Multi-scale tests for serial correlation

Ramazan Gençay *, Daniele Signori

Department of Economics, Simon Fraser University, 8888 University Drive, Burnaby, British Columbia, V5A 1S6, Canada

A R T I C L E   I N F O

Article history:
Received 8 January 2013
Received in revised form 5 August 2014
Accepted 18 August 2014
Available online 16 September 2014

JEL classification:
C1
C2
C12
C22
C58
F31
G0
G1

Keywords:
Serial correlation
Wavelets
Independence
Discrete wavelet transformation
Maximum overlap wavelet transformation
Variance ratio test
Variance decomposition

A B S T R A C T

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.

1. 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,

\[
\text{var}(y) = 2 \int_0^{1/2} S_y(f) \, df,
\]

which, paraphrasing, says that the contribution of the frequencies in a small interval \(\Delta f\) 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 \(\delta(a, b)\), defined as the ratio of the total variance contributed by the frequency band \((a, b)\) under the null of no serial correlation, \(\delta(a, b)\) 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 \(\delta\) – imposes practical limitations. We resort to wavelet analysis to address this...
need. For frequency bands of a particular form, the numerator of the statistic $E$ is a well known quantity, the wavelet variance, 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 appropriate to refer to $E(a, b)$ 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 are 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 moments, while at the same time the asymptotic theory is developed in great generality. Third, ours is the first test of serial correlation that directly utilizes the wavelet coefficients of the observed time series to construct the wavelet-based test statistics. The tests we design generalize, on the one hand, variance ratios tests (Lo and MacKinlay, 1988), and 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 density, the rate of convergence issues relating to the nonparametric spectral density are not of the first order of importance.

One of the well-known time-domain portmanteau tests for serial correlation is the Box and Pierce 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 observations 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 is violated, possibly leading to a 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 et al. (2002). Robust inference can also be achieved using bootstrapping methods. Building on the block bootstrap inference for autocorrelations of Romano and Thombs (1996), Horowitz et al. (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 the tests 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 2, 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 are obtained. In Section 3, we introduce and motivate our tests. In Section 4 we study their large sample distributions. In Section 5, we analyze the small sample properties through several Monte Carlo simulations. A brief conclusion follows afterwards.

2. Preliminaries

Let $y_t$ be a stochastic sequence with $E(y_t) = 0$ and $\text{var}(y_t) = \sigma^2_t$. If $y_t$ is homoskedastic, that is $\sigma^2_t = \sigma^2$ for all $t$, and uncorrelated, that is $\text{cov}(y_s, y_t) = 0$ 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, $H_0 : \text{cov}(y_s, y_t) = 0$ for all $s \neq t$, against correlated alternatives, $H_1 : \text{cov}(y_s, y_t) \neq 0$ 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 $y^T$, or simply $y$, leaving $T$ understood when there is no chance for confusion. Throughout the paper we impose periodic boundary conditions on $[y_t]$, that is

$$y_t \equiv y_{t+T},$$

and we define $s^2_y(y)$ as

$$s^2_y(y) = \sum_{i=1}^{n} \text{var}(y_i) + 2 \sum_{i=1}^{n-1} \text{cov}(y_i, y_{i-k}).$$

2.1. Wavelet transformations

The multiscale variance ratio at the core of our test is built from a specific frequency decomposition of the variance. The components of such decomposition can be estimated using the Maximum Overlap Discrete Wavelet Transform (MODWT), which we introduce in this section. A vector $[h]$ is a quadrature mirror relationship of $[y]$, the convolution of $[h]$, and $[y]$ is the sequence

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

where we define $h_0 = 0$ for all $l < 0$ and $l \leq L$.

A wavelet filter is a linear time invariant filter $[h]$ of length $L$, such that for all $n \neq 0$:

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

In words, $h$ sums to zero, has norm 1/2, and is orthogonal to its even shifts. The natural complement to the wavelet filter $[h]$ is the scaling filter $[g]$ determined by the quadrature mirror relationship $g_l = (-1)^l h_{L-l-1}$ for $l = 0, \ldots, L - 1$.

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

$$\|x_t - \mathbb{E}[x_t | F_{t-m}^e(\epsilon)]\|_p \leq d_t v_m$$

where $v_m \to 0$ as $m \to \infty$ and $d_m$ is a sequence of positive real numbers such that $d_t = O(||x_t||_p)$.

Any process $x_t$ that satisfies Definition 1 will be referred to as “Lp-NED on $\epsilon_t$". 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 Ibrahimov (1962). As pointed out by Davidson (1995), near-epoch dependence is not an alternative to mixing assumptions, instead it allows the establishment of useful memory properties of $x_t$ in terms of those of $\epsilon_t$.

When the innovation process $\epsilon_t$ is mixing, powerful laws of large numbers and central limit theorems can be established for NED processes. 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 $-\phi_1$ and $-\phi_2$ respectively, then $x_t y_t$ is $L_p/2$-NED of size $-\min(\phi_1, \phi_2)$ on $\{\epsilon_t\}$.

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.


7 This section closely follows Gençay et al. (2001), see also Percival and Walden (2000, Chap. 5). It is common in the literature to distinguish the objects related to the Discrete Wavelet Transform from those related to the Maximum Overlap Discrete Wavelet Transform by placing a tilde (~) 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 notation $a - b \mod T$ stands for “$a - b$ modulo $T$". If $j$ is an integer such that $1 \leq j \leq T$, then $j \mod T = j$. If $j$ is another integer, then $j \mod T = j + nT$ where $nT$ is the unique integer multiple of $T$ such that $1 \leq j + nT \leq T$.\footnote{4}

4 The notation $a - b \mod T$ stands for “$a - b$ modulo $T$". If $j$ is an integer such that $1 \leq j \leq T$, then $j \mod T = j$. If $j$ is another integer, then $j \mod T = j + nT$ where $nT$ is the unique integer multiple of $T$ such that $1 \leq j + nT \leq T$.\footnote{4}

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.


7 This section closely follows Gençay et al. (2001), see also Percival and Walden (2000, Chap. 5). It is common in the literature to distinguish the objects related to the Discrete Wavelet Transform from those related to the Maximum Overlap Discrete Wavelet Transform by placing a tilde (~) 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 Eq. (3):

\[
\sum_{i=0}^{l-1} g_i = 1, \quad \sum_{i=0}^{l-1} g_i^2 = 1/2,
\]

\[
\sum_{l=-\infty}^{\infty} g_l g_{l+2n} = 0, \quad \sum_{l=-\infty}^{\infty} g_l h_{l+2n} = 0,
\]

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 constraints, 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 \( g_l \) are 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:

\[
w = W y
\]

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

\[
W = \begin{bmatrix} W_1, W_2, \ldots, W_M, V_M \end{bmatrix},
\]

whose action is defined in terms of the wavelet filter \( h_l \) and scaling filter \( g_l \). Specifically,

\[
(W_m y)_i = \sum_{l=0}^{l_m} h_{m,l} v_{m,t-l \mod T},
\]

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

\[
h_m = h * g * \cdots * g,
\]

where \( g \) is the scaling filter and \( * \) denotes a convolution.

The MODWT of the observed time series \( y \) can be organized into \( M + 1 \) vectors of length \( T \):

\[
w = (w_1, \ldots, w_M, v_M)’,
\]

where \( M \leq \log_2 T \) is the decomposition level of the MODWT. In practice, \( w \) 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 filters to the data to obtain the first level wavelet and scaling coefficients:

\[
w_{1,t} = (w_1)_t = \sum_{l=0}^{l-1} h_l y_{t-l \mod T} \quad \text{and}
\]

\[
v_{1,t} = (v_1)_t = \sum_{l=0}^{l-1} g_l y_{t-l \mod T} \quad \text{for all } t = 1, \ldots, T.
\]

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

\[
w_{m+1,t} = (w_1)_t = \sum_{l=0}^{l-1} h_l v_{m,t-l \mod T} \quad \text{and}
\]

\[
v_{m+1,t} = (v_1)_t = \sum_{l=0}^{l-1} g_l v_{m,t-l \mod T} \quad \text{for all } t = 1, \ldots, T.
\]

Keeping all vectors of wavelet coefficients, and the level \( M \) scaling coefficients, we obtain the decomposition of Eq. (5).

3. Multi-scale variance ratios

Consider the general variance ratio

\[
\varepsilon(a, b) = 2 \int_a^b S_y(f) df / \text{var}(y).
\]

The numerator of \( \varepsilon(a, b) \) can, for specific intervals, be expressed in terms of the wavelet variance. Indeed, neglecting the leakage of the wavelet filter, the following approximation holds\(^8\)

\[
\text{wvar}(y) \approx 2 \int_{1/2}^{1/2+1} S_y(f) df.
\]

For \( m = 1 \), the integral in Eq. (7) corresponds to the area \( \varepsilon_1 \) in Fig. 1. Formally, the wavelet variance for a stationary process \( y \) is defined as

\[
\text{wvar}(y) \equiv \text{var}(w_{m,t}).
\]

From Eq. (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 \( h_l \) (see Brockwell and Davis, 2009, Page 121, Eq. 4.4.3.). It follows that

\[
\text{wvar}(y) = \int_{-1/2}^{1/2} S_m(f) df = \int_{-1/2}^{1/2} |H_m(f)|^2 S_y(f) df.
\]

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

\[
\text{wvar}(y) = \sigma_y^2 \int_{-1/2}^{1/2} |H_m(f)|^2 df = \sigma_y^2 \|h_m\|_2^2
\]

\[
= \sigma_y^2 \|g\|_2 \prod_{i=1}^{m-1} \|h_i\|_2 = \sigma_y^2 2^{-m}.
\]

The second equality uses Parseval’s identity, and the last equality follows from the normalization Eq. (3). In conclusion, we have proved the following:

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

\[
\varepsilon_m(y) \equiv \frac{\text{wvar}(y)}{\text{var}(y)} = \frac{1}{2^m}.
\]

When there is no risk of confusion, we will write \( \varepsilon_m \) for \( \varepsilon_m(y) \). In the remainder of this section we introduce a family of statistics that detects serial correlation by testing the implications of Theorem 3.

---

\(^8\) See Percival and Walden (2000), Eq. (297a), page 297.
3.1. Sample multiscale variance ratios: Scale one

The Maximum Overlap Discrete Wavelet Transform (MODWT) consists of a set of linear filters, which given a time series of a given length, generates a collection of vectors of the same length as the data. 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.\(^9\)

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 \(h_1^{(1)} = (1/2, -1/2)\) and the corresponding scaling filter \(g_1^{(1)} = (1/2, 1/2)\). The wavelet and scaling coefficients of a time series \(\{y_t\}_{t=1}^T\) are given by

\[
w_{1,t} = \frac{1}{2}(y_t - y_{t-1}), \quad t = 1, 2, \ldots, T,\]

\[
v_{1,t} = \frac{1}{2}(y_t + y_{t-1}), \quad t = 1, 2, \ldots, T.
\]

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

\[
\hat{E}_{1,T} = \frac{\text{wvar}_T y}{\text{var} y} = \frac{\sum_{t=1}^T w_{1,t}^2}{\sum_{t=1}^T v_{1,t}^2}.
\]

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

The definition of the variance ratio \(\hat{E}_{1,T}\) can be applied to the wavelet decomposition obtained from a generic filter wavelet \(h_1^{(1)}\). As before, we expect \(\hat{E}_{1,T}\) to be close to 1/2 under \(H_0\).

When the Haar wavelet is used, the scale one energy ratio is equivalent to the variance ratio of period 2 up to an affine transformation. This result follows after exploiting the orthogonality of wavelet decomposition and writing the energy ratio \(\hat{E}_{1,T}\) in terms of the scaling coefficients, rather than the wavelet coefficients,

\[
\hat{E}_{1,T} = \frac{\sum_{t=1}^T w_{1,t}^2}{\sum_{t=1}^T v_{1,t}^2} = 1 - \frac{\sum_{t=1}^T y_{1,t}^2}{\sum_{t=1}^T y_{2,t}^2} = 1 - \frac{\sum_{t=1}^T (y_t + y_{t+1})^2}{4 \sum_{t=1}^T y_{2,t}^2} \xrightarrow{p} 1 - \frac{1}{2} \text{VR}(2)
\]

where \(\text{VR}(k)\) is the \(k\)-period variance ratio (Lo and MacKinlay, 1988) and convergence is in probability.

3.2. Sample multiscale variance ratios: Scale \(m\)

The intuitive results that we discussed in the previous section 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

\[
\hat{E}_{m,T} = \frac{\text{wvar}_m y}{\text{var} y} = \frac{\sum_{t=1}^T w_{m,t}^2}{\sum_{t=1}^T y_{m,t}^2},
\]

where \(w_m\) are the \(m\)th level wavelet coefficients of \(y\).

To formalize the above discussion, we need to prove that \(\hat{E}_{m,T}\) 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 \(\hat{E}\) 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 a stochastic sequence that is \(L_1\) bounded for \(r > 2\) and \(L_r\)-NED on an \(\alpha\)-mixing process for \(p \geq 2\).

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

\[
\hat{E}_{m,T} \xrightarrow{p} \frac{1}{2^m}.
\]

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

\[
x_t = \alpha \epsilon_t, \quad \sigma_t^2 = \omega + \beta \sigma_{t-1}^2 + \gamma \epsilon^2_{t-1}
\]

for some real numbers \(\omega, \beta, \gamma\). Hansen (1991) shows that

\[
\mathbb{E}\left[(\beta + \gamma \epsilon^2_{t-1})^p | \mathcal{F}_{t-1}\right] \leq c < 1 \quad \text{a.s. for all } t,
\]

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 \(\beta^2 + 2\alpha \beta \mu_2^2 + \alpha^2 \mu_4^2 < 1\) \quad \text{a.s. for all } t,

in which \(\mu_2^2 = \mathbb{E}(\epsilon^2_{t-1})\) is the conditional kurtosis.

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

\[
x_t = \alpha \epsilon_t, \quad \sigma_t^2 = \omega + \sum_{i=1}^\infty \alpha_i \epsilon_{t-i}^2.
\]

This specification is called the ARCH(\(\infty\)) model. It encompasses several nonlinear time series, including GARCH (Bollerslev, 1986), IGARCH (Engle and Bollerslev, 1986), and FIGARCH (Baillie et al., 1996).

Assume that \(\mathbb{E}\epsilon^4 \exists\) and \(\lim_{t \to \infty} \sum_{i=1}^\infty \alpha_i < (\mathbb{E}\epsilon^4)^2\). Davidson (2004) shows that if \(0 \leq \alpha_i \leq \lambda_0^{-1}\) for some \(\lambda_0 > 0\), then \(x_t\) is \(L_\infty\)-NED on \(\epsilon_t\), of size \(-\lambda_0\).

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

\[
x_t = \sum_{j=1}^m \alpha_x \epsilon_{t-j} + \sum_{j=1}^m \beta_x \epsilon_{t-j} \epsilon_{t-1-j} + \sum_{j=1}^m \gamma \epsilon_{t-j}.
\]

This parametric family is referred to as BL(p, r, m, 1) and it is discussed in detail in Priestley (1988, Chapter 4). Davidson (2002)
indicates 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
\[
x_t = \beta x_{t-2} \epsilon_{t-1} + \epsilon_t \sim \text{i.i.d}(0, 1).
\]
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 \( \hat{\delta}_{m,t} \).

### 4. Asymptotic analysis

In the remainder of the paper, the process \( \{z_{m,t}\} \) is defined as the cross-product component of the square of each wavelet coefficient
\[
z_{m,t} := \sum_{l=0}^{L_m} h_{m,l} h_{m,l} y_{t-l} y_{t-l}.
\]
When there is no risk of confusion, we omit the index \( m \). Our next result establishes the asymptotic distribution of the wavelet variance ratio \( \hat{\delta}_{m,t} \).

**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 \( 0 \leq k < l \leq L_m \), \( \{y_{t-i-j} y_{t-l} y_{t-l} / M_{4,t}\} \) is uniformly \( L_r \)-bounded for \( r > 1 \), where
\[
M_{4,t} := \sum_{i=0}^{L_m} \sum_{j=0}^{L_m} \sum_{k=0}^{L_m} \sum_{l=0}^{L_m} h_i h_j h_k h_l \mathbb{E}(y_{t-i-j} y_{t-l} y_{t-l}).
\]

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

B3. \( \text{var}(z_t) \sim t^p \) and \( s^2_t(z) \sim n^{1+p}, \beta \leq \gamma \).

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 the rate of growth of 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 heterogeneously) 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 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
\[
\sqrt{\frac{T \sigma^4}{4 s^2_t(z)}} \left( \hat{\delta}_{m,t} - \frac{1}{2m} \right) \xrightarrow{d} N(0, 1),
\]
where \( s^2_t(z) \) is defined in the same way as in Eq. (1).

**Proposition 8** suggests the following definition for a test statistic
\[
G_{S_m} := \sqrt{\frac{T \sigma^4}{4 \text{aivar}(z)}} \left( \hat{\delta}_{m,t} - \frac{1}{2m} \right).
\]

**Fig. 2.** Let \( \{\epsilon_t\} \) be identically and independently normally distributed. Let \( x_t = \sigma_t \epsilon_t \) and \( \sigma_t^2 = \alpha + \beta \sigma_{t-1}^2 + \gamma \sigma_{t-1}^4 \) for some real numbers \( \alpha, \beta, \) and \( \gamma \). The pink region depicts the solution to the inequality \( \beta^4 + 2\alpha\beta + \gamma \beta^2 < 1 \). In this case \( x_t \) satisfies Assumption A. The purple region depicts the solution to the inequality \( \beta^4 + 4\beta^3 \alpha + \beta^2 \alpha^2 + 6\beta \alpha^3 + 105 \alpha^4 \leq 1 \). In this case \( x_t \) satisfies Assumption B.

The asymptotic distribution of \( G_{S_m} \) is given as follows:

**Example 9 (GARCH\((1, 1)\) with \( \alpha \)-mixing innovations).** Consider again Example 5. A straightforward generalization 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
\[
\beta^4 + 4\beta^3 \alpha + 18\beta^2 \alpha^2 + 60\beta \alpha^3 + 105 \alpha^4 \leq 1 \quad \text{a.s. for all } t,
\]
in which \( \mu_t := \mathbb{E}[t^{1/2}] \). Let \( \epsilon_t \sim \text{i.i.d}(0, 1) \), then 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 Fig. 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 the test can be computed exactly. More specifically, let \( X_t^{ijkl} = (X_{t-i}, X_{t-j}, X_{t-k}, X_{t-l}) \) and \( \xi \) be a vector in \( \mathbb{R}^4 \) and \( M(\xi) \) be the moment generating function \( X_t^{ijkl} \)
\[
M_4^{ijkl}(\xi) = \exp(\mathbb{E}^{X_t^{ijkl}}),
\]
which has as coefficients of its Taylor expansion
\[
M(\xi) = \sum_a \xi_a \mathbb{E}^a + \frac{1}{2!} \sum_{a,b} \xi_a \xi_b \mathbb{E}^{ab} + \frac{1}{3!} \sum_{a,b,c} \xi_a \xi_b \xi_c \mathbb{E}^{abc} + \cdots
\]
\footnote{See Andrews (1991) for a general theory of kernel estimators. Among several approaches and kernel choices, we did not find significant differences pointing to a strong preference for one method over the others.}
The cumulants of $\chi^2_{ijkl}$ are defined as the coefficients $\kappa^{(n)}$ 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 a white noise process with zero fourth order cumulants. Then
\[
\sqrt{\frac{T}{a_m}} (\hat{\xi}_{m,T} - \frac{1}{2^m}) \to \mathcal{N}(0, 1)
\]
with
\[a_m = \sum_{i=0}^{i_{\text{max}}} \sum_{j=0}^{j_{\text{max}}} \sum_{l=i}^{l_{\text{max}}} h_m \cdot h_m, \quad \text{where } h_m \text{ is the wavelet filter used in the construction of } \hat{\xi}_m \text{ and}
\]
\[i_{\text{min}} = \max(0, s), \quad i_{\text{max}} = \min(l_m, l_n + s) - 2, \quad j_{\text{max}} = \min(l_m, l_n + s) - 1 .
\]

The computation of $a_m$ is trivial but tedious.\(^{11}\) The following Corollary contains several asymptotic results for the Haar filter.

**Corollary 11.** Let $\{y_t\}$ be a white noise process with zero fourth order cumulants and let $h_1$ be the Haar filter $(\frac{1}{2}, -\frac{1}{2})$. Then, $GSM_m$ test statistics for the scales $1$ to $4$ are
\[
\sqrt{\frac{4T}{15}} (\hat{\xi}_{1,T} - \frac{1}{2}), \quad \sqrt{\frac{32T}{3}} (\hat{\xi}_{2,T} - \frac{1}{4}), \quad \sqrt{\frac{256T}{15}} (\hat{\xi}_{3,T} - \frac{1}{8}), \quad \sqrt{\frac{2048T}{71}} (\hat{\xi}_{4,T} - \frac{1}{16}) ,
\]
respectively. Their asymptotic distribution is the standard normal.

### 4.1. Multivariate multiscale tests

Each test in the $GS$ family has 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 remainder 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
\[
\Sigma_{ij} = \frac{\text{acov}(GZ_j, GZ_i)}{\text{avar}(GZ_i) \text{avar}(GZ_j)} .
\]
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 $\chi^2_N$ distribution.

The proof of this result follows closely the proof of Proposition 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 $\chi^2$ 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 $\Sigma$, $X^T \Sigma^{-1}X$ is distributed as a $\chi^2_n$. Accordingly, we define the test statistic
\[
GSM_N = (GS_1, \ldots, GS_N) \Sigma^{-1}(GS_1, \ldots, GS_N)^T,
\]
whose asymptotic distribution is a $\chi^2_N$.

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
\[
\gamma_{m,n}(s) = \sqrt{\frac{a_m}{\sigma^4}} \sum_{i=1}^{i_{\text{max}}} \sum_{j=1}^{j_{\text{max}}} h_m \cdot h_m, \text{ with}
\]
\[i_{\text{min}} = \max(0, s), \quad i_{\text{max}} = \min(l_m, l_n + s) - 2, \quad j_{\text{max}} = \min(l_m, l_n + s) - 1 .
\]
Define, furthermore,
\[
a_{m,n} = \frac{1}{\sqrt{a_m a_n}} \gamma_{m,n}(s)
\]
and let $A$ be an $N \times N$ matrix with ones on the main diagonal and off-diagonal entries
\[
A_{m,n} = \frac{a_{m,n}}{\sqrt{a_m a_n}} .
\]

**Corollary 13.** Let $\{y_t\}$ be a white noise process with zero fourth order cumulants. 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.** Let $\{y_t\}$ be a white noise process with zero fourth order cumulants and let $h_1$ be the Haar filter $(\frac{1}{2}, -\frac{1}{2})$. Then, \[\frac{GS_1}{GS_2} \to \mathcal{N}(0, A),
\]
with
\[
A = \begin{pmatrix}
1 & -1/\sqrt{6} & -5/\sqrt{6} \\
-1/\sqrt{6} & 1 & 2/\sqrt{360} \\
-5/\sqrt{6} & 2/\sqrt{360} & 1
\end{pmatrix}.
\]

### 5. Asymptotic local power and finite sample performance

In this section, we evaluate the GSM test family using two criteria: firstly, we study its asymptotic local power. Secondly, we analyze its finite sample performance when the Haar filter is used.

First, we illustrate through an example, the inconsistency of the family $GSM_N$. Consider the spectrum $S_y$ of the stochastic process $y$:
\[
S_y(f) = \begin{cases}
\frac{1}{2} + \frac{1}{4} \sin(8\pi f) & \text{if } f \in \left(\frac{1}{4}, \frac{1}{2}\right) \\
\frac{1}{2} + \frac{1}{8} \sin(16\pi f) & \text{if } f \in \left(\frac{1}{8}, \frac{1}{4}\right) \\
\cdots 
\end{cases}
\]
(14)

The spectrum $S_y(y)$ is shown in Fig. 3 and it is non-flat; hence, the corresponding time series is correlated. At the same time the area underneath $S_y$ within any of the blocks considered by the

\(^{11}\) 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.
dyadic decomposition of the frequency space is consistent with the equipartition of the 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 $H \ast S_x = S_y$, where $\ast$ 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) 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):

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

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

$$ \frac{P(f)}{Q(f)} = \frac{1}{2} = 0, \quad \text{and let } f_{\text{min}} \text{ be } f_{\text{min}} = \min_{f > 0} (f \in \mathcal{F}). $$

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

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

then the test ${\text{GSM}}_k$ 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 $S_k$ on the process $y$ with spectrum $S_y$ is $\mathbb{E}[{\text{GSM}}_k(X_f)] \neq 0$.

5.1. Asymptotic local power

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

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

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

$$ \delta_k = \int_{-2^{-k-1}}^{2^{-k-1}} \left| H_k(f) \right|^2 S(f) \, df $$

and that, in probability, $\delta_k \rightarrow \varepsilon_k$ and, therefore, $\text{GSM}_k(X) \rightarrow \varepsilon_k/\varepsilon_0 - 1/2^k$. Let

$$ \text{TGSM}_N = \sqrt{TE}(\text{GSM}_1(X), \ldots, \text{GSM}_N(X)) $$

$$ = (\varepsilon_1/\varepsilon_0 - 1/2^1, \ldots, \varepsilon_N/\varepsilon_0 - 1/2^N). $$

Since the estimator of the covariance matrix of $(\text{GSM}_1(X), \ldots, \text{GSM}_N(X))$ is consistent under $H_1$, it follows that the distribution of the test $\text{GSM}_N$ is the non-central $\chi^2_N(c)$, where

$$ c = T \ast \text{GSM}_N(X) $$

$$ = \text{TGSM}_N^\prime \text{avar}(\text{TGSM}_1(X))^{-1} \text{TGSM}_1(X). $$

Therefore the asymptotic local power of $\text{GSM}_N$ 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 $\chi^2_N$ distribution.

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

$$ y_1 = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \epsilon_t $$

where $\epsilon_t$ is Gaussian white noise. The first and second plots (left and center) depict the asymptotic power of the univariate tests $\text{GSM}_1$ and $\text{GSM}_2$ 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 consistent. The third plot (right) shows the asymptotic power of the bivariate test $\text{GSM}_2$: 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).

From the asymptotic rejection surfaces of Fig. 4, it follows that the inconsistency loci (depicted in black) and the asymptotic rejection surfaces are a function of the centrality parameter $c$, which, in turn, is a function of the time series $y$ and the wavelet filter. Therefore, given the complexity of such dependence, it is not possible to make general statements regarding relative asymptotic efficiency; rather, the choice of a family of alternative hypotheses will play a key role in assessing asymptotic relative efficiency.
Fig. 4. Asymptotic rejection rates at the nominal level \( \alpha = 0.10 \) against a two-dimensional AR family with Haar filter. 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 \( \alpha_1 = \alpha_2 = 0 \).

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

<table>
<thead>
<tr>
<th></th>
<th>( N(0,1) )</th>
<th>( N(0,1) )-GARCH(1,1)</th>
<th>( t_0 )-GARCH(1,1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( T )</td>
<td>100 300 1000 5000</td>
<td>100 300 1000 5000</td>
<td>100 300 1000 5000</td>
</tr>
<tr>
<td>( GSM_2^\varphi )</td>
<td>0.82 0.92 0.87 1.04</td>
<td>1.32 1.81 1.62 1.86</td>
<td>1.76 2.84 4.04 5.16</td>
</tr>
<tr>
<td>( GSM_2^A )</td>
<td>2.75 1.47 1.12 1.21</td>
<td>2.60 1.65 1.12 1.22</td>
<td>2.14 1.04 1.20 1.12</td>
</tr>
<tr>
<td>( GSM_2 )</td>
<td>5.25 2.75 1.87 1.54</td>
<td>5.13 2.72 1.64 1.55</td>
<td>4.26 2.14 1.70 1.25</td>
</tr>
<tr>
<td>( Q_5 )</td>
<td>0.86 0.88 0.95 1.06</td>
<td>1.22 1.94 1.81 2.17</td>
<td>1.81 3.59 5.67 7.51</td>
</tr>
<tr>
<td>( Q_{10} )</td>
<td>0.90 1.02 1.03 1.08</td>
<td>1.60 2.34 2.26 2.70</td>
<td>1.95 4.24 7.08 10.51</td>
</tr>
<tr>
<td>( Q_{20} )</td>
<td>0.88 1.15 1.02 1.14</td>
<td>1.51 2.40 2.59 2.63</td>
<td>1.59 4.98 8.51 12.08</td>
</tr>
<tr>
<td>( EL )</td>
<td>2.73 2.28 1.71 1.23</td>
<td>2.65 2.68 1.80 1.36</td>
<td>2.17 1.94 1.78 1.25</td>
</tr>
</tbody>
</table>

5.2. Monte Carlo simulations

Feasible tests are obtained from Theorem 12 replacing the matrix \( \Sigma \) 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 Corollaries 11 and 14. We denote these statistics \( GS^\varphi \) and \( GS^A \) in the univariate case and multivariate case, respectively.
2. Second, each level \( GS \) can be computed using an estimator of the long run variance (again, we use the Newey and
Table 2
Rejection rates under the null hypothesis at 5% nominal level.

<table>
<thead>
<tr>
<th>Model</th>
<th>N(0, 1)</th>
<th>N(0, 1)-GARCH(1, 1)</th>
<th>t_5-GARCH(1, 1)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>100</td>
<td>300</td>
<td>1000</td>
</tr>
<tr>
<td>GSM_\Delta_2</td>
<td>4.77</td>
<td>4.71</td>
<td>4.74</td>
</tr>
<tr>
<td>GSM_\Delta</td>
<td>9.21</td>
<td>6.32</td>
<td>5.53</td>
</tr>
<tr>
<td>GSM</td>
<td>13.32</td>
<td>8.37</td>
<td>7.12</td>
</tr>
<tr>
<td>Q_5</td>
<td>4.12</td>
<td>4.75</td>
<td>4.74</td>
</tr>
<tr>
<td>Q_10</td>
<td>4.06</td>
<td>4.53</td>
<td>4.64</td>
</tr>
<tr>
<td>Q_20</td>
<td>3.20</td>
<td>4.29</td>
<td>4.49</td>
</tr>
<tr>
<td>EL</td>
<td>7.80</td>
<td>6.70</td>
<td>5.47</td>
</tr>
</tbody>
</table>

West's estimator) while using the asymptotic covariance matrix implied by vanishing fourth order cumulants. This feasible statistic is denoted with \( GSM_{\Delta} \).

The \( GSM^s \) and \( GSM^\Delta \) 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 \( GSM^s \), \( GSM^\Delta \), and \( GSM^\Delta \), 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 \( GSM^s \), \( GSM^\Delta \), and \( GSM^\Delta \), the Q\_k test of Box and Pierce (1970), and the Escanciano–Lobato test (EL, see Escanciano and Lobato, 2009). We consider sample sizes of 100, 300, 1000, and 5000 observations and compute the empirical rejection rates from 50,000 replications of the following 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,
   \[
   y_t = \sigma_t \epsilon_t, \quad \epsilon_t \sim N(0, 1), \quad \sigma_t^2 = 0.001 + 0.05\sigma_{t-1}^2 + 0.90\sigma_{t-1}^2;
   \]
3. A GARCH(1, 1) process with i.i.d innovations following a Student’s t with 5 degrees of freedom;
   \[
   y_t = \sigma_t \epsilon_t, \quad \epsilon_t \sim N(0, 1), \quad \log \sigma_t^2 = 0.001 + 0.5|\epsilon_t| - 0.2\epsilon_t + 0.95\sigma_{t-1}^2;
   \]
4. An EGARCH(1, 1) process with i.i.d standard normal innovations
   \[
   y_t = \sigma_t \epsilon_t, \quad \epsilon_t \sim N(0, 1), \quad \epsilon_t \sim t(5), \quad \log \sigma_t^2 = 0.001 + 0.5|\epsilon_t| - 0.2\epsilon_t + 0.95\sigma_{t-1}^2;
   \]
5. An EGARCH(1, 1) process with i.i.d standard normal innovations
   \[
   y_t = \sigma_t \epsilon_t, \quad \epsilon_t \sim N(0, 1), \quad \epsilon_t \sim t(5), \quad \log \sigma_t^2 = 0.001 + 0.5|\epsilon_t| - 0.2\epsilon_t + 0.95\sigma_{t-1}^2;
   \]
6. A mixture of two normals \( N(0, 1/2) \) and \( N(0, 1) \) with mixing probability 1/2;
7. A heterogeneous normal with trending mean: \( y_t \sim N(0, 1) \).

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

For a small sample size (100 observations), the \( GSM^\Delta \) 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 \( GSM^\Delta \) is accurate across all models (in particular at the 5% level). In general, the test \( GSM^\Delta \) performs satisfactorily across all models: at the 1% level \( GSM^\Delta \) dominates \( EL \) in all cases; but against
Fig. 5. Empirical power functions of the tests $GS_1^2$, $GS_2^2$, and $GSM_2^2$ (first, second, and third row, respectively) against AR(1) and AR(2) alternatives (first and second columns, respectively). The rejection rates are based on 5000 replications with 1% nominal size for sample sizes of 100 (circle), 300 (triangle), and 1000 (square) observations.

Fig. 6. Contours of the power surface of the tests $GS_1^2$, $GS_2^2$, and $GSM_2^2$ against the Gaussian AR(2) alternative. Simulations are carried out for a grid of values of the parameters obtained varying $\alpha_1$ in the interval $(-0.50, 0.50)$ and $\alpha_2$ 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 25th, 50th, 75th and 100th quantiles, while each gray line corresponds to a 5% increment.
Table 3

Size-adjusted power against Gaussian AR(2) processes.

<table>
<thead>
<tr>
<th>$\alpha_1$</th>
<th>0.30</th>
<th>0.20</th>
<th>0.10</th>
<th>0.00</th>
<th>-0.10</th>
<th>-0.20</th>
<th>-0.30</th>
</tr>
</thead>
<tbody>
<tr>
<td>y</td>
<td>94.3</td>
<td>76.2</td>
<td>51.6</td>
<td>43.8</td>
<td>62.0</td>
<td>85.7</td>
<td>96.9</td>
</tr>
<tr>
<td>0.20</td>
<td>85.1</td>
<td>54.1</td>
<td>23.3</td>
<td>17.2</td>
<td>33.4</td>
<td>64.6</td>
<td>89.6</td>
</tr>
<tr>
<td>0.10</td>
<td>69.7</td>
<td>32.8</td>
<td>8.7</td>
<td>4.3</td>
<td>12.2</td>
<td>40.3</td>
<td>74.9</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>0.00</td>
<td>54.7</td>
<td>18.3</td>
<td>3.2</td>
<td>1.2</td>
<td>4.6</td>
<td>21.3</td>
</tr>
<tr>
<td>-0.10</td>
<td>39.7</td>
<td>11.0</td>
<td>2.7</td>
<td>2.5</td>
<td>5.2</td>
<td>17.1</td>
<td>46.5</td>
</tr>
<tr>
<td>-0.20</td>
<td>33.4</td>
<td>11.8</td>
<td>8.0</td>
<td>11.9</td>
<td>18.5</td>
<td>31.6</td>
<td>54.3</td>
</tr>
<tr>
<td>-0.30</td>
<td>40.5</td>
<td>27.2</td>
<td>29.2</td>
<td>37.3</td>
<td>48.3</td>
<td>60.8</td>
<td>76.3</td>
</tr>
</tbody>
</table>

EGARCH, while at the 5% level with $T \geq 300$ the two tests perform very similarly (although, EL maintains a significant edge against EGARCH).

Fig. 5 illustrates the empirical power functions of the tests $G_1$, $G_2$, and $G_3$ against two one-dimensional families of alternatives, an AR(1) model ($y_t = \alpha y_{t-1} + \epsilon_t$) and a restricted AR(2) model ($y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \epsilon_t$, $\epsilon_t \sim N(0, 1)$).

Simulations are carried out for a set of alternatives obtained varying $\alpha_1$ in the interval $(-0.50, 0.50)$ and $\alpha_2$ in $(-0.45, 0.45)$ in increments of 0.05.

<table>
<thead>
<tr>
<th>$\alpha_1$</th>
<th>0.30</th>
<th>0.20</th>
<th>0.10</th>
<th>0.00</th>
<th>-0.10</th>
<th>-0.20</th>
<th>-0.30</th>
</tr>
</thead>
<tbody>
<tr>
<td>y</td>
<td>84.7</td>
<td>58.8</td>
<td>31.7</td>
<td>19.9</td>
<td>25.7</td>
<td>45.9</td>
<td>77.8</td>
</tr>
<tr>
<td>0.20</td>
<td>64.4</td>
<td>33.3</td>
<td>12.3</td>
<td>6.8</td>
<td>9.6</td>
<td>26.1</td>
<td>56.2</td>
</tr>
<tr>
<td>0.10</td>
<td>39.3</td>
<td>15.3</td>
<td>5.0</td>
<td>1.7</td>
<td>3.6</td>
<td>12.1</td>
<td>32.4</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>0.00</td>
<td>21.6</td>
<td>7.3</td>
<td>2.3</td>
<td>1.1</td>
<td>4.7</td>
<td>17.9</td>
</tr>
<tr>
<td>-0.10</td>
<td>14.5</td>
<td>5.9</td>
<td>2.8</td>
<td>1.9</td>
<td>2.2</td>
<td>4.4</td>
<td>11.3</td>
</tr>
<tr>
<td>-0.20</td>
<td>16.1</td>
<td>9.9</td>
<td>7.1</td>
<td>6.7</td>
<td>7.2</td>
<td>9.2</td>
<td>14.7</td>
</tr>
<tr>
<td>-0.30</td>
<td>28.2</td>
<td>23.3</td>
<td>21.1</td>
<td>19.7</td>
<td>20.5</td>
<td>23.6</td>
<td>28.9</td>
</tr>
</tbody>
</table>

$Q_{20}$

<table>
<thead>
<tr>
<th>$\alpha_1$</th>
<th>0.30</th>
<th>0.20</th>
<th>0.10</th>
<th>0.00</th>
<th>-0.10</th>
<th>-0.20</th>
<th>-0.30</th>
</tr>
</thead>
<tbody>
<tr>
<td>y</td>
<td>82.7</td>
<td>55.5</td>
<td>29.3</td>
<td>17.7</td>
<td>23.0</td>
<td>45.5</td>
<td>75.1</td>
</tr>
<tr>
<td>0.20</td>
<td>60.9</td>
<td>30.5</td>
<td>11.2</td>
<td>6.1</td>
<td>8.6</td>
<td>24.0</td>
<td>52.8</td>
</tr>
<tr>
<td>0.10</td>
<td>35.5</td>
<td>13.7</td>
<td>4.4</td>
<td>1.6</td>
<td>3.3</td>
<td>10.9</td>
<td>30.6</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>0.00</td>
<td>19.1</td>
<td>6.5</td>
<td>2.1</td>
<td>1.0</td>
<td>5.0</td>
<td>15.9</td>
</tr>
<tr>
<td>-0.10</td>
<td>12.8</td>
<td>5.3</td>
<td>2.5</td>
<td>1.8</td>
<td>2.2</td>
<td>3.8</td>
<td>10.0</td>
</tr>
<tr>
<td>-0.20</td>
<td>14.3</td>
<td>8.9</td>
<td>6.4</td>
<td>6.3</td>
<td>6.5</td>
<td>8.3</td>
<td>12.8</td>
</tr>
<tr>
<td>-0.30</td>
<td>25.1</td>
<td>20.8</td>
<td>18.7</td>
<td>17.8</td>
<td>18.7</td>
<td>21.3</td>
<td>25.9</td>
</tr>
</tbody>
</table>

$L_B$

<table>
<thead>
<tr>
<th>$\alpha_1$</th>
<th>0.30</th>
<th>0.20</th>
<th>0.10</th>
<th>0.00</th>
<th>-0.10</th>
<th>-0.20</th>
<th>-0.30</th>
</tr>
</thead>
<tbody>
<tr>
<td>y</td>
<td>94.8</td>
<td>80.1</td>
<td>57.7</td>
<td>42.6</td>
<td>51.4</td>
<td>73.7</td>
<td>91.9</td>
</tr>
<tr>
<td>0.20</td>
<td>82.4</td>
<td>54.0</td>
<td>26.8</td>
<td>15.1</td>
<td>21.7</td>
<td>45.3</td>
<td>74.4</td>
</tr>
<tr>
<td>0.10</td>
<td>59.2</td>
<td>27.5</td>
<td>8.7</td>
<td>3.3</td>
<td>6.5</td>
<td>20.8</td>
<td>50.8</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>0.00</td>
<td>39.8</td>
<td>12.8</td>
<td>2.8</td>
<td>1.3</td>
<td>2.0</td>
<td>8.7</td>
</tr>
<tr>
<td>-0.10</td>
<td>29.2</td>
<td>9.8</td>
<td>4.6</td>
<td>1.0</td>
<td>4.1</td>
<td>9.8</td>
<td>23.3</td>
</tr>
<tr>
<td>-0.20</td>
<td>31.0</td>
<td>20.1</td>
<td>15.7</td>
<td>14.4</td>
<td>15.7</td>
<td>19.7</td>
<td>28.9</td>
</tr>
<tr>
<td>-0.30</td>
<td>55.1</td>
<td>50.3</td>
<td>45.1</td>
<td>41.9</td>
<td>43.9</td>
<td>48.7</td>
<td>55.1</td>
</tr>
</tbody>
</table>

$E_L$

<table>
<thead>
<tr>
<th>$\alpha_1$</th>
<th>0.30</th>
<th>0.20</th>
<th>0.10</th>
<th>0.00</th>
<th>-0.10</th>
<th>-0.20</th>
<th>-0.30</th>
</tr>
</thead>
<tbody>
<tr>
<td>y</td>
<td>50.0</td>
<td>34.1</td>
<td>20.6</td>
<td>12.7</td>
<td>19.4</td>
<td>34.1</td>
<td>50.0</td>
</tr>
<tr>
<td>0.20</td>
<td>40.1</td>
<td>24.3</td>
<td>14.4</td>
<td>9.0</td>
<td>16.2</td>
<td>30.2</td>
<td>45.2</td>
</tr>
<tr>
<td>0.10</td>
<td>26.3</td>
<td>15.6</td>
<td>9.1</td>
<td>5.2</td>
<td>12.7</td>
<td>22.6</td>
<td>33.5</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>0.00</td>
<td>17.8</td>
<td>11.9</td>
<td>7.3</td>
<td>4.6</td>
<td>11.9</td>
<td>23.3</td>
</tr>
<tr>
<td>-0.10</td>
<td>9.8</td>
<td>6.4</td>
<td>3.8</td>
<td>2.3</td>
<td>5.3</td>
<td>11.9</td>
<td>23.3</td>
</tr>
<tr>
<td>-0.20</td>
<td>11.9</td>
<td>8.4</td>
<td>5.8</td>
<td>4.6</td>
<td>7.8</td>
<td>15.3</td>
<td>28.9</td>
</tr>
<tr>
<td>-0.30</td>
<td>23.3</td>
<td>17.8</td>
<td>12.3</td>
<td>9.3</td>
<td>15.3</td>
<td>23.3</td>
<td>45.2</td>
</tr>
</tbody>
</table>

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

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

we plot in Fig. 6 the contours of the power surface obtained by varying $\alpha_1$ in the interval $(-0.50, 0.50)$ and $\alpha_2$ in $(-0.45, 0.45)$. Simulations are carried out for a grid of values of the parameters spaced by 0.05 and intermediate values are interpolated. The black lines correspond to 25%, 50%, 75%, and 100% percent power (starting form the center), while the gray lines correspond to 5% increments. Approximately, contour lines of the power function of the $G_S$ test (first panel) run vertically, an indication that the first scale test is not very sensitive to variations in the parameter $\alpha_2$. This picture is approximately reversed in the second panel: the contour lines for the $G_S$ test run horizontally. In the third panel we see that the contour lines of the multi-scale test $G_S$ are,
even in small samples, close to ellipses, the shape predicted by our asymptotic results.

In the remainder 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 Eq. (18). The first column contains the size adjusted power of each test for various alternatives. In the second column we report the relative power gains of the multi-scale test $GSM_g^2$ with respect to the LB tests, the BP test and the EL test. Against the great majority of the alternatives the $GSM_g^2$ test outperforms the BP and LB tests. The $GSM_g^2$ 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 $\alpha_1$ is positive, neither test has a clear edge, with variations in power against various alternatives between $+44\%$ and $-49\%$.

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 $GSM_g^2$ outperforms the BP and LB tests across a wide variety of alternatives.

---

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

13 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.

14 Despite our adjustments, sized-distortions remain because of the random nature of the Monte Carlo simulations.
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,
   \[ y_t = \sigma_t \epsilon_t, \quad \epsilon_t \sim N(0, 1), \quad \sigma_t^2 = 0.001 + 0.05\sigma_{t-1}^2 + 0.90\sigma_{t-1}^2. \]

All simulations are based on 10,000 replications.

<table>
<thead>
<tr>
<th>model</th>
<th>( N )</th>
<th>( GSM_2 )</th>
<th>( GSM_3 )</th>
<th>( GSM_4 )</th>
<th>( GSM_5 )</th>
<th>( GSM_6 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>norm</td>
<td>100</td>
<td>0.0518</td>
<td>0.0399</td>
<td>0.0428</td>
<td>0.0503</td>
<td>0.0664</td>
</tr>
<tr>
<td>t3</td>
<td>100</td>
<td>0.0385</td>
<td>0.0388</td>
<td>0.0428</td>
<td>0.0503</td>
<td>0.0664</td>
</tr>
<tr>
<td>garch</td>
<td>100</td>
<td>0.0566</td>
<td>0.0603</td>
<td>0.0667</td>
<td>0.0727</td>
<td>0.0896</td>
</tr>
<tr>
<td>norm</td>
<td>300</td>
<td>0.0468</td>
<td>0.0503</td>
<td>0.0547</td>
<td>0.0574</td>
<td>0.0650</td>
</tr>
<tr>
<td>t3</td>
<td>300</td>
<td>0.0425</td>
<td>0.0431</td>
<td>0.0433</td>
<td>0.0496</td>
<td>0.0565</td>
</tr>
<tr>
<td>garch</td>
<td>300</td>
<td>0.0444</td>
<td>0.0673</td>
<td>0.0705</td>
<td>0.0729</td>
<td>0.0831</td>
</tr>
<tr>
<td>norm</td>
<td>1000</td>
<td>0.0493</td>
<td>0.0500</td>
<td>0.0506</td>
<td>0.0518</td>
<td>0.0543</td>
</tr>
<tr>
<td>t3</td>
<td>1000</td>
<td>0.0486</td>
<td>0.0485</td>
<td>0.0493</td>
<td>0.0503</td>
<td>0.0532</td>
</tr>
<tr>
<td>garch</td>
<td>1000</td>
<td>0.0736</td>
<td>0.0755</td>
<td>0.0797</td>
<td>0.0798</td>
<td>0.0814</td>
</tr>
</tbody>
</table>

Table 6  
Power for higher order wavelet decompositions.

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

\[ y_t = 0.1y_{t-p} + \epsilon_t, \quad \text{for } p = 1, 2, 4. \]

All simulations are based on 10,000 replications.

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

(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.

In econometric practice, it is 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 the wavelet filter. We report size and power simulations comparing \( N \) up to 6. Table 5 shows that in general there is a tradeoff between the depth of the wavelet decomposition and the sample size: for small sample size, a shallower wavelet decomposition has better size properties.

To investigate the empirical 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.

Finally, wavelet filters of different lengths, namely, Haar filter (length two) to Daubechies filters D(4), D(6) and D(8) are studied in Table 7. The power of the wavelet test in small samples (100–1000 observations) does not exhibit significant differences, although there is a small deterioration in the empirical size and the power for the D(8) filter. The empirical power exhibits 15%–21% difference across restricted AR(1) and AR(2) models for a sample size of 1000 observations.

6. Applications to high frequency financial data

In this section we apply our test and the EL test to high frequency market data, specifically to returns from transactions of Apple Inc. (AAPL). We use intraday data from January 2, 2012 to December 28, 2012 and restrict our sample to the 10 min time
Table 7
Size and power for various wavelet families.

<table>
<thead>
<tr>
<th>Model</th>
<th>N</th>
<th>Haar</th>
<th>D(4)</th>
<th>D(6)</th>
<th>D(8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>garch</td>
<td>100</td>
<td>0.0477</td>
<td>0.044100</td>
<td>0.055700</td>
<td>0.074400</td>
</tr>
<tr>
<td>garch</td>
<td>300</td>
<td>0.0471</td>
<td>0.049600</td>
<td>0.053700</td>
<td>0.081000</td>
</tr>
<tr>
<td>garch</td>
<td>1000</td>
<td>0.0474</td>
<td>0.046000</td>
<td>0.075900</td>
<td>0.073700</td>
</tr>
<tr>
<td>norm</td>
<td>100</td>
<td>0.0418</td>
<td>0.040500</td>
<td>0.035200</td>
<td>0.052700</td>
</tr>
<tr>
<td>norm</td>
<td>300</td>
<td>0.0425</td>
<td>0.043500</td>
<td>0.043200</td>
<td>0.074200</td>
</tr>
<tr>
<td>norm</td>
<td>1000</td>
<td>0.0486</td>
<td>0.051700</td>
<td>0.052400</td>
<td>0.073800</td>
</tr>
<tr>
<td>t3</td>
<td>100</td>
<td>0.0563</td>
<td>0.059900</td>
<td>0.056200</td>
<td>0.087100</td>
</tr>
<tr>
<td>t3</td>
<td>300</td>
<td>0.0706</td>
<td>0.068600</td>
<td>0.085700</td>
<td>0.103000</td>
</tr>
<tr>
<td>t3</td>
<td>1000</td>
<td>0.0702</td>
<td>0.068800</td>
<td>0.083500</td>
<td>0.080700</td>
</tr>
</tbody>
</table>

Panel B: simulations under the alternative

<table>
<thead>
<tr>
<th>Test</th>
<th>N</th>
<th>Haar</th>
<th>D(4)</th>
<th>D(6)</th>
<th>D(8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>rarl(1)</td>
<td>100</td>
<td>0.1009</td>
<td>0.0976489</td>
<td>0.0929508</td>
<td>0.0693518</td>
</tr>
<tr>
<td>rarl(1)</td>
<td>300</td>
<td>0.2966</td>
<td>0.2987647</td>
<td>0.2975818</td>
<td>0.2803464</td>
</tr>
<tr>
<td>rarl(1)</td>
<td>1000</td>
<td>0.8152</td>
<td>0.8232647</td>
<td>0.8194157</td>
<td>0.8068786</td>
</tr>
<tr>
<td>rarl(2)</td>
<td>100</td>
<td>0.0912</td>
<td>0.0838653</td>
<td>0.0799517</td>
<td>0.0849552</td>
</tr>
<tr>
<td>rarl(2)</td>
<td>300</td>
<td>0.2376</td>
<td>0.2329933</td>
<td>0.2180051</td>
<td>0.2295319</td>
</tr>
<tr>
<td>rarl(2)</td>
<td>1000</td>
<td>0.7069</td>
<td>0.7109645</td>
<td>0.7119742</td>
<td>0.6670314</td>
</tr>
</tbody>
</table>

interval from 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 the one hand an indication of liquidity (as an indirect estimate of the bid–ask spread, see Roll, 1984) and on the other hand is a measure of market efficiency (see, for example, Jegadeesh and Titman, 2001).

The average p-values over the 251 testing days for the tests GM₃, GSM₄, and EL 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 EL test. This example shows that our test can be useful in econometric practice.

7. Conclusions

We use the wavelet coefficients of the observed time series to construct test statistics in the spirit of the 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

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

where the lag order is allowed to grow with T. On the 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 (1988), 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 to optimally choose 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.
Appendix A. Proofs

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

\[ z_{m,t} := \sum_{j=-\infty}^{\infty} h_{m,i} h_{m,j} y_{t+j} y_{t-j} \]

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

**Proof of Proposition 2.** Recall that on a measure space \((\mathcal{X}, \mu, \gamma)\), for any \( f \in L^p(\mathcal{X}) \) and \( g \in L^q(\mathcal{X}) \), the generalized Hölder inequality holds (see, for example, Reed and Simon, 1972, page 82):

\[ \|fg\|_r \leq \|f\|_p \|g\|_q \alpha \quad \text{with} \quad p^{-1} + q^{-1} = r^{-1}, \]

(19)

in particular, if \( p = q \), \( \|fg\|_{L^p} \leq \|f\|_p \|g\|_p \). For the remainder of this proof let \( \mathbb{E}[-] = \mathbb{E}[\cdot; F_{t-m}(\epsilon)] \). 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):

\[ \|x_i y_i - E x_i y_i\|_{L^p} \]

\[ = \|x_i y_i - x_i E y_i + (x_i - E x_i) y_i - E x_i y_i\|_{L^p} \]

\[ \leq \|x_i (y_i - E y_i)\|_{L^p} + \|(x_i - E x_i) E y_i\|_{L^p} \]

\[ \leq \|x_i\|_p \|y_i - E y_i\|_{L^p} \]

\[ + \|x_i - E x_i\|_p \|E y_i\|_{L^p} \]

\[ \leq \|x_i\|_p \|d_i\|_{L^m} \]

\[ + \|y_i\|_p \|d_i\|_{L^m} \]

\[ + \|d_i\|_{L^m} \|d_i\|_{L^m} \] \( \leq d_i v_m \),

where \( d_i = \max(\|x_i\|_{L^p}, \|y_i\|_{L^p}, \|d_i\|_{L^m}) \) and \( v_m = O(m^{-\min(\phi, \phi)}) \). \( \Box \)

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

\[ z_{m,t} = w_{m,t}^2 - \sum_{i=0}^{T} h_{m,i} y_{t-i}^2 \]

Again, since the linear combination of NED processes is an NED process, \( \{z_{m,t}\} \) is \( L_1 \)-NED. Notice that \( \tilde{\varepsilon}_{m,T} - 2^{-m} \) can be written in terms of \( z_i \) and \( y_i \):

\[ \tilde{\varepsilon}_{m,T} = \frac{1}{2^m} \frac{1}{\sum_{t=1}^{T} y_i^2} \]

and

\[ \tilde{\varepsilon}_{m,T} = \frac{1}{2^m} \frac{1}{\sum_{t=1}^{T} y_i^2} \]

(20)

\[ \sum_{t=1}^{T} \left( \sum_{i=0}^{T} h_{m,i}^2 y_{t-i}^2 + 2 \sum_{j=1}^{T} \sum_{i=0}^{T} h_{m,i} h_{m,j} y_{t-j} y_{t-i} \right) \]

\[ = \sum_{t=1}^{T} \left( \sum_{i=0}^{T} h_{m,i}^2 y_{t-i}^2 + 2 \sum_{j=1}^{T} \sum_{i=0}^{T} h_{m,i} h_{m,j} y_{t-j} y_{t-i} \right) \]

\[ = \sum_{t=1}^{T} \frac{1}{2^m} \frac{1}{\sum_{t=1}^{T} y_i^2} \]

(21)

Step (22) uses the fact that filtering is cyclic, therefore the sum \( \sum_{i=1}^{T} y_{t-i} \) does not depend on \( i \) and is the same as the denominator \( \sum_{i=1}^{T} y_{t-i} \). The last equality holds because the norm of a convolution is the product of the norms. Since \( h_{m,i} \) 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

\[ \frac{2 \sum_{t=1}^{T} z_{m,t}}{\sum_{t=1}^{T} y_i^2} \]

\[ \rightarrow 0 \quad \text{as} \quad p \rightarrow \infty \]

and the theorem is proven. \( \Box \)

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,

\[ \sum_{t=1}^{n} n \]

\[ \sum_{t=1}^{n} y_i^2 \]

\[ \rightarrow 0 \]

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

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

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

for all \( i, j, k, \) and \( l \) such that \( 0 \leq i < l \leq L \) and \( 0 \leq k < l \leq L \). Then,

\[ \text{var}(z_i) = \sum_{l=0}^{L} \sum_{j=0}^{L} \sum_{k=0}^{L} h_{h} h_{h} h_{h} h_{h} \mathbb{E}[y_{t-j} y_{t-k} y_{t-l} y_{t-i}] \]

and

\[ \text{cov}(z_i, z_{i-k}) = \sum_{l=0}^{L} \sum_{j=0}^{L} \sum_{k=0}^{L} h_{h} h_{h} h_{h} h_{h} \mathbb{E}[y_{t-j} y_{t-k} y_{t-l} y_{t-i}] \]

(22)
**Proof.** The proof relies on a direct computation. First, we compute the variance:

\[
\begin{align*}
\text{var}(z_t) &= \text{var} \left( \sum_{t=0}^{l-1} \sum_{j=1}^{l} h_j h_{t-j} y_{t-j} y_{t-j} \right) \\
&= \text{cov} \left( \sum_{t=0}^{l-1} \sum_{j=1}^{l} h_j h_{t-j} y_{t-j} y_{t-j} \right) \\
&= \sum_{t=0}^{l-1} \sum_{j=1}^{l} \sum_{k=0}^{l} h_j h_k h_{t-k} h_{t-j-k} \text{Cov}(y_{t-j} y_{t-j-k} y_{t-j-k} y_{t-j-k}) \\
&= \sum_{t=0}^{l-1} \sum_{j=1}^{l} \sum_{k=0}^{l} \sum_{h=0}^{l} h_j h_k h_{t-k} h_{t-j-k} (y_{t-j} y_{t-j-k} y_{t-j-k} y_{t-j-k}).
\end{align*}
\]

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

\[
\begin{align*}
\text{cov}(z_t, z_{t-s}) &= \text{cov} \left( \sum_{t=0}^{l-1} \sum_{j=1}^{l} h_j h_{t-j} y_{t-j} y_{t-j} \right) \\
&= \sum_{t=0}^{l-1} \sum_{j=1}^{l} \sum_{k=0}^{l} h_j h_k h_{t-k} h_{t-j-k} \text{Cov}(y_{t-j} y_{t-j-k} y_{t-j-k} y_{t-j-k}) \\
&= \sum_{t=0}^{l-1} \sum_{j=1}^{l} \sum_{k=0}^{l} \sum_{h=0}^{l} h_j h_k h_{t-k} h_{t-j-k} (y_{t-j} y_{t-j-k} y_{t-j-k} y_{t-j-k}).
\end{align*}
\]

Proof of Proposition 8. Since \( \{z_{m,t}\} \) is a linear combination of processes of the form \( \{y_l y_{l-1}\} \) and since the NED property is preserved under linear combinations, it follows that under Assumption B2, \( \{z_{m,t}\} \) is L2-NED of size \(-1/2\) on \( \varepsilon_t \).

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

\[
\text{var}(z_{m,t}) = \sum_{t=0}^{l-1} \sum_{j=1}^{l} \sum_{k=0}^{l} \sum_{h=0}^{l} h_j h_k h_{t-k} h_{t-j-k} (y_{t-j} y_{t-j-k} y_{t-j-k} y_{t-j-k}).
\]

Then,

\[
\begin{align*}
\begin{vmatrix}
\sum_{t=0}^{l-1} \sum_{j=1}^{l} \sum_{k=0}^{l} \sum_{h=0}^{l} h_j h_k h_{t-k} h_{t-j-k} (y_{t-j} y_{t-j-k} y_{t-j-k} y_{t-j-k}) \\
\end{vmatrix}_p = \text{var}(z_{m,t}) = \frac{\sigma^2}{\sigma_{m,t}}.
\end{align*}
\]

which implies that \( z_{m,t}/\sigma_{m,t} \) is \( L_\alpha \)-bounded for \( \alpha = 2p \).

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

\[
\sum_{t=1}^{T} \frac{z_{m,t}}{\sigma(z)} \xrightarrow{d} N(0, 1).
\]

Therefore,

\[
\begin{align*}
\sum_{t=1}^{T} \frac{y_t^2}{525(z)} \left( \hat{\theta}_{m,T} - \frac{1}{2m} \right) &\xrightarrow{d} N(0, 1) \\
\sqrt{\frac{T \sigma^2}{452(z)}} \left( \hat{\theta}_{m,T} - \frac{1}{2m} \right) &\xrightarrow{d} N(0, 1),
\end{align*}
\]

where \( \sigma^2 = T^{-1} \sum_{t=1}^{T} \text{EY}_t^2 \).

**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 \( \sigma^2 \), and vanishing fourth joint cumulants. Let \( \{h_l\}_{l=1}^{L} \) be an \( L \)-dimensional vector. Then the stochastic sequence \( \{z_t\} \) has variance

\[
\text{var}(z_t) = \sigma^4 \sum_{l=0}^{L-1} \sum_{j=1}^{L} (h_l h_j)^2,
\]

and autocovariances

\[
\text{cov}(z_t, z_{t-s}) = \begin{cases} 
\sigma^4 \sum_{l=0}^{L-1} \sum_{j=1}^{L} h_l h_{t-j} h_{t-j-s} h_{t-j-s} & \text{if } s \leq L - 1 \\
0 & \text{otherwise}
\end{cases}
\]

where

\[
\begin{align*}
i_{\text{min}} &= \max(0, s), \quad i_{\text{max}} = L + \min(0, s), \\
L_{\text{max}} &= L + \min(0, s).
\end{align*}
\]

**Proof.** When fourth cumulants are zero, the fourth moment \( \kappa^{r,s,t,u}_{m} \) of \( y_t \) can be expressed in terms of the second moments \( \kappa^{r,s}_m \) and \( \kappa^{r,s,t,u}_{m} \). Such decomposition is valid whenever the fourth cumulant \( \kappa^{r,s,t,u}_{m} \) is zero. Indeed (see for example McCullagh (1987))

\[
\kappa^{r,s,t,u}_{m} = \kappa^{r,s,t,u}_{m} + \kappa^{r,s,t,u}_{m} + \kappa^{r,s,t,u}_{m} + \kappa^{r,s,t,u}_{m},
\]

where 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 \( \kappa^{r,s}_m = 0 \) as \( y_t \) is a zero mean sequence. Continuing from Eq. (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 Eq. (23) correspond to the two possibilities \( (i, j, k, l) \) and \( (i, k, j, l) \) (from the second summation). Therefore, when \( i = j \) and \( j = k \),

\[
i \neq j \Rightarrow i \neq j,
\]

which contradicts the condition \( j > i \) (from the second summation). Thus, \( \delta_j = 0 \) whenever \( i = j \) and \( 0 \) otherwise. Thus,

\[
\begin{align*}
\sum_{i=0}^{L-1} \sum_{j=1}^{L} \sum_{k=0}^{L} \sum_{h=0}^{L} h_i h_j h_k h_{t-j-k} h_{t-j-k-l} (\delta_j \delta_j + \delta_j \delta_k) \\
&= \sum_{i=0}^{L-1} \sum_{j=1}^{L} \sum_{h=0}^{L} h_i^2 h_j^2 (\kappa^{r,s}_m),
\end{align*}
\]

\[
\sum_{i=0}^{L-1} \sum_{j=1}^{L} \sum_{h=0}^{L} h_i^2 h_j^2 (\kappa^{r,s}_m) = \sigma^4 \sum_{l=0}^{L-1} \sum_{j=1}^{L} (h_l h_j)^2.
\]
A very similar computation yields the autocorrelation function $\gamma_{m(s)}$:

$$\gamma_{m(s)} = \text{Cov} \left( \sum_{i=0}^{l-1} \sum_{j>l} h_{i,j} y_{l-j}, \sum_{i=0}^{l-1} \sum_{j>l} h_{i,j} y_{l-j} \right)$$

$$= \sum_{i=0}^{l-1} \sum_{j>l} \sum_{k>l} h_{i,j} h_{i,k} \text{Cov}(y_{l-j}, y_{l-k})$$

$$= \sum_{i=0}^{l-1} \sum_{j>l} \sum_{k>l} h_{i,j} h_{i,k} E(y_{l-j} y_{l-k})$$

$$= \sum_{i=0}^{l-1} \sum_{j>l} \sum_{k>l} h_{i,j} h_{i,k} E(y_{l-j} y_{l-k}) \delta_{i,k}$$

$$= \sigma^4 \max_{i \leq \min} \max_{j > l} h_{i,j} h_{i,k}$$

where

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

$$j_{\text{max}} = L + \min(0, s).$$

At equality (28) we used the fact that the contribution of $\delta_{i,k} h_{i,k}$ is zero. The argument is the same as for the analogous contribution to $\gamma_{m(s)}$.

Notice that the autocovariance $\gamma(s)$ is zero when $i_{\text{min}} > i_{\text{max}}$. 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 an $(L - 1)$-dependent sequence (i.e., $z_t$ is independent of $z_{t-l}$ for $l > L - 1$). \qed

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

$$\sqrt{T} \left( \frac{\hat{\gamma}_{m,T} - \frac{1}{2^m}}{\sigma^2} \right) = \frac{2 \sum_{i=0}^{l-1} \sum_{j>l} h_{i,j} y_{l-j}}{\sqrt{T} \sum_{i=0}^{l-1} y_i^2}$$

$$= \frac{\sqrt{T} \sum_{i=0}^{l-1} 2 z_i}{\sum_{i=0}^{l-1} y_i^2} = \frac{\sqrt{T} \sum_{i=0}^{l-1} y_i^2}{\sum_{i=0}^{l-1} y_i^2} \sim \mathcal{N} \left( 0, \frac{\sigma^4 a_{0m}}{\sigma^2} \right) \sim \mathcal{N} \left( 0, 1 \right).$$

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 $a_{0m}$ from $\sigma$ follows directly from Eqs. (25) and (26). \qed

**Proof of Corollary 13.** Consider the vector $(G_{1,T}, \ldots, G_{N,T})$.

$$\left( \begin{array}{c} G_{1,T} \\ \vdots \\ G_{N,T} \end{array} \right) = \left( \begin{array}{c} \frac{1}{\sqrt{a_1}} \sqrt{T} \left( \frac{\hat{\gamma}_{1,T} - \frac{1}{2^m}}{\sigma^2} \right) \\ \vdots \\ \frac{1}{\sqrt{a_N}} \sqrt{T} \left( \frac{\hat{\gamma}_{N,T} - \frac{1}{2^m}}{\sigma^2} \right) \end{array} \right)$$

Let $q$ be the column $N$-vector with coordinates $\frac{1}{\sqrt{a_i}}$. Let $\text{diag}(v)$ be the square matrix with $v$ on the main diagonal and zero everywhere else. By definition $\text{diag}(q) \left( \sum_{i=0}^{L-1} \Gamma(s) \right) \text{diag}(q) = \sigma^2 A$. Indeed,

$$\left( \begin{array}{cccc} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{array} \right) \times \left( \begin{array}{cccc} \sigma^4 a_1 & \sigma^4 a_{12} & \cdots & \sigma^4 a_{1N} \\ \sigma^4 a_{21} & \sigma^4 a_{22} & \cdots & \sigma^4 a_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma^4 a_{N1} & \sigma^4 a_{N2} & \cdots & \sigma^4 a_{NN} \end{array} \right)$$

$$= \left( \begin{array}{cccc} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{array} \right) \times \left( \begin{array}{cccc} \sigma^4 a_1 & \sigma^4 a_{12} & \cdots & \sigma^4 a_{1N} \\ \sigma^4 a_{21} & \sigma^4 a_{22} & \cdots & \sigma^4 a_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma^4 a_{N1} & \sigma^4 a_{N2} & \cdots & \sigma^4 a_{NN} \end{array} \right)$$

$$= \sigma^4 A.$$

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

$$\left( \begin{array}{c} G_{1,T} \\ \vdots \\ G_{N,T} \end{array} \right) \xrightarrow{d} \frac{1}{\sigma^2} \mathcal{N} \left( 0, \text{diag}(q) \left( \sum_{j=0}^{L-1} \Gamma(j) \right) \text{diag}(q) \right)$$

$$\sim \mathcal{N} \left( 0, A \right).$$

\qed

**References**


