White noise testing using wavelets

Guy P. Nason and Delyan Savchev*

Received 29 October 2014; Accepted 17 November 2014

Testing whether a time series is consistent with white noise is an important task within time series analysis and for model fitting and criticism via residual diagnostics. We introduce three fast and efficient white noise tests that assess spectral constancy via the wavelet coefficients of a periodogram. The Haar wavelet white noise test derives the exact distribution of the Haar wavelet coefficients of the asymptotic periodogram under mild conditions. The single-coefficient white noise test uses a single Haar wavelet coefficient obtaining a test statistic as a linear combination of odd-indexed autocorrelations. The general wavelet white noise test uses compactly supported Daubechies' wavelets, shows that its coefficients are asymptotically normal and derives its theoretical power for an arbitrary spectrum. All our tests are available in the freely available hwwntest package for the R system. We present a comprehensive simulation study that shows the good performance of our new tests against alternatives commonly found in available software and show an example applied to a wind power time series. © 2014 The Authors. Stat published by John Wiley & Sons Ltd.

Keywords: white noise test; wavelet; periodogram

1 Introduction

Time series white noise tests are useful in their own right or as part of a procedure that assesses model fit from residuals. A stochastic process is white noise if and only if its spectrum is flat and if and only if the wavelet coefficients of the spectrum are all zero. Our tests work by assessing whether the well-known estimator of the spectrum, the periodogram, has wavelet coefficients that are significantly different from zero.

We develop a highly accurate approximation to the Haar wavelet coefficient distribution of the normalized periodogram. Using this approximation, and contemporary multiple hypothesis testing methods, Section 2 introduces a test for white noise based on all the Haar wavelet coefficients and also a single Haar wavelet coefficient test. A further general wavelet test, based on assessing wavelet coefficients produced by compactly supported Daubechies' wavelets, appears in Section 3. Here, an exact distribution is out of reach, so we use the asymptotic normality result due to Neumann (1996) to assess coefficient significance. Section 3 additionally develops a theoretical formula for the general wavelet test's power for any spectrum for a fixed number of observations T. The power formula is implemented in our software package and provides useful guidance on questions such as ‘how large does my sample have to be to detect departures from white noise against some specific alternative’?

Many white noise tests already exist, so why do we need another? The answer is that all tests are ‘directional’ in that they have excellent power for some alternatives, reasonable power for many alternatives and poor power for others. We...
show that our test achieves high power when others fail (and, naturally, vice versa) and hence is a useful complement to existing methods. Apart from specifying statistical size, our test requires no tuning parameters. Section 4 describes an extensive simulation study comparing our new tests with some established ones and a wind speed data example.

Next, we briefly summarize a few popular white noise tests and cite more extensive review articles. Sections 1.2 and 1.3 set up the basic components of our test and describe how they are combined to assess spectral constancy.

1.1. Brief overview of white noise tests

Box & Pierce (1970) introduced a popular white noise test that was designed to work with the residuals from an autoregressive integrated moving average model fit. They developed a mathematical approximation to the residual distribution and showed it to be chi-square-distributed with degrees of freedom equal to the lag up to which the autocorrelations are tested (minus the number of fitted autoregressive moving average (ARMA) parameters). This work formed the basis for two tests—Box–Pierce and Ljung–Box from Ljung & Box (1978), the latter a modification of the earlier test. The test statistics used were proportional to \( T \sum_{i=1}^{p} \hat{r}_i^2 \) where \( T \) is the number of observations, \( \hat{r}_i \) is the usual estimate of autocorrelation and \( \rho \) denotes an upper bound on the lags being tested. Both versions are portmanteau tests, and the null hypothesis is that \( \rho \) is zero for \( i = 1, \ldots, p \). The operation of these tests requires the user to prespecify a likely value of \( \rho \). Selecting a too small \( \rho \) causes the test to miss any significant autocorrelations at higher lags. Selecting a too large \( \rho \) can detect more significant autocorrelations, but often at reduced power. Many subsequent papers have attempted to address the problem of selecting the ‘best’ lag for a given problem or averaging results over several lags (e.g. Guay et al., 2013).

Frequency domain white noise tests assess the ‘flatness’ of a spectrum estimate. Early tests of this type are that of Bartlett (1954, 1955), which apply a Kolmogorov–Smirnov test to the cumulative periodogram ordinates. See also Durbin (1969, 1970). Some more recent tests and comprehensive reviews can be found in the works of Lobato (2001) and Guay et al. (2013).

1.2. Basic components of our tests

Suppose that \( \{X_t\}_{t \in \mathbb{Z}} \) is a real-valued second-order stationary stochastic process with mean \( \mu = 0 \), variance \( \sigma^2 < \infty \), autocovariance \( \gamma_X(k) \) for integers \( k \) and spectrum \( f(\omega) \) for \( \omega \in [-\pi, \pi) = \Pi \). In practice, we remove the mean for any time series for which \( \mu \neq 0 \). We address the problem of testing the hypothesis \( H_0 \) that \( \{X_t\} \) is a white noise versus the alternative \( H_A \) that \( \{X_t\} \) is not white noise given a realization \( X_t, t = 1, \ldots, T \), for some positive integer \( T \).

The process \( \{X_t\} \) is white noise if and only if its spectrum \( f(\omega) \) is a constant function on \( \Pi \). The spectrum, \( f \), can be estimated from a realization \( \{X_t\}_{t=1}^T \) via the periodogram

\[
I_T(\omega) = (2\pi T)^{-1} \left| \sum_{t=1}^{T} X_t e^{-i\omega t} \right|^2,
\]

which can be computed at the Fourier frequencies \( I_p = I_T(\omega_p) \), where \( \omega_p = 2\pi p T^{-1} \) for \( p = 1, \ldots, T/2 \).

Our tests will be based on wavelet decompositions of a spectrum defined as follows.

**Definition 1** (Wavelets)

Define \( \langle \cdot, \cdot \rangle \) to be the usual inner product on \( L^2(\Pi) \) given by \( \langle f, g \rangle = \int_{-\pi}^{\pi} f(\omega)g(\omega) d\omega \). Let \( \mathbb{N}_0 = \mathbb{N} \cup 0 \) and \( \mathbb{T}(j) = \{0, \ldots, 2^j - 1\} \) for \( j \in \mathbb{N}_0 \). Let \( \{\psi_{j,k}\}_{j \in \mathbb{N}_0, k \in \mathbb{T}(j)} \) be an orthonormal periodic wavelet basis for functions \( f \in L^2(\Pi) \), where \( \psi_{j,k}(x) = 2^j/2^{j/2} \psi(2^{j/2}x - k) \), where \( \psi \in L^2(\Pi) \) is a (suitably rescaled) Daubechies’
compactely supported mother wavelet. A spectrum’s wavelet expansion can be written as $f(\omega) = \langle f, \phi_{0,0} \rangle + \sum_{j \in \mathbb{N}_0} \sum_{k \in \mathbb{Z}(j)} v_{j,k} \psi_{j,k}(\omega)$, for $\omega \in \Pi$, where $\phi(\omega)$ is the scaling function associated with the wavelet basis and $\psi_{j,k}$ is defined similarly to $\psi_{j,k}$. The coefficients of the expansion are given by $v_{j,k} = \langle f, \psi_{j,k} \rangle$ for $j \in \mathbb{N}_0$ and $k \in \mathbb{Z}(j)$.

Here, wavelets are used provide a decomposition of a spectrum, $f(\omega)$, that is localized in frequency and ‘scale of frequency’ (in contrast to the usual localization in time and frequency for a function defined over time). The wavelet coefficient sets for a wide variety of functions (technically in some Besov class) are sparse. This means that the essential information present in a spectrum can potentially be represented using few wavelet coefficients, generally fewer than in other representations such as Fourier or orthogonal polynomials. Our procedures exploit this sparsity, particularly for the general wavelet test presented in Section 3. Wavelet transforms have further advantages in that their implementations are extremely fast and efficient and the coefficients provide useful information about location and scale of non-constancies. For further information on wavelets, see Daubechies (1992), Mallat (1998), Vidakovic (1999) or Nason (2008) in the statistical context.

### 1.3. Assessing spectral constancy

We investigate the constancy of a normalized spectrum by examining its wavelet coefficients given by

$$v_{j,k} = \sigma_X^{-2} \int_{-\pi}^{\pi} f(\omega) \psi_{j,k}(\omega) \, d\omega = \sigma_X^{-2} \langle f, \psi_{j,k} \rangle \quad \text{for } j \in \mathbb{N}_0, k \in \mathbb{Z}(j).$$

(2)

For example, $\psi_{j,k}(x)$ might be the usual Haar wavelet system defined by $\psi_{j,k}(x) = 2^{-j/2} \psi(2^j x - k)$ where $\psi(x) = 1$ for $x \in [0, 1/2)$, $\psi(x) = -1$ for $x \in [1/2, 1)$ and zero otherwise.

The white noise null hypothesis can be reformulated in terms of the wavelet coefficients: if $f(\omega)$ is constant, then every $v_{j,k} = 0$, and the null hypothesis is equivalent to $H_0 : v_{j,k} = 0$ for all $j \in \mathbb{N}_0, k \in \mathbb{Z}(j)$. This is because a defining property of wavelets is that the integral of every wavelet is zero.

In practice, we do not know the spectrum and hence estimate it using the periodogram. For the Haar white noise test, we apply the discrete Haar wavelet transform to the normalized periodogram to form the following estimates of $v_{j,k}$:

$$\hat{v}_{j,k} = 2^{-j/2} \sigma_X^{-2} \left( \sum_{r=0}^{2^{j-1}-1} l_{2^{j-r}} - \sum_{q=2^{j-1}}^{2^{j-1}} l_{2^{j-1}+q} \right),$$

(3)

for $(j,k) \in \mathbb{I}_T = \{(j,k) : j = 0, \ldots, \log_2(T), k \in \mathbb{Z}(j)\}$. Here, $T = 2^J$ for some $J \in \mathbb{N}$, but the test can be extended to arbitrary-length data sets by using the Haar wavelet transform for arbitrary $T$. Alternatively, with truncation or minor ‘zero-padding’, a series can be fed to the current software without a major effect. A version of (3) exists for more general Daubechies’ wavelets, and a fast transform, called the discrete wavelet transform, exists due to Mallat (1989).

To test the constancy of $f$, we perform multiple hypothesis tests for $H_0 : v_{j,k} = 0$ for all $(j,k) \in \mathbb{I}_T$ against $H_A : \exists (j,k)$ such that $v_{j,k} \neq 0$ using $\hat{v}_{j,k}$ as test statistic. Sections 2 and 3 explain how to compute the distribution of $\hat{v}_{j,k}$, to a high degree of accuracy, for both the Haar and general wavelet situations. From this, we can obtain a ‘per-test’ $p$-value with respect to each $j,k$ pair. We use multiple hypothesis test size adjustment techniques such as Bonferroni correction for a test of size $\alpha$ or the less conservative false discovery rate method of Benjamini & Hochberg (1995).
The idea of using Haar wavelet coefficients for testing constancy appeared in the work of von Sachs & Neumann (2000) in the context of stationarity testing. Here, we test for constancy over frequency, rather than over time, and, additionally, use more general wavelets to improve the statistical power of our tests.

2 A Haar wavelet test

Under the null hypothesis of \( \{X_t\} \) being independent and identically distributed with mean zero and variance \( \sigma_X^2 < \infty \), it is known that a fixed number of periodogram ordinates are distributed asymptotically as independent exponential random variables (Brockwell & Davis, 1991, p. 344; and, apart from \( p = T/2 \), identically). Further asymptotic results that also show the exponential limit for all periodogram ordinates under certain conditions can be found in the works of Freedman & Lane (1980), Kokoszka & Mikosch (2000), Fay & Soulier (2001) and Shao & Wu (2007). If the \( X_t \) are Gaussian, then the periodogram ordinates are independent.

We will model the periodogram ordinates as independent exponential random variables with appropriate means.

For now, suppose that \( \sigma_X^2 = 1 \). Then, under \( H_0 \), the distribution of \( \hat{v}_{j,k} \) from (3) is given by the following result.

**Proposition 1**

The distribution of \( \hat{v}_{j,k} \) for \( (j,k) \in I_T \) is

\[
g_m(x) = \frac{\sqrt{2m} \exp(-\sqrt{2m}|x|)}{2^{2m-1}(m-1)!} \sum_{j=1}^{m} \frac{(m+j-2)!}{(m-j)! (j-1)!} \left( 2 \sqrt{2m} \right)^{m-j},
\]

where \( m = 2^{j-i-1} \) for \( j = 0, \ldots, J-1 \) and \( \sigma^2 = 1 \).

**Proof:** See Simon (2002, p. 26) or Kotz et al. (2001) for the essential result. A tailored version is provided in the supplementary material, which also provides information on the asymptotic behaviour of \( g_m(x) \), which is useful later.

In practice, \( \sigma_X^2 \) is unknown, and so we estimate it using the standard sum-of-squares sample variance formula. Then we operate our test on the normalized periodogram \( \hat{\sigma}_X^{-2}I_p \), which is distributed, asymptotically, as an exponentially distributed random variable with a rate of 1. The test requires no tuning parameters.

**Haar wavelet white noise (HWWN) test procedure:** Compute the normalized periodogram, and then compute its Haar wavelet transform as specified by (3). The Haar coefficients have distribution specified, to a high degree of accuracy, by (4), and the \( p \)-value for testing \( H_0 : v_{j,k} = 0 \) versus \( H_A : v_{j,k} \neq 0 \) can be obtained by using the inverse cumulative distribution of \( g_m(x) \) from (4).

To demonstrate the accuracy of our approximations, Figure 1 shows probability and cumulative density estimates of the wavelet coefficients with their theoretical \( g_{2^j}(x) \) versions superimposed and associated empirical \( p \)-values obtained by the test procedure operating on \( 2^{14} \) independent Gaussian random variables. The empirical \( p \)-values appear to be close to uniform random variables, which is what one would expect for the test statistic under the null hypothesis.

In general, Proposition 1 is only approximate for the distribution of \( \hat{v}_{j,k} \) even for Gaussian \( X_t \). For finite \( T \), the normalization by \( \hat{\sigma}_X^{-2} \) causes the periodogram ordinates to obey an \( F \)-distribution (Koen, 1990), although, in practice, the difference between this and the exponential is small, even for moderate sample sizes. For non-Gaussian random variables, the ordinates will not be precisely exponential and will suffer small correlations, although these tend to be mitigated by the ‘decorrelation’ action of the wavelet transform (Craigmile & Percival, 2005).

Any single non-zero Haar coefficient in the test described earlier indicates departures from white noise. We developed a further test using the single coefficient \( v_{0,p} \), the coarsest scale wavelet coefficient that can be written in the frequency
domain as

\[ v_{0,0} = 2^{-1/2} \sigma_n^{-2} \left( \int_0^{\pi/2} f(\omega) \mathrm{d}\omega - \int_{\pi/2}^\pi f(\omega) \mathrm{d}\omega \right). \]  

(5)

Because of the Fourier relationship between autocovariance and spectrum, the coefficient \( v_{0,0} \) can be rewritten as

\[ v_{0,0} = (4/\sqrt{2\pi}) \sigma_n^{-2} \sum_{m=0}^{\infty} (-1)^m \gamma(2m+1)/(2m+1) \approx 0.90\rho(1) - 0.30\rho(3) + 0.18\rho(5) - 0.13\rho(7) + \cdots, \]  

(6)

where \( \rho(\tau) \) is the usual autocorrelation. A test statistic can obtained from \( v_{0,0} \) by substituting the sample autocorrelation \( r_\tau \) for \( \rho(\tau) \) and comparing \( |v_{0,0}| \) with a critical value obtained either from the \( g_{2n} \) distribution (4) from Proposition 1 or using the Gaussian approximation derived in the next section. Simulations (in the following sections) show that this single coefficient, SCWN, test is surprisingly effective in practice. The SCWN test is also particularly easy to compute from the sample autocorrelations, which, in many cases, would have already been computed and plotted.
A general wavelet test

Wavelet expansions for a wide variety of functions are sparse (e.g. Wasserman, 2005, p. 208; Nason, 2008). The bottom-right plot in Figure 1 shows the Haar wavelet coefficients of the AR(1) spectrum with $\alpha_1 = 0.9$: most coefficients are small or zero, and only four or five large ones are required to represent the spectrum accurately. One reason for the sparsity of representation can be explained via wavelets’ vanishing moments property. A wavelet $\psi(x)$ has $m$ vanishing moments if $\int x^\ell \psi(x) \, dx = 0$ for $\ell = 0, \ldots, m - 1$. The property means that coefficients of wavelets that overlap the spectrum, where it is locally like a polynomial of degree less than $m$, will be small and in many cases zero. Consequently, smooth spectral densities, such as those belonging to ARMA($p, q$) processes, have sparse wavelet representations with increasing sparsity for higher numbers of vanishing moments.

Hence, this section uses Daubechies’ compactly supported wavelets with ten vanishing moments in (2) instead of Haar wavelets. The wavelet coefficients, associated with these smoother wavelets, can be written as $\hat{v}_{j,k} = \hat{\delta}_X^{-2} \sum_l g_l^{(j,k)} l_i$ where $l_i$ are the usual periodogram ordinates and the $g_l$ are the weights of the discrete wavelet transform. Formula (3) can be put into this form with $g_l = \pm 2^{l/2}$ at scale $j$, and then this window of coefficients is moved across the $l_i$ sequence to obtain the $\hat{v}_{j,k}$ for $k$ across this scale. The fast wavelet transform due to Mallat (1989) computes $\hat{v}_{j,k}$ for all the $T/2$ periodogram ordinates using only $O(T)$ operations for all Daubechies’ wavelets.

For general wavelets, no simple closed form for the null distribution of $\hat{v}_{j,k}$, such as in Proposition 1 for Haar wavelets, seems to exist because the wavelet weights, $g_l$, are not constant and change from scale to scale. Alternatively, as long as we stay away from the finer scales, it turns out that the wavelet coefficients are asymptotically normal.

For Haar wavelets, the asymptotic normality can be seen immediately by noticing that the characteristic function of the $g_m$ distribution in (4) is essentially Student’s $t$-distribution, which tends asymptotically to the Gaussian as the number of degrees of freedom tends to infinity. Hence, asymptotically, the $g_m$ distribution tends to a Gaussian as that density is a fixed point of the Fourier transform. Asymptotic normality of coefficients arising from general Daubechies’ compactly supported wavelets can be established using the following result from Neumann (1996).

**Corollary 1** (Coefficient asymptotic normality)

Let $\hat{v}_{j,k}$ be the empirical wavelet coefficients of the normalized periodogram computed by (3) for Haar wavelets or $\sum l_i g_l$ for other Daubechies’ compactly supported wavelets. Then asymptotically, as $T \to \infty$, for $2^j = O(T^{-1/2})$, we have $\hat{v}_{j,k} \sim N(0,1)$ under $H_0$ and $\hat{v}_{j,k} \sim N(v_{j,k}, \eta_{j,k}^2)$ under $H_k$, where $\eta_{j,k}^2 = <\psi_{j,k}^2, (\pi \sigma_X^{-2} f)^2>$, the coefficients of the squared wavelet transform of the square of the normalized spectrum.

More details can be found in the work of Nason & Savchev (2014). The $2^j = O(T^{-1/2})$ ensures that we only consider coefficients away from the fine scales and is assumption (A1) from von Sachs & Neumann (2000). Our general wavelet test is called $\text{GENNWN}$, and its definition is identical to the earlier $\text{HWWN}$ except that we gauge the distribution of the $\hat{v}_{j,k}$ using the Gaussian approximation suggested by Corollary 1.

Few tests in the literature have freely available software to compute them, and none, as far as we know, provides functionality to compute theoretical power. Flexible computation of the theoretical power of a test for a given ARMA model is extremely useful as it sheds considerable light on the likely sample size required to detect departures of a given size in a given direction. The theoretical power function of $\text{GENNWN}$ is given by the following.

**Proposition 2** (Test power function)

Let the nominal size of the test be $\alpha$, the Bonferroni corrected size be $\alpha_c = N^{-1} \alpha$ and the Bonferroni critical value $C_{\alpha_c} = \Phi^{-1}(1 - \alpha_c/2)$, where $\Phi$ is the standard normal cumulative distribution function. Then the (approximate)
Power function of the test is given by

\[
P\{\text{Rej } H_0 |f(\omega)\} = P\left(C_{\alpha c} < \max_{(j,k)\in I_T} |\hat{\nu}_{j,k}|\right) = 1 - \prod_{(j,k)\in I_T} \{\Phi_{\eta_{j,k}}(C_{\alpha c} - \nu_{j,k}) - \Phi_{\eta_{j,k}}(-C_{\alpha c} - \nu_{j,k})\},
\]

where \(\nu_{j,k}\) and \(\eta_{j,k}\) are given by (2) and Corollary 1, respectively.

Figure 2 shows the utility and high accuracy of our theoretical power function (solid line) by comparing it with results (circles) from four of the models used in our simulation study later. The top left-hand plot in Figure 2 corresponds to the tests’ statistical size and agrees well with the 5% nominal value.

4 Implementation, simulations and real data example

4.1 Implementation

Our tests are implemented in an R package \texttt{hwwntest} where the spectrum and its normalization are computed using the fast Fourier transform and the regular \texttt{var} variance function and the wavelet coefficients computed using the
wavethresh package (Nason, 2013). The wavelet coefficients are then compared with the $g_m$ distribution from (4) (for Haar) or the normal distribution (for general wavelets), and then the coefficient $p$-values are adjusted by a Bonferroni correction (alternatively, other methods such as false discovery rate can be used).

To compute the theoretical power of the GENWWN test for a general spectrum, we first compute the wavelet coefficients of the normalized spectrum. In particular, the variances $\eta_{j,k}^2 = \langle \psi_{j,k}'(\pi \sigma_X^{-2} f) \rangle$ need to be calculated.

We choose to use the fast approximate methods due to Herrick et al. (2001) and Barber et al. (2002) where $x$ is represented as a linear combination of father wavelets, $\phi_{k,m}$ at finer scales. Then all the rescaled squared wavelets $\psi_{j,k}^2$ in the calculation of $\eta_{j,k}^2$ can be rapidly computed by using father wavelet coefficients at finer scales still.

4.2. Simulation results

We carried out an extensive simulation study to evaluate the empirical size and power of our new tests in comparison with the Box–Ljung, Bartlett and whiteNoise.test() test from Lobato & Velasco (2004), which we call LVWN in the tables. Tables I–III present a selection of results, and more can be found in the work of Nason & Savchev (2014). The nominal size of all hypothesis tests was set to 5%, and results were established via $10^5$ Monte Carlo replications.

Table I shows empirical size results for independent and identically distributed standard normal and Cauchy distributions for sample sizes of $T = 64, 256$ and $1024$. In most cases, the empirical size is close to, and never exceeds, the nominal size for all tests for the standard normal data. However, for Cauchy data, the empirical size for the LVWN test, and to a much lesser extent the Ljung–Box test, dramatically exceeds its nominal value, and hence, we would not recommend its use in heavy-tailed noise scenarios. The periodogram-based tests (HWWN, GENWWN and Bartlett) are not affected, probably because of the stable exponential asymptotic limit of the periodogram under mild distributional conditions.

Table II compares three AR(1) models and shows that results are similar for $\alpha = \pm 0.9$, except that LVWN is less powerful for $\alpha = 0.9$. For $\alpha = 0.3$, the Ljung–Box test with lag = 1 dominates, but this is not surprising because there is a large lag 1 coefficient. In practice, of course, the best lag is not known and, for example, for lag = 20, the SCWN, GENWWN and Bartlett tests dominate the Ljung–Box test, with Bartlett being better for larger $T$.

Table III compares three ARMA models. The GENWWN test dominates for the MA(2) model, the SCWN test dominates for the ARMA(1, 2) model but the GENWWN test does well for larger $T$. For the AR(12) model, the LVWN test performs well, but the HWWN test is second best, and the Bartlett test performs poorly. Nason & Savchev (2014) rerun some of these simulations with heavy-tailed innovations.

| Table I. Empirical size for white noise tests sample size, $T$. |
|---|
| Model | $T$ | NWNN | SCWN | GENWWN | Box $\text{lag = 1}$ | Box $\text{lag = 20}$ | Bartlett | LVWN |
|---|---|---|---|---|---|---|---|---|
| $N(0, 1)$ | 64 | 3.0 | 4.2 | 4.8 | 4.6 | 2.8 | 2.9 | 4.7 |
| | 256 | 3.8 | 4.8 | 4.7 | 4.9 | 4.3 | 3.7 | 4.7 |
| | 1024 | 4.1 | 4.9 | 4.9 | 4.9 | 4.9 | 4.1 | 4.9 |
| Cauchy | 64 | 0.4 | 2.4 | 2.0 | 2.7 | 0.3 | 1.5 | 30.7 |
| | 256 | 1.0 | 2.4 | 5.0 | 2.7 | 5.6 | 2.0 | 63.8 |
| | 1024 | 2.0 | 2.0 | 5.0 | 1.9 | 7.5 | 2.1 | 87.3 |

Truth: independent and identically distributed variates.
Table II. Empirical power for white noise tests, sample size, $T$.

| Model   | $T$ | HWWN | SCWN | GENWWN | Box lag = 1 | Box lag = 20 | Bartlett | LVWN |
|---------|-----|------|------|--------|-------------|-------------|----------|------|
|         |     |      |      |        | lag = 1     | lag = 20     |          |      |
| $\alpha = 0.3$ |     |      |      |        | 26.0        | 4.4         | 12.5     | 7.9  |
|         | 32  | 6.1  | 21.3 | 17.1   |             |             |          |      |
|         | 64  | 16.2 | 50.0 | 28.7   |             |             |          |      |
|         | 256 | 79.1 | 99.0 | 89.4   | 99.7        | 83.7        | 98.7     | 40.7 |
|         | 1024| 100.0| 100.0| 100.0  | 100.0       | 100.0       | 100.0    | 91.6 |
| $\alpha = 0.9$ |     |      |      |        | 26.0        | 4.4         | 12.5     | 7.9  |
|         | 32  | 86.0 | 98.0 | 95.5   | 98.9        | 84.3        | 96.3     | 61.1 |
|         | 64  | 100.0| 100.0| 100.0  | 100.0       | 99.9        | 100.0    | 84.0 |
| $\alpha = -0.9$ |     |      |      |        | 26.0        | 4.4         | 12.5     | 7.9  |
|         | 32  | 96.0 | 100.0| 98.0   | 99.9        | 95.2        | 99.7     | 96.9 |
|         | 64  | 100.0| 100.0| 100.0  | 100.0       | 100.0       | 100.0    | 100.0 |

Truth: AR(1), parameter $\alpha$, standard normal innovations.

Table III. Empirical power for white noise tests, sample size, $T$, standard normal innovations.

| Model   | $T$ | HWWN | SCWN | GENWWN | Box lag = 1 | Box lag = 20 | Bartlett | LVWN |
|---------|-----|------|------|--------|-------------|-------------|----------|------|
|         |     |      |      |        | lag = 1     | lag = 20     |          |      |
| MA(2)   | 64  | 26.7 | 8.3  | 55.0   | 17.0        | 34.7        | 38.3     | 29.8 |
|         | 256 | 93.6 | 8.6  | 99.9   | 17.8        | 99.8        | 98.8     | 77.0 |
|         | 1024| 100.0| 9.0  | 100.0  | 17.9        | 100.0       | 100.0    | 100.0 |
| ARMA(1, 2) | 16  | 42.5 | 93.0 | 19.4   | 93.3        | *           | 80.5     | 58.6 |
|         | 32  | 93.1 | 100.0| 98.0   | 100.0       | 79.1        | 99.7     | 87.6 |
|         | 64  | 100.0| 100.0| 100.0  | 100.0       | 100.0       | 100.0    | 99.3 |
| AR(12)  | 64  | 12.7 | 7.4  | 10.0   | 7.4         | 24.1        | 6.0      | 18.9 |
|         | 256 | 51.9 | 9.0  | 12.2   | 8.9         | 98.6        | 14.6     | 75.9 |
|         | 1024| 98.9 | 9.3  | 13.2   | 9.4         | 100.0       | 36.1     | 100.0 |

The MA(2) model has $\beta_1 = 0, \beta_2 = 0.5$, the ARMA(1, 2) model has $\alpha_1 = -0.4, \beta_1 = -0.8, \beta_2 = 0.4$ and the AR(12) model has $\alpha_1 = \cdots = \alpha_{11} = 0, \alpha_{12} = -0.4$.

The Asterisk indicates that the Box test cannot be carried out because the lag is greater than the number of observations.

Overall conclusions from the simulations are as follows: (i) one should avoid the LVWN test if one suspects heavy tails; (ii) the performance of the Box–Ljung test is highly variable and depends on the number of lags; and (iii) all of the tests perform well against some alternatives and not against others. The GENWWN test performs well, especially for a reasonable $T$.

4.3. Wind speed example

Figure 3 shows an hourly wind speed series of length $T = 128$ recorded at Aberporth, Wales, during 2010. The top left plot shows the series with a non-trivial trend, which we remove by first differencing (top right) and is then tested for consistency with white noise.
Figure 3, bottom left, shows the first differences’ autocorrelation function, which seems close to white noise, although there may be significant autocorrelations at lags 1 and 10. This kind of borderline situation, where the series is close to white noise and where the autocorrelations are close to being assessed as zero, is interesting and challenging for white noise tests. Most tests are less challenged by obvious non-white noise or exact white noise.

The $p$-value for the Ljung–Box test with the lag $p = 1$ is 0.004, which indicates strong evidence to reject the null hypothesis. For $p = 20$, the $p$-value is 0.08, which does not appear to be strong enough to reject $H_0$. Other values of lag $p$ give a variety of other $p$-values, making the assessment of $H_0$ hard. The $p$-values for the other tests are $10^{-15}$ for the LVWN test, 0.026 for the Bartlett test, 1.00 for the HWWN test, 0.011 for the SCWN test and 0.044 for the GENWWN test. Overall, our conclusion is that there is good evidence to reject the null hypothesis of white noise. (The cumulative normalized periodogram used to compute the Bartlett statistics is at the bottom-right plot of Figure 3.)

References

Barber, S, Nason, GP & Silverman, BW (2002), ‘Posterior probability intervals for wavelet thresholding’, Journal of the Royal Statistical Society: Series B (Statistical Methodology), 64, 189–205.

Bartlett, M (1954), ‘Problèmes de l’analyse spectrale des séries temporelles stationnaires’, Publications de l’Institut de statistique de l’Université de Paris, 3, 119–34.
Bartlett, M (1955), *An Introduction to Stochastic Processes with Special Reference to Methods and Applications*, Cambridge University Press, Cambridge.

Benjamini, Y & Hochberg, Y (1995), ‘Controlling the false discovery rate: a practical and powerful approach to multiple testing’, *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 57, 289–300. http://www.jstor.org/stable/2346101.

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

Brockwell, PJ & Davis, RA (1991), *Time Series: Theory and Methods*, Springer, New York.

Craigile, PF & Percival, DB (2005), ‘Asymptotic decorrelation of between-scale wavelet coefficients’, *IEEE Transactions on Information Theory*, 51, 1039–1048.

Daubechies, I (1992), *Ten Lectures on Wavelets*, SIAM, Philadelphia.

Durbin, J (1969), ‘Testing for serial correlation in regression analysis based on the periodogram of least-squares residuals’, *Biometrika*, 56, 1–56.

Durbin, J (1970), ‘Testing for serial correlation in least-squares regression when some of the regressors are lagged dependent variables’, *Econometrica*, 38, 410–421.

Fay, G & Soulier, P (2001), ‘The periodogram of an i.i.d. sequence’, *Stochastic Processes and Their Applications*, 92, 315–343.

Freedman, D & Lane, D (1980), ‘The empirical distribution of Fourier coefficients’, *Annals of Statistics*, 8, 1244–1251.

Guay, A, Guerre, E & Štěpána, L (2013), ‘Robust adaptive rate-optimal testing for the white noise hypothesis’, *Journal of Econometrics*, 176, 134–145.

Herrick, DRM, Nason, GP & Silverman, BW (2001), ‘Some new methods for wavelet density estimation’, *Sankhyā A*, 63, 391–411.

Koen, C (1990), ‘Significance testing of periodogram ordinates’, *Astrophysics Journal*, 348, 700–702.

Kokoszka, P & Mikosch, T (2000), ‘The periodogram at the Fourier frequencies’, *Stochastic Processes and Their Applications*, 86, 49–79.

Kotz, S, Kozubowski, T & Podgórski, K (2001), *The Laplace Distribution and Generalizations*, Springer, New York.

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

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

Lobato, I & Velasco, C (2004), *A simple and general test for white noise*, Econometric Society.

Mallat, SG (1989), ‘A theory for multiresolution signal decomposition: the wavelet representation’, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11, 674–693.

Mallat, SG (1998), *A Wavelet Tour of Signal Processing*, Academic Press, San Diego.

Nason, G (2013), *wavethresh: wavelets statistics and transforms*. R package version 4.6.6. http://CRAN.R-project.org/package=wavethresh.
Nason, GP (2008), *Wavelet Methods in Statistics with R*, Springer, New York.

Nason, G & Savchev, D. (2014). *Supplementary material for “White noise testing using wavelets”*, Technical Report 14:01, Statistics Group, University of Bristol.

Neumann, M (1996), ‘Spectral density estimation via nonlinear wavelet methods for stationary non-Gaussian time series’, *Journal of Time Series Analysis*, 17, 601–633.

Shao, X & Wu, WB (2007), ‘Asymptotic spectral theory for nonlinear time series’, *Annals of Statistics*, 35, 1773–1801.

Simon, MK (2002), *Probability Distributions Involving Gaussian Random Variables*, Springer, New York.

Vidakovic, B (1999), *Statistical Modeling by Wavelets*, Wiley, New York.

von Sachs, R & Neumann, MH (2000), ‘A wavelet-based test for stationarity’, *Journal of Time Series Analysis*, 21, 597–613.

Wasserman, L (2005), *All of Nonparametric Statistics*, Springer, New York.

**Supporting information**

Additional supporting information may be found in the online version of this article at the publisher’s web site.