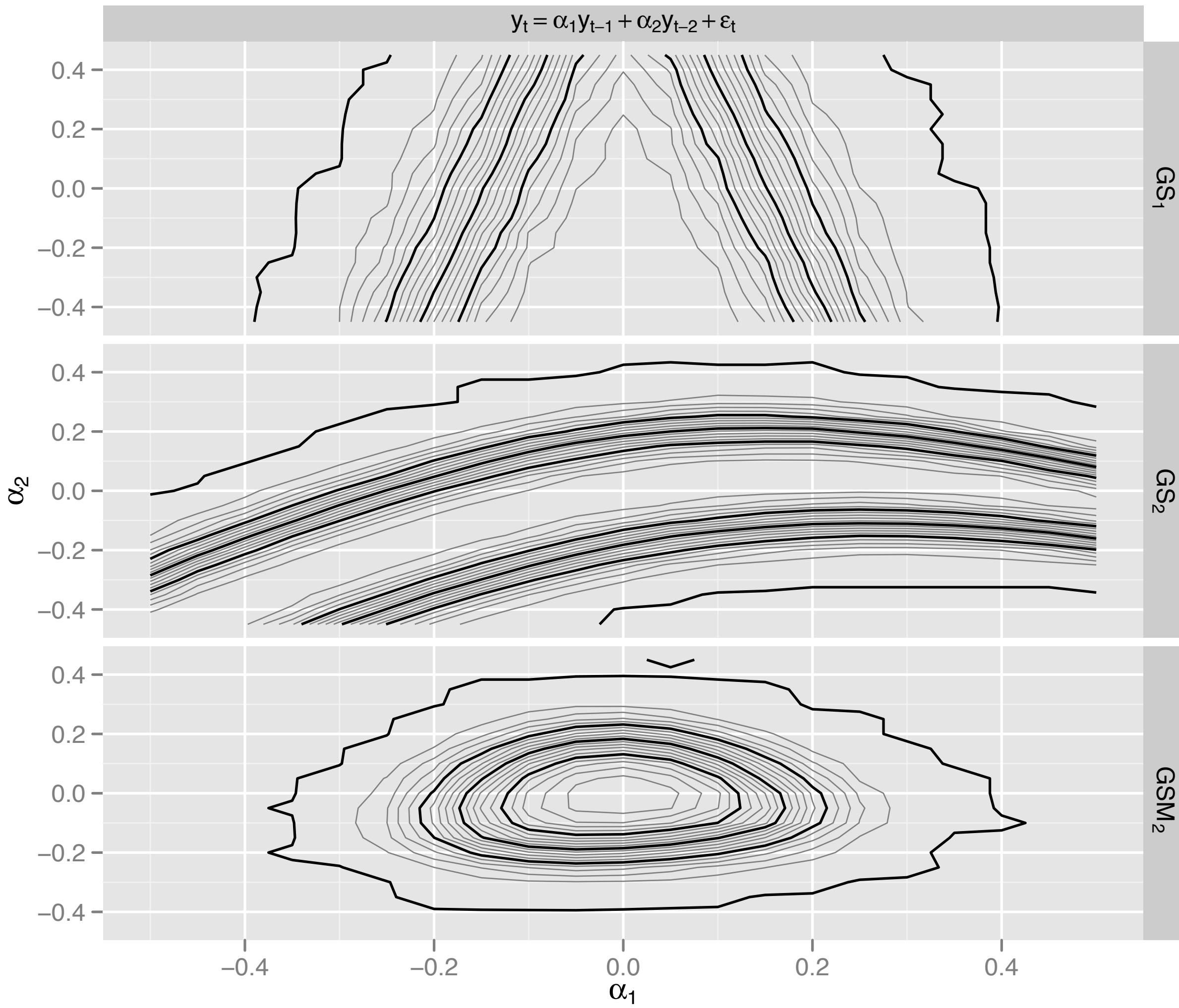

FIGURE 6. Contours of the power surface of the tests GS_1^g, GS_2^g, and GSM_2^g against the Gaussian AR(2) alternative. Simulations are carried out for a grid of values of the parameters obtained varying α_1 in the interval $(-0.50, 0.50)$ and α_2 in $(-0.45, 0.45)$ in steps of size 0.05. Intermediate values are interpolated. From the center of each graph, the black lines correspond to the 25-th, 50-th, 75-th and 100-th quantiles, while each grey line corresponds to a 5% increment.

Table 1
Rejection rates under the null hypothesis at 1% nominal level

Rejection probabilities in percentages of tests with nominal levels of 1% against five different data generating processes under the null hypothesis:

(1) A standard normal process $y_t \sim N(0,1)$;
(2) A GARCH(1,1) process with i.i.d. standard normal innovations,

$$y_t = \sigma_t \epsilon_t \ , \quad \epsilon_t \sim N(0,1) \ , \quad \sigma_t^2 = 0.001 + 0.05 y_{t-1}^2 + 0.90 \sigma_{t-1}^2 \ ;$$

(3) A GARCH(1,1) process with i.i.d innovations following a Student's t with 5 degrees of freedom;
(4) An EGARCH(1,1) process with i.i.d standard normal innovations

$$y_t = \sigma_t \epsilon_t \ , \quad \epsilon_t \sim N(0,1) \ , \quad \log \sigma_t^2 = 0.001 + 0.5|\epsilon_t| - 0.2\epsilon_t + 0.95 \sigma_{t-1}^2 \ ;$$

(5) An mixture of two normals $N(0,1/2)$ and $N(0,1)$ with mixing probability 1/2.
(6) An heterogeneous normal with trending mean: $y_t \sim N(0,t)$.

The tests GSM^g, GSM^Δ, and GSM are computed assuming zero fourth order cumulants, estimating the scaling coefficients, and estimating scaling coefficients and asymptotic covariance matrix, respectively; Q_k is the Box and Pierce test with k lags; EL is the Escanciano and Lobato test. All size simulations based on 50,000 replications.

	$N(0,1)$				$N(0,1)$-GARCH(1,1)				t_5-GARCH(1,1)			
T	100	300	1000	5000	100	300	1000	5000	100	300	1000	5000
GSM_2^g	0.82	0.92	0.87	1.04	1.32	1.81	1.62	1.86	1.76	2.84	4.04	5.16
GSM_2^Δ	2.75	1.47	1.12	1.21	2.60	1.65	1.12	1.22	2.14	1.04	1.20	1.12
GSM_2	5.25	2.75	1.87	1.54	5.13	2.72	1.64	1.55	4.26	2.14	1.70	1.25
Q_5	0.86	0.88	0.95	1.06	1.22	1.94	1.81	2.17	1.81	3.59	5.67	7.51
Q_{10}	0.90	1.02	1.03	1.08	1.60	2.34	2.26	2.70	1.95	4.24	7.08	10.51
Q_{20}	0.88	1.15	1.02	1.14	1.51	2.40	2.59	2.63	1.59	4.98	8.51	12.08
EL	2.73	2.28	1.71	1.23	2.65	2.68	1.80	1.36	2.17	1.94	1.78	1.25
	$N(0,1)$-EGARCH(1,1)				Mixture of Normals				Trending σ			
T	100	300	1000	5000	100	300	1000	5000	100	300	1000	5000
GSM_2^g	8.45	19.67	32.22	45.52	0.86	0.90	0.98	1.24	2.63	2.87	2.88	3.05
GSM_2^Δ	1.65	0.66	0.46	0.55	2.64	1.40	1.32	1.40	1.72	1.20	1.11	1.19
GSM_2	4.14	1.84	1.06	0.91	4.98	2.63	1.93	1.73	4.48	2.19	1.84	1.55
Q_5	12.15	30.15	49.57	67.07	0.93	0.79	1.09	1.03	3.49	3.99	4.34	4.86
Q_{10}	12.15	36.45	59.61	79.79	0.88	0.94	0.97	1.02	4.28	5.57	6.22	6.93
Q_{20}	8.40	35.13	63.07	84.86	0.83	0.90	1.04	0.99	4.38	7.90	9.62	10.18
EL	1.98	1.93	1.38	1.17	2.58	2.28	1.83	1.55	2.56	2.31	1.87	1.26

Table 2
Rejection rates under the null hypothesis at 5% nominal level

Rejection probabilities in percentages of tests with nominal levels of 5% against five different data generating processes under the null hypothesis:

(1) A standard normal process $y_t \sim N(0,1)$;
(2) A GARCH(1,1) process with i.i.d. standard normal innovations,

$$y_t = \sigma_t \epsilon_t \ , \quad \epsilon_t \sim N(0,1) \ , \quad \sigma_t^2 = 0.001 + 0.05 y_{t-1}^2 + 0.90 \sigma_{t-1}^2 \ ;$$

(3) A GARCH(1,1) process with i.i.d innovations following a Student's t with 5 degrees of freedom;
(4) An EGARCH(1,1) process with i.i.d standard normal innovations

$$y_t = \sigma_t \epsilon_t \ , \quad \epsilon_t \sim N(0,1) \ , \quad \log \sigma_t^2 = 0.001 + 0.5|\epsilon_t| - 0.2\epsilon_t + 0.95 \sigma_{t-1}^2 \ ;$$

(5) An mixture of two normals $N(0,1/2)$ and $N(0,1)$ with mixing probability 1/2.
(6) An heterogeneous normal with trending mean: $y_t \sim N(0,t)$.

The tests GSM^g, GSM^Δ, and GSM are computed assuming zero fourth order cumulants, estimating the scaling coefficients, and estimating scaling coefficients and asymptotic covariance matrix, respectively; Q_k is the Box and Pierce test with k lags; EL is the Escanciano and Lobato test. All size simulations based on 50,000 replications.

	$N(0,1)$				$N(0,1)$-GARCH(1,1)				t_5-GARCH(1,1)			
T	100	300	1000	5000	100	300	1000	5000	100	300	1000	5000
GSM_2^g	4.77	4.71	4.74	5.11	5.63	7.06	7.02	7.53	6.56	8.94	11.58	13.34
GSM_2^Δ	9.21	6.32	5.53	5.43	8.29	6.53	5.48	5.61	7.46	5.85	5.49	5.14
GSM_2	13.32	8.37	7.12	6.33	12.39	8.84	7.11	6.30	11.26	7.64	7.01	5.69
Q_5	4.12	4.75	4.74	5.01	6.00	7.54	7.69	8.37	6.41	10.71	14.60	17.97
Q_{10}	4.06	4.53	4.64	5.19	5.91	8.11	8.74	9.44	6.00	11.87	16.97	22.87
Q_{20}	3.20	4.29	4.49	5.05	4.93	7.61	9.02	10.46	4.82	11.51	18.13	25.32
EL	7.80	6.70	5.47	5.50	7.66	6.83	5.52	5.39	7.33	5.86	5.56	5.07
	$N(0,1)$-EGARCH(1,1)				Mixture of Normals				Trending σ			
T	100	300	1000	5000	100	300	1000	5000	100	300	1000	5000
GSM_2^g	18.37	33.17	46.74	59.24	4.55	5.05	4.86	5.41	8.74	9.60	10.16	10.42
GSM_2^Δ	6.22	3.83	3.21	3.97	9.04	6.40	5.77	5.74	7.61	5.90	5.46	5.46
GSM_2	10.80	6.43	4.73	4.92	13.06	8.57	7.26	6.69	11.75	7.86	6.74	5.98
Q_5	24.32	45.80	64.90	79.30	3.93	4.81	4.67	5.48	11.24	12.31	13.92	14.05
Q_{10}	23.53	51.90	73.62	88.80	3.69	4.45	5.10	5.14	12.05	15.57	17.65	18.31
Q_{20}	16.86	50.37	77.03	92.82	3.15	4.52	4.61	5.31	11.65	19.49	23.98	25.32
EL	6.64	5.65	5.11	4.94	7.92	6.57	5.75	5.44	7.92	6.34	5.54	4.97

Table 3
Size-adjusted power against Gaussian AR(2) processes

Power and relative power against the two-parameter family

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \epsilon_t \;, \quad \epsilon_t \sim N(0,1) \;,$$

Simulations are carried out for set of alternatives obtained varying α_1 in the interval $(-0.50, 0.50)$ and α_2 in $(-0.45, 0.45)$ in increments of 0.05.

GSM_2^g

	α_1:	0.30	0.20	0.10	0.00	−0.10	−0.20	−0.30
	0.30	94.3	76.2	51.6	43.8	62.0	85.7	96.9
	0.20	85.1	54.1	23.3	17.2	33.4	64.6	89.6
	0.10	69.7	32.8	8.7	4.3	13.2	40.3	74.9
α_2	0.00	53.7	18.3	3.2	1.2	4.6	21.3	56.1
	−0.10	39.7	11.0	2.7	2.5	5.2	17.1	46.5
	−0.20	33.4	11.8	8.0	11.9	18.5	31.6	54.3
	−0.30	40.5	27.2	29.2	37.3	48.3	60.8	76.3

Q_{20}

	α_1:	0.30	0.20	0.10	0.00	−0.10	−0.20	−0.30
	0.30	84.7	58.8	31.7	19.9	25.7	49.5	77.8
	0.20	64.4	33.3	12.3	6.8	9.6	26.1	56.2
	0.10	39.3	15.3	5.0	1.7	3.6	12.1	33.4
α_2	0.00	21.6	7.3	2.2	1.1	1.7	5.6	17.9
	−0.10	14.5	5.9	2.8	1.9	2.2	4.2	11.3
	−0.20	16.1	9.9	7.1	6.7	7.2	9.2	14.7
	−0.30	28.2	23.3	21.1	19.7	20.5	23.6	28.9

Relative power: (GSM_2^g/Q_{20}) - 1

	α_1:	0.30	0.20	0.10	0.00	−0.10	−0.20	−0.30
	0.30	0.11	0.30	0.63	1.20	1.42	0.73	0.25
	0.20	0.32	0.62	0.89	1.53	2.50	1.47	0.59
	0.10	0.77	1.14	0.75	1.49	2.68	2.33	1.24
α_2	0.00	1.49	1.52	0.48	0.07	1.74	2.82	2.14
	−0.10	1.74	0.87	−0.04	0.27	1.32	3.08	3.10
	−0.20	1.08	0.19	0.13	0.79	1.57	2.44	2.70
	−0.30	0.43	0.17	0.38	0.89	1.36	1.57	1.64

LB

	α_1:	0.30	0.20	0.10	0.00	−0.10	−0.20	−0.30
	0.30	82.7	55.5	29.3	17.7	23.0	45.5	75.1
	0.20	60.9	30.5	11.2	6.1	8.6	24.0	52.8
	0.10	35.5	13.7	4.4	1.6	3.3	10.9	30.6
α_2	0.00	19.1	6.5	2.1	1.0	1.6	5.0	15.9
	−0.10	12.8	5.3	2.5	1.8	2.2	3.8	10.0
	−0.20	14.3	8.9	6.4	6.3	6.5	8.3	12.8
	−0.30	25.1	20.8	18.7	17.8	18.7	21.3	25.9

Relative power: (GSM_2^g/LB) - 1

	α_1:	0.30	0.20	0.10	0.00	−0.10	−0.20	−0.30
	0.30	0.14	0.37	0.76	1.47	1.70	0.88	0.29
	0.20	0.40	0.78	1.07	1.83	2.87	1.69	0.70
	0.10	0.96	1.39	0.95	1.65	3.04	2.69	1.45
α_2	0.00	1.81	1.80	0.52	0.18	1.80	3.25	2.53
	−0.10	2.11	1.10	0.05	0.38	1.39	3.55	3.66
	−0.20	1.34	0.32	0.25	0.91	1.83	2.79	3.24
	−0.30	0.61	0.31	0.56	1.10	1.59	1.85	1.95

EL

	α_1:	0.30	0.20	0.10	0.00	−0.10	−0.20	−0.30
	0.30	94.8	80.1	57.7	42.6	51.4	73.7	91.9
	0.20	82.4	54.0	26.8	15.1	21.7	45.3	74.4
	0.10	59.2	27.5	8.7	3.3	6.5	20.8	50.8
α_2	0.00	39.8	12.8	2.8	1.3	2.0	8.7	32.1
	−0.10	29.2	9.8	4.6	3.5	4.1	7.8	23.3
	−0.20	31.0	20.1	15.7	14.4	15.7	19.7	28.9
	−0.30	55.1	50.3	45.1	41.9	43.9	48.7	55.1

Relative power: (GSM_2^g/EL) - 1

	α_1:	0.30	0.20	0.10	0.00	−0.10	−0.20	−0.30
	0.30	0.00	−0.05	−0.11	0.03	0.21	0.16	0.06
	0.20	0.03	0.00	−0.13	0.14	0.54	0.42	0.20
	0.10	0.18	0.19	0.00	0.32	1.04	0.94	0.47
α_2	0.00	0.35	0.44	0.13	−0.12	1.25	1.44	0.75
	−0.10	0.36	0.13	−0.43	−0.30	0.28	1.19	1.00
	−0.20	0.08	−0.41	−0.49	−0.17	0.18	0.61	0.88
	−0.30	−0.27	−0.46	−0.35	−0.11	0.10	0.25	0.39

Table 4
Size-adjusted power against GARCH(1,1)-AR(2) processes

Power and relative power against the two-parameter family

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \epsilon_t \ ,$$
$$\epsilon_t = \sigma_t z_t \ , \quad z \sim N(0,1) \ , \quad \sigma_t^2 = 0.001 + 0.05 y_{t-1}^2 + 0.90 \sigma_{t-1}^2 \ .$$

Simulations are carried out for set of alternatives obtained varying α_1 in the interval $(-0.50, 0.50)$ and α_2 in $(-0.45, 0.45)$ in increments of 0.05.

GSM_2^g

		α_1						
		0.30	0.20	0.10	0.00	−0.10	−0.20	−0.30
	0.30	92.5	70.7	46.6	40.5	57.4	82.7	96.3
	0.20	80.7	48.8	19.9	15.0	32.6	62.5	88.6
	0.10	65.1	28.1	7.2	4.2	12.6	36.7	71.0
α_2	0.00	46.9	15.3	2.3	1.0	4.3	19.2	52.1
	−0.10	34.4	8.7	1.8	2.0	4.7	14.8	41.2
	−0.20	27.1	8.4	6.4	9.8	16.3	27.1	50.2
	−0.30	30.0	20.9	23.1	32.5	41.9	54.0	70.3

Q_{20}

		α_1						
		0.30	0.20	0.10	0.00	−0.10	−0.20	−0.30
	0.30	83.1	55.2	28.3	17.5	22.9	46.4	74.3
	0.20	60.4	30.2	11.4	6.4	9.7	24.7	52.8
	0.10	38.3	14.6	4.6	2.0	3.4	11.5	30.0
α_2	0.00	19.8	6.2	2.2	1.1	2.1	5.1	17.0
	−0.10	12.3	4.6	2.6	1.9	2.3	4.2	10.9
	−0.20	14.0	8.8	6.3	5.8	5.8	8.4	13.1
	−0.30	25.6	20.5	18.0	17.4	17.6	20.0	25.5

Relative power: (GS_2/Q_{20}) - 1

		α_1						
		0.30	0.20	0.10	0.00	−0.10	−0.20	−0.30
	0.30	0.11	0.28	0.64	1.31	1.51	0.78	0.30
	0.20	0.33	0.62	0.75	1.36	2.34	1.53	0.68
	0.10	0.70	0.93	0.55	1.10	2.70	2.20	1.37
α_2	0.00	1.37	1.46	0.06	−0.02	1.03	2.73	2.06
	−0.10	1.80	0.89	−0.29	0.04	1.08	2.50	2.76
	−0.20	0.94	−0.04	0.01	0.69	1.81	2.21	2.83
	−0.30	0.17	0.02	0.28	0.87	1.39	1.70	1.75

LB

		α_1						
		0.30	0.20	0.10	0.00	−0.10	−0.20	−0.30
	0.30	81.0	52.0	26.7	15.8	21.2	43.5	72.0
	0.20	57.2	27.7	10.6	5.8	9.1	22.9	49.6
	0.10	35.5	13.4	4.4	1.9	3.3	10.3	27.8
α_2	0.00	18.0	5.7	2.2	1.0	2.1	4.9	15.3
	−0.10	11.5	4.4	2.5	2.0	2.3	4.0	9.8
	−0.20	12.4	8.1	6.0	5.5	5.4	7.7	12.2
	−0.30	23.2	18.8	16.5	16.2	16.0	18.4	23.1

Relative power: (GSM_2^g/LB) - 1

		α_1						
		0.30	0.20	0.10	0.00	−0.10	−0.20	−0.30
	0.30	0.14	0.36	0.74	1.56	1.70	0.90	0.34
	0.20	0.41	0.76	0.88	1.59	2.56	1.73	0.79
	0.10	0.83	1.10	0.63	1.24	2.81	2.58	1.56
α_2	0.00	1.61	1.70	0.08	0.02	1.01	2.93	2.40
	−0.10	2.00	1.00	−0.28	0.02	1.10	2.69	3.20
	−0.20	1.18	0.04	0.07	0.77	2.00	2.54	3.11
	−0.30	0.29	0.11	0.40	1.01	1.62	1.94	2.04

EL

		α_1						
		0.30	0.20	0.10	0.00	−0.10	−0.20	−0.30
	0.30	94.0	78.2	54.6	40.8	48.7	72.3	90.6
	0.20	79.7	51.8	25.5	14.4	21.6	44.2	73.1
	0.10	57.8	25.2	7.6	3.6	6.7	19.4	47.6
α_2	0.00	37.1	11.5	2.3	1.4	1.8	8.5	29.0
	−0.10	26.2	8.3	3.5	2.9	3.6	6.4	21.4
	−0.20	28.1	17.4	14.1	11.9	13.5	17.6	27.0
	−0.30	50.0	45.5	41.4	38.6	39.7	45.3	51.5

Relative power: (GSM_2^g/EL) - 1

		α_1						
		0.30	0.20	0.10	0.00	−0.10	−0.20	−0.30
	0.30	−0.02	−0.10	−0.15	−0.01	0.18	0.14	0.06
	0.20	0.01	−0.06	−0.22	0.05	0.51	0.41	0.21
	0.10	0.13	0.11	−0.05	0.16	0.90	0.90	0.49
α_2	0.00	0.26	0.34	0.00	−0.24	1.34	1.26	0.80
	−0.10	0.31	0.05	−0.47	−0.30	0.33	1.30	0.92
	−0.20	−0.04	−0.52	−0.55	−0.18	0.21	0.54	0.86
	−0.30	−0.40	−0.54	−0.44	−0.16	0.05	0.19	0.37

Table 5
Size for higher order wavelet decompositions

Rejection probabilities of tests with nominal levels of 5% against the following models for the null

(1) A standard normal process $y_t \sim N(0,1)$;
(2) A Student-t process y_t with 3 degrees of freedom;
(3) A GARCH(1,1) process with i.i.d. standard normal innovations,

$$y_t = \sigma_t \epsilon_t \ , \quad \epsilon_t \sim N(0,1) \ , \quad \sigma_t^2 = 0.001 + 0.05 y_{t-1}^2 + 0.90 \sigma_{t-1}^2 \ ;$$

All simulations based on 10,000 replications.

model	N	GSM_2	GSM_3	GSM_4	GSM_5	GSM_6
norm	100	0.0518	0.0499	0.0563	0.0642	0.0787
t3	100	0.0385	0.0388	0.0428	0.0503	0.0664
garch	100	0.0566	0.0603	0.0667	0.0727	0.0896
norm	300	0.0468	0.0503	0.0547	0.0574	0.0650
t3	300	0.0425	0.0431	0.0433	0.0496	0.0565
garch	300	0.0644	0.0673	0.0705	0.0729	0.0831
norm	1000	0.0493	0.0500	0.0506	0.0518	0.0543
t3	1000	0.0486	0.0485	0.0493	0.0503	0.0532
garch	1000	0.0736	0.0755	0.0797	0.0798	0.0814

Table 6
Power for higher order wavelet decompositions

Rejection probabilities of tests with nominal levels of 5% against the restricted autoregressive model $rar(p)$

$$y_t = 0.1y_{t-p} + \epsilon_t \ , \qquad \text{for } p = 1, 2, 4 \ .$$

All simulations based on 10,000 replications.

model	N	GSM_2^g	GSM_3^g	GSM_4^g	GSM_5^g	GSM_6^g
rar(1)	100	0.1009	0.0673	0.0496	0.0408	0.0479
rar(2)	100	0.0912	0.0769	0.0614	0.0506	0.0553
rar(4)	100	0.0404	0.0645	0.0663	0.0602	0.0641
rar(5)	100	0.0389	0.0568	0.0591	0.0558	0.0632
rar(1)	300	0.2966	0.2226	0.1716	0.1359	0.1079
rar(2)	300	0.2376	0.2004	0.1643	0.1296	0.1049
rar(4)	300	0.0395	0.1308	0.1192	0.0971	0.0826
rar(1)	1000	0.8152	0.7504	0.6931	0.6338	0.5752
rar(2)	1000	0.7069	0.6558	0.5866	0.5292	0.4743
rar(4)	1000	0.0399	0.4084	0.3807	0.3280	0.2838

Appendix A. Proofs

Recall that the process $\{z_{m,t}\}$ is defined as the cross-product component of the square of each wavelet detail

$$z_{m,t} := \sum_{i=0}^{L-1} \sum_{j>i}^{L} h_{m,i} h_{m,j} y_{t-i} y_{t-j}$$

and that when there is no risk of confusion we omit the index m.

Proof of Proposition 2. Recall that on a measure space $\{X, \mu\}$, for any $f \in L_p(\Omega)$ and $g \in L_q(\Omega)$, the generalized Hölder inequality holds (see, for example, Reed and Simon, 1972, page 82):

$$\|fg\|_r \leq \|f\|_p \|g\|_q, \text{ with } p^{-1} + q^{-1} = r^{-1} \,, \tag{19}$$

in particular, if $p = q$, $\|fg\|_{p/2} \leq \|f\|_p \|g\|_p$. For the remainder of this proof let $\mathbb{E}[\cdot] = \mathbb{E}[\cdot|\mathcal{F}_{t-m}^{t+m}(\epsilon)]$. The following computation follows almost exactly the proof of Theorem 17.9 in Davidson (1995). Using the triangle inequality and the generalized Hölder inequality (19):

$$\begin{aligned}
&\|x_t y_t - \mathbb{E}x_t y_t\|_{p/2} \\
&= \|(x_t y_t - x_t \mathbb{E}y_t) + (x_t \mathbb{E}y_t - \mathbb{E}x_t \mathbb{E}y_t) - \mathbb{E}(x_t - \mathbb{E}x_t)(y_t - \mathbb{E}y_t)\|_{p/2} \\
&\leq \|x_t(y_t - \mathbb{E}y_t)\|_{p/2} + \|(x_t - \mathbb{E}x_t)\mathbb{E}y_t\|_{p/2} + \|\mathbb{E}(x_t - \mathbb{E}x_t)(y_t - \mathbb{E}y_t)\|_{p/2} \\
&\leq \|x_t\|_p \|y_t - \mathbb{E}y_t\|_p + \|x_t - \mathbb{E}x_t\|_p \|\mathbb{E}y_t\|_p + \|x_t - \mathbb{E}x_t\|_p \|y_t - \mathbb{E}y_t\|_p \\
&\leq \|x_t\|_p d_t^y \nu_m^y + \|y_t\|_p d_t^x \nu_m^x + d_t^x \nu_m^x d_t^y \nu_m^y \leq d_t \nu_m,
\end{aligned}$$

where $d_t = \max(\|x_t\|_p d_t^y, \|y_t\|_p d_t^x, d_t^x d_t^y)$ and $\nu_m = O(m^{-\min \phi_x, \phi_y})$. ☐

Proof of Theorem 4. Let $\{\epsilon_t\}$ be the driving mixing process of $\{y_t\}$. Since the NED property is preserved under linear combinations (Davidson, 1995, Theorem 17.8, page 267), $\{w_{m,t}\}$ is L_2-NED on ϵ_t. It follows that $\{w_{m,t}^2\}$ is L_1-NED on ϵ_t (Davidson, 1995, Theorem 17.9, page 268). Recall that

$$z_{m,t} = w_{m,t}^2 - \sum_{i=1}^{L_m} h_{m,i}^2 y_{t-i}^2 \,.$$

Again, since the linear combination of NED processes is a NED process, $\{z_{m,t}\}$ is L_1-NED. Notice that $\hat{\mathcal{E}}_{m,T} - 2^{-m}$ can be written in terms of z_t and y_t:

$$\hat{\mathcal{E}}_{m,T} - \frac{1}{2^m} = \frac{2\sum_t z_{m,t}}{\sum_t y_t^2} \,,$$

indeed

$$\begin{aligned}
\widehat{\mathcal{E}}_{m,T} &= \frac{\|w_m^T\|}{\|y_t^T\|} = \frac{\sum_{t=1}^T \left(\sum_{i=0}^{L_m} h_{m,i} y_t\right)^2}{\sum_{t=1}^T y_t^2} &(20)\\
&= \frac{\sum_{t=1}^T \left(\sum_{i=0}^{L_m} h_{m,i}^2 y_{t-i}^2 + 2\sum_{i=0}^{L_m-1} \sum_{j>i}^{L_m} h_{m,i} h_{m,j} y_{t-i} y_{t-j}\right)}{\sum_{t=1}^T y_t^2} &(21)\\
&= \frac{\sum_{i=0}^{L_m} h_{m,i}^2 \sum_{t=1}^T y_{t-i}^2}{\sum_{t=1}^T y_t^2} + \frac{2\sum_{t=1}^T z_{m,t}}{\sum_{t=1}^T y_t^2}
\end{aligned}$$

$$= \sum_{i=0}^{L_m} h_{m,i}^2 + \frac{2\sum_{t=1}^T z_{m,t}}{\sum_{t=1}^T y_t^2} = \frac{1}{2^m} + \frac{2\sum_{t=1}^T z_{m,t}}{\sum_{t=1}^T y_t^2} \tag{22}$$

Step (22) uses the fact that filtering is cyclic, therefore the sum $\sum_{t=1}^T y_{t-i}$ does not depend on i and is the same as the denominator $\sum_{t=1}^T y_t$. The last equality holds because the norm of a convolution is the product of the norms. Since $h_{m,t}$ is the cascade filter obtained by convolution of m filters with norm 1/2, the result holds. Now, the Law of Large Numbers for NED processes (see Davidson, 1995, , page 302) together with Slutsky's Theorem imply

$$\frac{2\sum_{t=1}^T z_{m,t}}{\sum_{t=1}^T y_t^2} \xrightarrow{p} 0$$

and the theorem is proven. □

In the stationary case, Theorem 4 follows easily from the Law of Large Numbers for NED processes (see Davidson, 1995, , page 302) and Slutsky's Theorem. Indeed,

$$\frac{\sum_{t=1}^n w_{m,n}^2}{\sum_{t=1}^n y_t^2} \xrightarrow{p} \frac{2^{-m}\sigma^2}{\sigma^2} = \frac{1}{2^m} ,$$

as $\mathbb{E}w_{m,n}^2 = 2^{-m}\sigma^2$ for all m and n.

Lemma 15. *Let $\{y_t\}$ be a stochastic sequence with zero means with finite joint fourth cumulants, i.e.*

$$\mathbb{E}[y_{t-i}y_{t-j}y_{t-k}y_{t-l}] < \infty ,$$

for all $i, j, k,$ and l such that $0 \le i < l < L$ and $0 \le k < l < L$. Then,

$$\text{var}(z_t) = \sum_{i=0}^{L-1}\sum_{j>i}^{L}\sum_{k=0}^{L-1}\sum_{l>k}^{L} h_i h_j h_k h_l \mathbb{E}(y_{t-i}y_{t-j}y_{t-k}y_{t-l})$$

and

$$\text{cov}(z_t, z_{t-s}) = \sum_{i=0}^{L-1}\sum_{j>i}^{L}\sum_{l=0}^{L-1}\sum_{k>l}^{L} h_i h_j h_{l-s} h_{k-s} \mathbb{E}(y_{t-i}y_{t-j}y_{t-s-l}y_{t-s-k})$$

Proof. The proof relies on a direct computation. First, we compute the variance:

$$\begin{aligned}
\text{var}(z_t) &= \text{var}\left(\sum_{i=0}^{L-1}\sum_{j>i}^{L} h_i h_j \; y_{t-i}y_{t-j}\right) \\
&= \text{cov}\left(\sum_{i=0}^{L-1}\sum_{j>i}^{L} h_i h_j \; y_{t-i}y_{t-j}, \sum_{k=0}^{L-1}\sum_{l>k}^{L} h_k h_l \; y_{t-k}y_{t-l}\right) \\
&= \sum_{i=0}^{L-1}\sum_{j>i}^{L}\sum_{k=0}^{L-1}\sum_{l>k}^{L} h_i h_j h_k h_l \text{Cov}(y_{t-i}y_{t-j}, y_{t-k}y_{t-l}) \\
&= \sum_{i=0}^{L-1}\sum_{j>i}^{L}\sum_{k=0}^{L-1}\sum_{l>k}^{L} h_i h_j h_k h_l \mathbb{E}(y_{t-i}y_{t-j}y_{t-k}y_{t-l}) , \qquad (23)
\end{aligned}$$

where at step (23) we used the fact that y_t has zero mean.

The autocovariances of $\{z_t\}$ are computed similarly. Let $h_l = 0$ for all $l > L$, then

$$\begin{aligned}
\operatorname{cov}(z_t, z_{t-s}) &= \operatorname{cov}(\sum_{i=0}^{L-1}\sum_{j>i}^{L} h_i h_j \; y_{t-i}y_{t-j}, \sum_{l=0}^{L-1}\sum_{k>l}^{L} h_{l-s}h_{k-s} \; y_{t-s-l}y_{t-s-k}) \\
&= \sum_{i=0}^{L-1}\sum_{j>i}^{L}\sum_{l=0}^{L-1}\sum_{k>l}^{L} h_i h_j h_{l-s} h_{k-s} \operatorname{Cov}(y_{t-i}y_{t-j}, y_{t-s-l}y_{t-s-k}) \\
&= \sum_{i=0}^{L-1}\sum_{j>i}^{L}\sum_{l=0}^{L-1}\sum_{k>l}^{L} h_i h_j h_{l-s} h_{k-s} \mathbb{E}(y_{t-i}y_{t-j}y_{t-s-l}y_{t-s-k}) \; . \qquad (24)
\end{aligned}$$

□

Proof of Proposition 8. Since $\{z_{m,t}\}$ is linear combination of processes of the form $\{y_t y_{t-i}\}$ and since the NED property is preserved under linear combinations, it follows that under under Assumption (B2), $\{z_{m,t}\}$ is L_2-NED of size $-1/2$ on ϵ_t.

To see that Assumption B1 implies condition (a) of Central Limit Theorem for NED processes (De Jong, 1997, page 358, Corollary 1) recall that from lemma 15

$$var(z_{m,t}) = \sum_{i=0}^{L_m}\sum_{j>1}^{L_m}\sum_{k=0}^{L_m}\sum_{l>1}^{L_m} h_i h_j h_k h_l \mathbb{E}(y_{t-i}y_{t-j}y_{t-k}y_{t-l}) \; .$$

Then,

$$\left\| \frac{y_{t-i}y_{t-j}y_{t-k}y_{t-l}}{\sum_{i=0}^{L_m}\sum_{j>1}^{L_m}\sum_{k=0}^{L_m}\sum_{l>1}^{L_m} h_i h_j h_k h_l \mathbb{E}(y_{t-i}y_{t-j}y_{t-k}y_{t-l})} \right\|_p \sim$$

$$\left\| \frac{\sum_{i=0}^{L_m}\sum_{j>1}^{L_m}\sum_{k=0}^{L_m}\sum_{l>1}^{L_m} h_i h_j h_k h_l y_{t-i}y_{t-j}y_{t-k}y_{t-l}}{\sum_{i=0}^{L_m}\sum_{j>1}^{L_m}\sum_{k=0}^{L_m}\sum_{l>1}^{L_m} h_i h_j h_k h_l \mathbb{E}(y_{t-i}y_{t-j}y_{t-k}y_{t-l})} \right\|_p = \left\| \frac{z_{t,m}^2}{\operatorname{var}(z_{m,t})} \right\|_p = \left\| \frac{z_{t,m}}{\sigma_{m,t}} \right\|_{2p} ,$$

which implies that $z_{m,t}/\sigma_{m,t}$ is L_r-bounded for $r = 2p > 2$.

Thus, $z_{m,t}$ satisfies the conditions of Central Limit Theorem for NED processes (De Jong, 1997, page 358, Corollary 1) and

$$\sum_{t=1}^{T} z_{m,t}/s_T(z) \xrightarrow{d} N(0,1) \; .$$

Therefore,

$$\frac{\sum_t y_t^2}{2 s_T(z)} \left(\hat{\mathcal{E}}_{m,T} - \frac{1}{2^m} \right) \xrightarrow{d} N(0,1)$$

$$\sqrt{\frac{T\sigma_T^4}{4 s_T^2(z)}} \left(\hat{\mathcal{E}}_{m,T} - \frac{1}{2^m} \right) \xrightarrow{d} N(0,1) \text{ , where } \sigma_T^2 = T^{-1}\sum_{t=1}^{T} \mathbb{E} y_t^2 \; .$$

□

Proof of Corollary 10. In order to prove Corollary 10, we require the following lemma.

Lemma 16. *Let $\{y_t\}$ be a stochastic sequence with zero means, identical variances $\sigma_t = \sigma$, and vanishing fourth order joint cumulants. Let $\{h_l\}_0^{L-1}$ be an L-dimensional vector. Then the stochastic* *sequence z_t is has variance*

$$\operatorname{var}(z_t) = \sigma^4 \sum_{i=0}^{L-1} \sum_{j>i}^{L} (h_i h_j)^2 \ , \tag{25}$$

and autocovariances

$$\operatorname{cov}(z_t, z_{t-s}) = \begin{cases} \sigma^4 \sum_{i=i_{min}}^{i_{max}} \sum_{j>i}^{j_{max}} h_i h_j h_{i-s} h_{j-s} \ , & \textit{if } s \leq L-1 \\ 0 \ , & \textit{otherwise} \end{cases} \tag{26}$$

where

$$i_{min} = \max(0, s) \ , \quad i_{max} = L - 1 + \min(0, s) \ , \quad j_{max} = L + \min(0, s) \ .$$

Proof. When fourth cumulants are zero, the fourth moment κ^{rstu} of y_t can be expressed in terms of the second moments $\kappa^{rs} = \sigma^2$. Such decomposition is valid whenever the fourth cumulant $\kappa^{r,s,t,u}$ is zero. Indeed (see for example McCullagh (1987))

$$\begin{aligned} \kappa^{rstu} &= \kappa^{r,s,t,u} + \kappa^{r,s,t}\kappa^{u}[4] + \kappa^{r,s}\kappa^{t,u}[3] + \kappa^{r,s}\kappa^{t}\kappa^{u} \\ &= \kappa^{r,s,t,u} + \kappa^{r,s}\kappa^{t,u}[3] \end{aligned}$$

where the the bracket notation $[n]$ indicates the number of terms in implicit summation over distinct partitions having the same block sizes. The second equality follows since $\kappa^s = 0$ as y_t is a zero mean sequence. Continuing form (23), since y_t is independently distributed and since $i \neq j$ and $k \neq l$ (from the second and fourth summations), the only non vanishing contributions in (23) correspond to the two possibilities $(i = k, j = l)$ and $(i = l, j = k)$. The second scenario never arises. Indeed, when $i = l$ and $j = k$, using $l > k$ (from the fourth summation)

$$i = l > k = j \quad \Longrightarrow \quad i > j,$$

which contradicts the condition $j > i$ (from the second summation). Let δ_{ij} be equal to 1 whenever $i = j$ and 0 otherwise. Thus,

$$\begin{aligned} &\sum_{i=0}^{L-1} \sum_{j>i}^{L} \sum_{k=0}^{L-1} \sum_{l>k}^{L} h_i h_j h_k h_l \mathbb{E}(y_{t-i} y_{t-j} y_{t-k} y_{t-l})(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}) \\ &= \sum_{i=0}^{L-1} \sum_{j>i}^{L} h_i^2 h_j^2 \mathbb{E}(y_{t-i}^2 y_{t-j}^2) \\ &= \sum_{i=0}^{L-1} \sum_{j>i}^{L} h_i^2 h_j^2 \mathbb{E}(y_{t-i}^2) \mathbb{E}(y_{t-j}^2) \\ &= \sigma^4 \sum_{i=0}^{L-1} \sum_{j>i}^{L} h_i^2 h_j^2 \ . \end{aligned}$$

A very similar computation yields the autocorrelation function γ_s:

$$\gamma_m(s) = \operatorname{Cov}\Big(\sum_{i=0}^{L-1} \sum_{j>i}^{L} h_i h_j \ y_{t-i} y_{t-j}, \sum_{l=0}^{L-1} \sum_{k>l}^{L} h_{l-s} h_{k-s} \ y_{t-s-l} y_{t-s-k}\Big) \tag{27}$$

$$= \sum_{i=0}^{L-1} \sum_{j>i}^{L} \sum_{l=0}^{L-1} \sum_{k>l}^{L} h_i h_j h_{l-s} h_{k-s} \mathrm{Cov}(y_{t-i} y_{t-j}, y_{t-s-l} y_{t-s-k})$$

$$= \sum_{i=0}^{L-1} \sum_{j>i}^{L} \sum_{l=0}^{L-1} \sum_{k>l}^{L} h_i h_j h_{l-s} h_{k-s} \mathbb{E}(y_{t-i} y_{t-j} y_{t-s-l} y_{t-s-k})$$

$$= \sum_{i=0}^{L-1} \sum_{j>i}^{L} \sum_{l=0}^{L-1} \sum_{k>l}^{L} h_i h_j h_{l-s} h_{k-s} \mathbb{E}(y_{t-i} y_{t-j} y_{t-s-l} y_{t-s-k}) (\delta_{i,s+l} \delta_{j,s+k} + \delta_{i,s+k} \delta_{j,s+l})$$

$$= \sum_{i=0}^{L-1} \sum_{j>i}^{L} \sum_{l=0}^{L-1} \sum_{k>l}^{L} h_i h_j h_{l-s} h_{k-s} \mathbb{E}(y_{t-i} y_{t-j} y_{t-s-l} y_{t-s-k}) \delta_{i,s+l} \delta_{j,s+k} \tag{28}$$

$$= \sum_{i=i_{\min}}^{i_{\max}} \sum_{j>i}^{j_{\max}} h_i h_j h_{i-s} h_{l-s} \mathbb{E}(y_{t-i}^2) \mathbb{E}(y_{t-j}^2)$$

$$= \sigma^4 \sum_{i=i_{\min}}^{i_{\max}} \sum_{j>i}^{j_{\max}} h_i h_j h_{i-s} h_{l-s} \ .$$

where

$$i_{\min} = \max(0, s) \ , \quad i_{\max} = L - 1 + \min(0, s) \ , \quad j_{\max} = L + \min(0, s) \ .$$

At equality (28) we used the fact that the contribution of $\delta_{i,s+k} \delta_{j,s+l}$ is zero. The argument is the same as for the analogous contribution to $\gamma_m(0)$.

Notice that the autocovariance $\gamma(s)$ is zero when $i_{\min} > i_{\max}$. For $s > 0$, this condition holds when

$$max(0, s) > L - 1 + \min(0, s)$$

$$s > L - 1 \ .$$

In particular, the sequence z_t is a $(L-1)$-dependent sequence (i.e. z_t is independent of z_{t-l} for $l > L - 1$). □

Using Equation 21 and the fact that $\hat{\mathcal{E}}_{m,T} \xrightarrow{p} \frac{1}{2^m}$ (see Proposition 8) we can write

$$\sqrt{T} \left(\hat{\mathcal{E}}_{m,T} - \frac{1}{2^m} \right) = \sqrt{T} \frac{2 \sum_{t=1}^{T} \sum_{i=0}^{2^m-2} \sum_{j>i}^{2^m-1} h_i h_j y_{t-i} y_{t-j}}{\sum_{t=1}^{T} y_t^2}$$

$$= \sqrt{T} \frac{\sum_{t=1}^{T} 2 z_t}{\sum_{t=1}^{T} y_t^2} = \frac{\sqrt{T} (2 \bar{z}_t)}{\frac{1}{T} \sum_{t=1}^{T} y_t^2}$$

$$\xrightarrow{d} \frac{\mathcal{N}\left(0, 4 \sum_{j=-L+1}^{L-1} \gamma(j)\right)}{\sigma^2} \sim \frac{\mathcal{N}\left(0, \sigma^4 a_n\right)}{\sigma^2} \sim \sqrt{a_n} \mathcal{N}(0, 1) \ . \tag{29}$$

In step (29) we used the Continuous Mapping Theorem and the Central Limit Theorem for stationary time series (see Hamilton, 1994, Theorem 7.11). Independence of a_m from σ follows directly from Equations (25) and (26). □

Proof of Theorem 13. Consider the vector $(GS_{1,T}, \dots, GS_{N,T})$.

$$\begin{pmatrix} GS_{1,T} \\ \vdots \\ GS_{N,T} \end{pmatrix} = \begin{pmatrix} \sqrt{\frac{T}{a_1}} \left(\hat{\mathcal{E}}_{1,T} - \frac{1}{2^1} \right) \\ \vdots \\ \sqrt{\frac{T}{a_N}} \left(\hat{\mathcal{E}}_{N,T} - \frac{1}{2^N} \right) \end{pmatrix} = \frac{\sqrt{T}}{\sum_{t=1}^T y_t^2} \begin{pmatrix} \frac{1}{\sqrt{a_m}} \sum_{t=1}^T z_{1,t} \\ \vdots \\ \frac{1}{\sqrt{a_N}} \sum_{t=1}^T z_{N,t} \end{pmatrix} = \frac{\sqrt{T}}{\frac{1}{T} \sum_{t=1}^T y_t^2} \begin{pmatrix} \frac{1}{\sqrt{a_1}} \bar{z}_{1,T} \\ \vdots \\ \frac{1}{\sqrt{a_N}} \bar{z}_{N,T} \end{pmatrix}$$

Let $\boldsymbol{q}$ be the column N-vector with coordinates $\frac{1}{\sqrt{a_i}}$. Let $\operatorname{diag}(v)$ be the square matrix with v on the main diagonal and zero everywhere else. By definition $\operatorname{diag}(\boldsymbol{q}) \left(\sum_{s \in \mathbb{Z}} \boldsymbol{\Gamma}(s) \right) \operatorname{diag}(\boldsymbol{q}) = \sigma^2 A$. Indeed,

$$\begin{pmatrix} \frac{1}{\sqrt{a_1}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{a_2}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sqrt{a_N}} \end{pmatrix} \begin{pmatrix} \sigma^4 a_1 & \sigma^4 a_{12} & \cdots & \sigma^4 a_{1N} \\ \sigma^4 a_{21} & \sigma^4 a_2 & \cdots & \sigma^4 a_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma^4 a_{N1} & \sigma^4 a_{N2} & \cdots & \sigma^4 a_N \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{a_1}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{a_2}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sqrt{a_N}} e \end{pmatrix} = \sigma^4 A$$

The joint asymptotic distribution of the vector of multi-scale energy ratios is

$$\begin{pmatrix} GS_{1,T} \\ \vdots \\ GS_{N,T} \end{pmatrix} \xrightarrow{d} \frac{1}{\sigma^2} \mathcal{N} \left(0, \operatorname{diag}(\boldsymbol{q}) \left(\sum_{j=-\infty}^{+\infty} \boldsymbol{\Gamma}(j) \right) \operatorname{diag}(\boldsymbol{q}) \right) \sim \mathcal{N}\left(0, A\right) .$$

□