A simple test of normality for time series

Ignacio N. Lobato and Carlos Velasco
Centro de Investigación Económica
Instituto Tecnológico Autónomo de México (ITAM)
Av. Camino Sta. Teresa 930, México D.F. 10700, Mexico
e-mail:ilobato@itam.mx
Departamento de Econometría y Estadística
Universidad Carlos III de Madrid
Avda. de la Universidad 30, 28911 Leganés (Madrid), Spain.
e-mail:cavelas@est-econ.uc3m.es

Abstract

This article considers testing for normality for correlated data. A common testing procedure employs the skewness-kurtosis test statistic, which has an asymptotic chi-square distribution when the considered series is uncorrelated. However, with time series data it often happens that the model is not correctly specified, so the residual series may exhibit serial correlation, and in other cases the researcher might not be interested in modeling the serial correlation at all. The skewness-kurtosis test is invalid in these situations because it does not take the serial correlation into account. In this paper we propose a simple nonparametric modification of the skewness-kurtosis test that is robust to the presence of serial correlation of a general form. The main feature of our proposed test is its simplicity since it does not require the selection of any user-chosen parameter such as a smoothing number or the order of an approximating model.

Keywords: autocorrelation, Gaussianity, kurtosis, nonparametric, skewness.

1 Introduction

There has been a recent interest in testing for normality for economic and financial data. For instance, Bai and Ng (2001) test for normality in a set of macroeconomic series while Bontemps and Meddahi (2002) emphasize financial applications. Kilian and Demiroglu (2001) present a variety of cases where testing for normality is of interest for econometricians. These applications include financial and economic ones where, for instance, assessing whether abnormal financial profits or economic growth rates are normal is important for the specification of financial and economic models. They also present methodological applications where testing for normality is a previous step for the design of some tests, such as tests for structural stability or tests of forecast encompassing.

In econometrics, testing for normality is customarily performed by means of the skewness-kurtosis test. The main reason for its widespread use is its straightforward implementation and interpretation. The skewness-kurtosis test statistic is the sum of the square of the sample skewness and the excess kurtosis coefficients properly standardized by their asymptotic variances in the white noise case, 6 and 24, respectively. Implementing the skewness-kurtosis test is very simple since it compares the skewness-kurtosis test statistic against upper critical values of a chi-squared distribution with two degrees of freedom (χ_2^2). This test is typically applied to the residual series of dynamic econometric models, e.g. Lütkepohl (1991, Section 4.5).

In many empirical studies with time series data, the application of the skewness-kurtosis test is questionable, though. The reason is that the previous asymptotic variances are correct under the assumption that the model is correctly specified, implying that the sequence under examination is uncorrelated. However, in many occasions, either the researcher might specify incorrectly the model or she might not even be interested in modeling the serial correlation. In both cases, when the considered data is correlated, the asymptotic variances are no longer 6 and 24 but some functions of all the autocorrelations. In this situation the skewness-kurtosis test is invalid since it does not control asymptotically the type I error. There are two strategies to potentially perform an asymptotically valid test. The first consists on carrying out a two step test where the skewness-kurtosis is applied after testing that the considered series is uncorrelated. The second is done by modifying the skewness-kurtosis test to account for the possibility of serial correlation. The former approach is problematic because there is an obvious pre-test problem in such a sequential procedure and, furthermore, testing for uncorrelatedness for non-Gaussian series is rather challenging, see Lobato, Nankervis and Savin (2002).

In this paper we follow the latter approach and propose a modification of the skewness-kurtosis test statistic that is valid for serially correlated data. The proposed test statistic is a very simple modification of the skewness-kurtosis test statistic and it also has an asymptotic χ_2^2 null distribution under weak dependent conditions. The modification is based on straightforward consistent estimators of the asymptotic variances of the sample skewness and the sample excess kurtosis. Besides its simplicity, the main feature of our procedure is that, as opposed to most of the literature concerning consistent variance estimation, e.g. Robinson and Velasco (1997), we are able to provide consistent estimators without introducing any user-chosen object such as a smoothing number, a kernel function or an approximating parametric model. These user-chosen tools are theoretical devices that are useful for establishing asymptotic results, but they are a nuisance for the applied researcher who faces the problem of choosing them for her particular problem. Certainly, in some cases asymptotic theory has been established to justify the automatic selection of these tools using some optimality criteria. However, these criteria are typically designed for estimation problems and they can be questionable in a testing framework, see the discussion in Robinson (1998, p.1165).

Our test can employ either frequency or time domain estimators of the asymptotic variances of the sample skewness and the sample excess kurtosis. Although the proposed test is based on a time domain estimator, in the technical part in the appendices we stress a frequency domain estimator since it is relatively easier to handle theoretically. In addition, for conciseness of exposition, we only analyze the univariate case.

We end this section with a brief comment on two papers related to ours. First, Bai and Ng (2001) also consider a modification of the skewness-kurtosis test statistic that is able to account for serial correlation. However, they rely on standard consistent variance estimators such as kernel methods, that present the difficulty of arbitrarily selecting some inputs, as we have commented above. Second, Bontemps and Meddahi (2002) test for normality using Hermite polynomials of arbitrary orders. In their framework, the skewness-kurtosis test emerge when the considered Hermite polynomials are of orders 3 and 4. However, similarly to Bai and Ng, Bontemps and Meddahi address the problem of serial correlation by employing common variance estimators such as a kernel estimator which is briefly described at the beginning of Section 4.

The plan of the article is the following. Section 2 presents the framework. Section 3 introduces the proposed test statistic and studies its asymptotic theory. Section 4 discusses the proposed variance estimators. Section 5 examines the case where the considered series are the residuals of regression and time series models. Section 6 considers the finite sample performance of the proposed test in a brief Monte Carlo exercise. The technical material is included in the Appendices.

2 Framework

Notation. Let x_t be an ergodic strictly stationary process with mean μ and centered moments denoted by $\mu_k = E(x_t - \mu)^k$ for k natural, with $\widehat{\mu}_k = n^{-1}\sum_{t=1}^{n}(x_t - \overline{x})^k$ being the corresponding sample moments where $\overline{x}$ is the sample mean and n is the sample size. In addition, $\gamma(j)$ denotes the population autocovariance of order j, $\gamma(j) = E[(x_1 - \mu)(x_{1+j} - \mu)]$, and $\widehat{\gamma}(j)$ is the corresponding sample autocovariance, $\widehat{\gamma}(j) = n^{-1}\sum_{t=1}^{n-|j|}(x_t - \overline{x})(x_{t+|j|} - \overline{x})$. Notice that $\mu_2 = \gamma(0)$. Let $f(\lambda)$ be the spectral density function of x_t, defined by

$$\gamma(j) = \int_{\Pi} f(\lambda)\exp(ij\lambda)d\lambda \quad j = 0, 1, 2, \ldots, \tag{1}$$

where $\Pi = [-\pi, \pi]$, and let $I(\lambda)$ denote the periodogram $I(\lambda) = |w(\lambda)|^2$ where $w(\lambda) = (2\pi n)^{-1/2}\sum_{t=1}^{n} x_t \exp(it\lambda)$. In addition, $\kappa_q(j_1, \ldots, j_{q-1})$ denotes the q-th order cumulant of $x_1, x_{1+j_1}, \ldots, x_{1+j_{q-1}}$ and the marginal cumulant of order q is $\kappa_q = \kappa_q(0, \ldots, 0)$.

Null and alternative hypotheses. In principle, the null hypothesis of interest is that the marginal distribution of x_t is normal. For the independent case, omnibus tests for this null hypothesis such as the Shapiro-Wilk (1965) which is based on order statistics, or tests based on the distance between the empirical distribution function and the normal cumulative distribution function such as the Kolmogorov-Smirnov or the Cramér von-Mises have been proposed, see Mardia (1980) for a survey. For the independent case, these omnibus tests are consistent, but it has been shown that their finite sample performance can be very poor, e.g. Shapiro et al.(1968). For the weak dependent case, no such analysis exists because inference with these omnibus test statistics is problematic since their asymptotic distributions are nonstandard and case dependent. Hence, the standard application of these tests to weak dependent time series sequences is invalid, see Gleser and Moore (1983). The only developed test of which we are aware is the one by Epps (1987) which is based on the characteristic function. However, Epps' procedure is based on restrictive theoretical assumptions and in practice its implementation is complicated.

Instead of testing that the marginal distribution function is normal, in practice the common procedure tests whether the third and fourth marginal moments coincide with those of the normal distribution. Equivalently, in terms of the cumulants, it is tested that the third and fourth marginal cumulants are zero instead of testing that all higher order marginal cumulants are zero. We follow this practice, and in this paper the considered null hypothesis is

$$H_0 : \mu_3 = 0 \text{ and } \mu_4 = 3\mu_2^2, \tag{2}$$

that is, both the skewness and the excess-kurtosis are zero. The alternative hypothesis is the negation of the null, that is,

$$\overline{H}_0 : \mu_3 \neq 0 \text{ or } \mu_4 \neq 3\mu_2^2. \tag{3}$$

The skewness-kurtosis test statistic. The null hypothesis (2) is commonly tested using the skewness-kurtosis test statistic, e.g. Bowman and Shenton (1975),

$$SK = \frac{n\widehat{\mu}_3^2}{6\widehat{\mu}_2^3} + \frac{n(\widehat{\mu}_4 - 3\widehat{\mu}_2)^2}{24\widehat{\mu}_2^4},$$

which is typically compared against upper critical values of a χ_2^2 distribution. Apart from the fact that Jarque and Bera (1987) have shown the optimality of this test within the Pearson family of distributions, the popularity of this approach resides in its simplicity as we mentioned above. In fact, nowadays most econometrics packages report customarily the SK test which is called the Jarque-Bera test.

The SK test procedure is justified in the following grounds. When the considered series x_t is an uncorrelated Gaussian process, the following limiting result holds

$$\sqrt{n}\begin{pmatrix} \widehat{\mu}_3 \\ \widehat{\mu}_4 - 3\widehat{\mu}_2^2 \end{pmatrix} \to_d N\begin{pmatrix} 6\mu_2^3 & 0 \\ 0 & 24\mu_2^4 \end{pmatrix}, \tag{4}$$

where the symbol $\to_d$ denotes convergence in distribution. However, when x_t is a Gaussian process satisfying the weak dependent condition

$$\sum_{j=0}^{\infty} |\gamma(j)| < \infty, \tag{5}$$

the result (4) is replaced by

$$\sqrt{n}\begin{pmatrix} \widehat{\mu}_3 \\ \widehat{\mu}_4 - 3\widehat{\mu}_2^2 \end{pmatrix} \to_d N\begin{pmatrix} 6F^{(3)} & 0 \\ 0 & 24F^{(4)} \end{pmatrix}, \tag{6}$$

where

$$F^{(k)} = \sum_{i=-\infty}^{\infty} \gamma(i)^k, \tag{7}$$

for $k = 3, 4$, see Lomnicki (1961) and Gasser (1975). Notice that condition (5) guarantees that all $F^{(k)}$ are well defined since it entails that $\sum |\gamma(j)|^r < \infty$, for all natural r.

Hence, when the series exhibits serial correlation, the SK test is invalid since the denominators of its components do not estimate consistently the true asymptotic variances in (6), implying that asymptotically its rejection probabilities do not coincide with the desired nominal levels under the null hypothesis.

Given the complicated form of the asymptotic variances in (6), Epps (1987) conjectured that "these moment results do not of themselves support an operational test of the Gaussian hypothesis". In the next section we show that Epps' conjecture is wrong and provide a simple and operational test for the null hypothesis (2).

3 The generalized skewness-kurtosis test

In the previous section we have seen that the SK test is invalid when the considered process x_t exhibits serial correlation. Looking at (6) two natural solutions appear. The first one consists on modifying the SK test statistic by including consistent estimators of $F^{(3)}$ and $F^{(4)}$ in the denominators of its components. This solution is proposed in Gasser (1975, section 6) who suggested truncating the infinite sums which appear in the asymptotic variances. However, he did not provide any formal analysis or any recommendation about the selection of the truncation number. As we will see, our proposed procedure overcomes these difficulties since it does not require the selection of any truncation number. The second solution is estimating the unknown asymptotic variances with the bootstrap, that is, employing the SK test statistic with bootstrap based critical values. In this paper we follow the first approach because implementing the bootstrap in a time series context is problematic since generally valid bootstrap procedures require the introduction of an arbitrary user chosen number, typically a block length, e.g. Davison and Hinkley (1997, chapter 8). Furthermore, in our case the bootstrap does not present a clear theoretical advantage since the SK statistic is not asymptotically pivotal.

Before introducing our test statistic, let consider the following estimator of $F^{(k)}$ which is the sample analog of (7)

$$\widehat{F}^{(k)} = \sum_{j=1-n}^{n-1} \widehat{\gamma}(j)^k. \tag{8}$$

In the next section we consider alternative versions of this estimator and study their large sample properties, in particular, Theorem 1 establishes the consistency of $\widehat{F}^{(k)}$ for $F^{(k)}$ for Gaussian processes that satisfy condition (5). Then, our proposed test statistic, the Generalized SK statistic, is

$$G = \frac{n\widehat{\mu}_3^2}{6\widehat{F}^{(3)}} + \frac{n(\widehat{\mu}_4 - 3\widehat{\mu}_2)^2}{24\widehat{F}^{(4)}}.$$

The G statistic does not require the introduction of any user chosen number and, in view of (6) and Theorem 1 in the next section, the proposed test for (2) consists on comparing the G test statistic against upper critical values from a χ_2^2 distribution.

In the next assumption we introduce the class of processes under the alternative hypothesis for which both $\tilde{F}^{(k)}$ and $\hat{F}^{(k)}$ converge to bounded positive constants, and hence the G test is consistent. Notice that the conditions of Gasser (1975) which involve summability conditions of cumulants of all orders are relaxed to cumulants up to order 16 using an extension of Theorem 3 in Rosenblatt (1985, p.58).

Assumption A. The process x_t *satisfies* $Ex_t^{16} < \infty$, *and for* $q = 2, 3, \ldots, 16$

$$\sum_{j_1=-\infty}^{\infty} \cdots \sum_{j_{q-1}=-\infty}^{\infty} |\kappa_q(j_1, \ldots, j_{q-1})| < \infty, \tag{9}$$

and for $k = 3, 4$,

$$\sum_{j=1}^{\infty} \left[E \left| \left(Ex_0^k \mid \Im_{-j} \right) - \mu_k \right|^2 \right]^{1/2} < \infty, \tag{10}$$

where $\Im_{-j}$ *denotes the* σ*-field generated by* x_t, $t \leq -j$, *and, for* $k = 3, 4$,

$$E(x_0^k - \mu_k)^2 + 2\sum_{j=1}^{\infty} E\left[\left(x_0^k - \mu_k \right) \left(x_j^k - \mu_k \right) \right] > 0. \tag{11}$$

Assumption A is a weak dependent assumption that implies that the higher order spectral densities up to the sixteenth order are bounded and continuous. For the case $q = 2$, expression (9) implies that condition (5) holds. We require finite moments up to the sixteenth order because we need to evaluate the variance of the fourth power of the sample autocovariances. Notice that condition (11) assures that the asymptotic variances of estimates are positive.

The following lemma establishes the asymptotic properties of the G test.

Lemma 1. a) Under the null hypothesis and for Gaussian processes that satisfy condition (5), $G \to_d \chi^2_2$. b) *Under the alternative hypothesis (3) and under Assumption A, the test based on G is consistent.*

The asymptotic null distribution is straightforward to derive given the consistency of $\widehat{F}^{(k)}$ for $F^{(k)}$ that is proved in Theorem 1 in the next section. The proof of b) is omitted since it follows easily using that under the alternative hypothesis $\widehat{F}^{(k)}$ converges to a bounded positive constant (by (9) and (11)), while the numerator of G diverges.

4 Consistent variance estimators

Following the literature on nonparametric estimation of asymptotic covariance matrices, the standard approach to estimate consistently $F^{(k)}$ employs a smoothed estimator such as

$$\sum_{j=1-n}^{n-1} w_j \widehat{\gamma}(j)^k. \tag{12}$$

In (12) the weights $\{w_j\}$ are usually obtained through a lag window $\{w_j = w(j/M)\}$ such that the weight function $w(\cdot)$ verifies some regularity properties and M is a smoothing number which grows slowly with n. Note that the introduction of the smoothing number leads to estimators whose rate of convergence is usually slower than the parametric rate. We stress that in this approach the weights $\{w_j\}$ provide a nonstochastic dampening on the $\widehat{\gamma}(j)^k$ for large j. Due to this dampening, the estimator in (12) is consistent for (7) as it happens in the case $k = 0$, where $f(0)$ is consistently estimated by autocorrelation robust estimators, e.g. Robinson and Velasco (1997).

As commented in the introduction, the main problem with the smoothing approach is that statistical inference can be very sensitive to the selection of the user-chosen weights; in our context, the discussion in Section I in Robinson (1998) is especially relevant. In the absence of a clear and rigorously justified procedure to select the smoothing number in our testing framework, we prefer to analyze estimators which do not require any smoothing.

Our first estimator $\widehat{F}^{(k)}$ introduced in equation (8) also admits a frequency domain version, see Appendix A. For technical reasons, in this paper we consider a second estimator that can be motivated by writing $F^{(k)}$ in terms of the spectral density function of the x_t

process using (1)

$$
\begin{aligned}
F^{(k)} &= \sum_{j=-\infty}^{\infty} \gamma(j)^k = \sum_{j=-\infty}^{\infty} \prod_{h=1}^{k} \left\{ \int_{\Pi} f(v_h) \exp(ijv_h) dv_h \right\} \\
&= 2\pi \int_{\Pi^{k-1}} f(v_1 + \cdots + v_{k-1}) \prod_{h=1}^{k-1} \{ f(v_h) dv_h \} . \qquad (13)
\end{aligned}
$$

The sample analog of the previous equation renders the following alternative estimator for $F^{(k)}$,

$$
\tilde{F}^{(k)} = \frac{(2\pi)^k}{n^{k-1}} \sum_{j_1=1}^{n-1} \cdots \sum_{j_{k-1}=1}^{n-1} I(\lambda_{j_1}) \cdots I(\lambda_{j_{k-1}}) I(\lambda_{j_1} + \cdots + \lambda_{j_{k-1}}), \qquad (14)
$$

where $\lambda_j = 2\pi j/n$. The estimator $\tilde{F}^{(k)}$ can also be written in the time domain by plugging

$$
I(\lambda_j) = \frac{1}{2\pi} \sum_{t=1-n}^{n-1} \exp(it\lambda_j) \hat{\gamma}(t), \quad j \neq 0, \bmod n, \qquad (15)
$$

into equation (14). After some algebra, in Appendix A it is shown that

$$
\tilde{F}^{(k)} = \sum_{t=1-n}^{n-1} \hat{\gamma}(t) \left\{ \hat{\gamma}(t) + \hat{\gamma}(n - |t|) \right\}^{k-1} . \qquad (16)
$$

Notice that both expressions for $\tilde{F}^{(k)}$ are numerically identical, but in the appendices, for technical reasons, we stress the frequency domain version (14). Expression (14) guarantees that $\tilde{F}^{(k)}$ is positive in finite samples.

The next theorem states the consistency of $\tilde{F}^{(k)}$ and $\hat{F}^{(k)}$ for $F^{(k)}$. This theorem is the main technical contribution of the paper. Its proof is in Appendix B.

Theorem 1. Under the null hypothesis, for Gaussian time series processes that satisfy condition (5), (a) $\tilde{F}^{(k)} = F^{(k)} + o_p(1)$ *and (b)* $\hat{F}^{(k)} - \tilde{F}^{(k)} = o_p(1)$ *for* $k = 3, 4$.

At first look, consistency of $\widehat{F}^{(k)}$ and $\tilde{F}^{(k)}$ could be surprising since no smoothing parameter has been introduced. Robinson (1998) analyzed a special regression model where smoothing was not necessary for establishing consistency of asymptotic covariance matrix estimators. The reason was that the specific form of the covariance matrix that he considered, see his equation (1.2), allowed for a stochastic dampening of some sample autocovariances by other sample autocovariances. The time domain versions (8) and (16) provide a similar intuition where the powers of the sample autocovariances provide the stochastic dampening factors.

In the frequency domain, (13) provides a complementary explanation. Recall that in time series the standard problem is that the relevant asymptotic variance depends on the spectral density function evaluated at a unique point, typically the zero frequency, $f(0)$. However, in our case (13) shows that the asymptotic variance, $F^{(k)}$, is a convolution of the spectral density function, instead of a single value. Intuitively, in the first case a user-chosen smoothing number is required to estimate the local quantity, $f(0)$, whereas in our case no such number is needed because we are estimating a global quantity.

5 Residual testing

The previous sections analyze the case where raw data are under examination. However, in practice the test is commonly applied to the residuals of regression or time series models. Again, two approaches can be used: first, the G test that we propose and second, employing the SK statistic with bootstrap based critical values. The bootstrap has been employed by Kilian and Demiroglu (2000). However, as commented in Section 3, application of the bootstrap is not an obvious task in a time series context. Kilian and Demiroglu perform a parametric bootstrap that could be justified if the model were correctly specified, although in this case the SK test would also be asymptotically valid. However, in the absence of the knowledge of the true data generating process, a parametric bootstrap is invalid, that is, there is no guarantee that the type I error is controlled properly asymptotically. As commented above, bootstrap procedures generally valid in time series require the introduction of a user chosen number, typically a block number, complicating statistical inference in finite samples.

Next, we introduce a general assumption that validates the use of the G statistic applied to the residuals of many dynamic econometric models where the correlation structure is not correctly specified or it is not specified at all. In this section, $\hat{x}_t$ denote the residuals of the regression or time series model, and x_t denote the true disturbances.

Assumption B. Let the Gaussian process x_t satisfy (5) and let $e_t = x_t - \hat{x}_t$ satisfy

$$\sum_{t=1}^{n} e_t^2 = O_p(1), \text{ and } \sum_{t=1}^{n} e_t^4 = o_p(n^{-1/4}). \tag{17}$$

The first condition in (17) guarantees the consistency of the estimates of $F^{(k)}$ based on residuals while the second guarantees that the residual SK test has the same asymptotic distribution as the original SK test. The previous *Assumption B* is very general and covers many interesting cases such as linear regressions with possible trending stochastic and deterministic regressors which satisfy Grenander's conditions and weakly dependent errors. In this case $e_t = (\hat{\beta} - \beta)' Z_t$, where Z_t is a p-dimensional sequence of regressors, so (17) implies that $\sum_{t=1}^{n} e_t^2 = (\hat{\beta} - \beta)' Z' Z (\hat{\beta} - \beta) = O_p(1)$, allowing for the components of $\hat{\beta}$ to have different convergence rates. A leading example with stochastic Z_t is a regression between cointegrated variables. For stationary Z_t, another interesting application is when $\hat{x}_t$ are the residuals obtained through possibly misspecified AR(p) regressions, that is, $\hat{x}_t = y_t - \hat{\beta}' Z_t$ with $Z_t = (y_{t-1}, \ldots, y_{t-p})'$, and $\sqrt{n}(\hat{\beta} - \beta) = O_p(1)$ for some vector β such that the polynomial $\beta(\omega) = 1 - \sum_{j=1}^{p} \beta_j \omega^j$ has no roots on or inside the unit circle. For this case, if *Assumption B* holds for y_t, the limit process $x_t = y_t - \beta' Z_t = \beta(L) y_t$ inherits the weak dependence properties of y_t, but notice that x_t is autocorrelated unless y_t follows an AR(q) process with $q \leq p$.

In Appendix C we prove the following theorem which shows that the use of residuals does not affect the consistent studentization that we propose in this paper.

Theorem 2. Under the null hypothesis (2) and Assumption B, for $k = 3, 4$,

$$\sum_{1-n}^{n-1} \hat{\gamma}_{\hat{x}}(j)^k = \sum_{1-n}^{n-1} \hat{\gamma}_x(j)^k + o_p(1).$$

Finally, using the previous theorem and Hölder's inequality, it is straightforward to prove the next lemma which establishes that the asymptotic null distribution of the G test statistic applied to the residuals of many dynamic econometric models whose correlation structure is ignored or misspecified is still χ^2_2, and that the G test is consistent under the broad class of alternatives considered in *Assumption B*.

Lemma 2. Let $\widehat{G}$ *be the test statistic* G *calculated from residuals* $\hat{x}_t$*. a) Under the null hypothesis (2) and Assumption B,* $\widehat{G} \to_d \chi^2_2$. b) *Under the alternative hypothesis (3) and Assumptions A and B, the test based on* $\widehat{G}$ *is consistent.*

6 Finite Sample Performance

This section compares briefly the finite sample behavior of the previous tests. Under the null hypothesis we generate data from an AR(1) process $x_t = \phi x_{t-1} + \varepsilon_t$ where ε_t is independent and identically distributed N(0,1) and the autoregressive parameter ϕ takes three values: $-0.5, 0$, and 0.5.

Along with the null hypothesis (2), we consider also testing the null that the skewness is zero by using the first components of the SK and G statistics. Namely we compute the skewness test statistic $S = n\widehat{\mu}_3^2/6\widehat{\mu}_2^3$ and the generalized skewness test statistic $GS = n\widehat{\mu}_3^2/6\widehat{F}^{(3)}$ and compare them with upper critical values from a χ_1^2. We have not reported the results of a kurtosis test because of the well-known slow convergence of the sample kurtosis to the normal asymptotic distribution even in the white noise case, e.g. Bowman and Shenton (1975, p.243). In Table I we report the empirical rejection probabilities for the tests for three sample sizes, n=100, 500 and 1000, and three nominal levels, α=0.10, 0.05 and 0.01. In these experiments 5,000 replications are carried out.

The main conclusions derived from this table are the following. First, the S test is not reliable since it severely under-rejects when $\phi = -0.5$ and substantially over-rejects when $\phi = 0.5$. This result could be expected because when ϕ is negative $\sum_{j=1}^{\infty} \gamma_j^3$ is negative, leading to overestimation of the asymptotic variance, and then to under-rejection of the S test; while when ϕ is positive the opposite effect occurs. The most interesting evidence is the magnitude of these distortions that are very large for $\phi = -0.5$ and all sample sizes, while for $\phi = 0.5$ the distortions are increasing steadily with the sample size. On the contrary, for the GS test the empirical rejection probabilities are very close to the nominal levels for all the parameter values and all sample sizes. Finally, the SK test, which is the sum of the skewness test and the kurtosis test, inherits their characteristics. Notice that for the specific case where $\phi = -0.5$, there is a fair amount of compensation between the skewness and kurtosis, making the distortions of the SK test much smaller that those of its components. The G test inherits the slow convergence from the kurtosis, but using the white noise case as benchmark, it appears to be robust to the presence of serial correlation.

We also conducted power experiments for data generated by the previous AR(1) model for five different distributions: standard log-normal, student's t with 10 and 20 degrees of freedom, χ_1^2 and χ_{10}^2. Table II reports the power results for three sample sizes, n=50, 100, 200, for a test with a 5% nominal level. In these experiments 2,000 replications are carried out. Table II indicates that the sign of the autocorrelation has no relevance in terms of power. As could be expected for heavily skewed distributions such as the lognormal or the χ_1^2, the empirical rejection probabilities are close to one even for very moderate sample sizes.

We end with a suggestion on further research. In this section we have seen that for small sample sizes, due to the slow convergence of the sample kurtosis coefficient, the G test presents significant size distortions even in the white noise case. One potential way of improving the finite sample performance is by using the bootstrap. Since the G test statistic is asymptotically pivotal, it can be expected that application of the bootstrap will deliver an asymptotic refinement. Hence, it would be interesting to study the implementation of the G statistic with bootstrap based critical values.

7 Appendices

Appendix A

This appendix provides the alternative versions of $\hat{F}^{(k)}$ and $\tilde{F}^{(k)}$. First, the $\hat{F}^{(k)}$ estimator can be written in the frequency domain as follows

$$\begin{aligned}
\hat{F}^{(k)} &= \sum_{j=1-n}^{n-1} \hat{\gamma}(j)^k = \sum_{j=1-n}^{n-1} \prod_{h=1}^{k} \left\{ \int_{\Pi} I(v_h) \exp(ijv_h) dv_h \right\} \\
&= \prod_{h=1}^{k} \left\{ \int_{\Pi} I(v_h) dv_h \right\} \sum_{j=1-n}^{n-1} \exp\{ij(v_1 + \cdots + v_k)\} \\
&= \int_{\Pi^k} I_{x-\bar{x}}(v_1) \cdots I_{x-\bar{x}}(v_k) D_n(v_1 + \cdots + v_k) dv_1 \cdots dv_k,
\end{aligned}$$

where $D_n(v) = \sum_{j=1-n}^{n-1} \exp(ijv)$ satisfies $\int_{\Pi} D_n(v)dv = 2\pi$ and $D_n(v) \to 2\pi\delta(v=0)$ as $n \to \infty$, where δ represents the Dirac's delta function. Hence, for n large we obtain the following approximate expression for $\hat{F}^{(k)}$ in the frequency domain

$$\hat{F}^{(k)} \approx 2\pi \int_{\Pi^{k-1}} I_{x-\bar{x}}(\lambda_1) \cdots I_{x-\bar{x}}(\lambda_{k-1}) I_{x-\bar{x}}(\lambda_1 + \cdots + \lambda_{k-1}) d\lambda_1 \cdots d\lambda_{k-1}. \tag{18}$$

Equation (14) is the natural discrete approximation of (18).

Second, in order to obtain the time domain expression of $\tilde{F}^{(k)}$ we just plug in (15) into equation (14) to get

$$\tilde{F}^{(k)} = \frac{1}{n^{k-1}} \sum_{t_1=1-n}^{n-1} \hat{\gamma}(t_1) \cdots \sum_{t_{k-1}=1-n}^{n-1} \hat{\gamma}(t_{k-1}) \sum_{t_k=1-n}^{n-1} \hat{\gamma}(t_k)$$

$$\times \sum_{j_1=1}^{n} \cdots \sum_{j_{k-1}=1}^{n} \exp\left\{i\left(t_1\lambda_{j_1} + \cdots + t_{k-1}\lambda_{j_{k-1}} + t_k(\lambda_{j_1} + \cdots + \lambda_{j_{k-1}})\right)\right\}$$

$$= \frac{1}{n^{k-1}} \sum_{t_1=1-n}^{n-1} \hat{\gamma}(t_1) \cdots \sum_{t_{k-1}=1-n}^{n-1} \hat{\gamma}(t_{k-1}) \sum_{t_k=1-n}^{n-1} \hat{\gamma}(t_k)\phi_n(\lambda_{t_1} + \lambda_{t_k}) \cdots \phi_n(\lambda_{t_{k-1}} + \lambda_{t_k}),$$

where $\phi_n(\lambda) = \sum_{t=1}^{n} \exp(it\lambda)$. Finally, using that $\phi_n(\lambda_j) = 0$ if $\lambda_j = 2\pi j/n$, $j \neq 0 \bmod n$ and $\phi_n(0) = n$, we obtain (16).

Appendix B

Proof of Theorem 1(a)

We just report the analysis for $\tilde{F}^{(3)}$ because the one for $\tilde{F}^{(4)}$ is similar, but notationally more involved. We prove consistency by checking the sufficient conditions that $\tilde{F}^{(3)}$ is asymptotically unbiased and that its variance goes to zero as $n \to \infty$.

First, we consider the expectation of $\tilde{F}^{(3)}$,

$$E[\tilde{F}^{(3)}] = \frac{(2\pi)^3}{n^2} \sum_{j_1=1}^{n-1} \sum_{j_2=1}^{n-1} E\left[I(\lambda_{j_1})I(\lambda_{j_2})I(\lambda_{j_1} + \lambda_{j_2})\right].$$

Using the definition of $I(\lambda)$,

$$\begin{aligned} & E\left[I(\lambda_{j_1})I(\lambda_{j_2})I(\lambda_{j_1} + \lambda_{j_2})\right] \\ = \; & E\left[w(\lambda_{j_1})w(-\lambda_{j_1})w(\lambda_{j_2})w(-\lambda_{j_2})w(\lambda_{j_1} + \lambda_{j_2})w(-\lambda_{j_1} - \lambda_{j_2})\right] \qquad (19) \\ = \; & \sum_{\nu} cum(\nu_1) \cdots cum(\nu_q), \end{aligned}$$

where the summation in ν runs for all possible partitions $\nu = \nu_1 \cup \cdots \cup \nu_q$, $q = 1, 2, 3$ of the 6-tuple

$$\{j_1,\ -j_1,\ j_2,\ -j_2,\ j_1 + j_2,\ -j_1 - j_2\} \qquad (20)$$

such that $\nu_i = \{\nu_i(1), \ldots, \nu_i(p_i)\}$ and $\sum_{i=1}^{q} p_i = 6$, and where $cum(\nu_i)$ stands for $cum(w(\lambda_{\nu_i(1)}), \ldots, w(\lambda_{\nu_i(p_i)}))$, see Brillinger (1981, pp. 20-21).

In order to evaluate the expectation (19), by Gaussianity the only cumulants different from zero are second order cumulants, κ_2, with $q = 3$. Hence

$$E[\tilde{F}^{(3)}] = \frac{1}{n^5} \sum_{j_1=1}^{n-1} \sum_{j_2=1}^{n-1} \left\{ \sum_{\kappa_2^3} \int_{\Pi^3} \prod_{i=1,2,3} \left[f(\mu_i)\phi_n(\mu_i + \lambda_{\nu_i(1)})\phi_n(\lambda_{\nu_i(2)} - \mu_i) d\mu_i \right] \right\}, \qquad (21)$$

where the sum in κ_2^3 is for all the different 3-tuples $\nu_1 \cup \nu_2 \cup \nu_3$ of pairs $\nu_i = (\nu_i(1), \nu_i(2))$ formed with all the permutations of the coefficients in (20). In fact, following Brillinger

(1981, Theorem 4.3.1), the only relevant combinations in the sum in κ_2^3 are those for which $\nu_i(1)+\nu_i(2) = 0 \bmod n$, $i = 1,2,3$. Therefore, using that $|\phi_n(\mu)| \leq 2\min\{|\mu|^{-1}, n\}$, see Zygmund (1977, pp.49-51), and the continuity of $f(\lambda)$ implied by (5), we obtain that (21) is

$$\begin{aligned} E[\tilde{F}^{(3)}] &= \frac{(2\pi)^3}{n^2} \sum_{j_1=1}^{n-1} \sum_{j_2=1}^{n-1} \left\{ \sum_{\kappa_2^3} \int_{\Pi^3} \prod_{i=1,2,3} \left[f(\mu_i) \Phi_n^{(2)}(\mu_i - \lambda_{\nu_i}) d\mu_i \right] \right\} + o(1) \\ &= \frac{(2\pi)^3}{n^2} \sum_{j_1=1}^{n-1} \sum_{j_2=1}^{n-1} f(\lambda_{j_1}) f(\lambda_{j_2}) f(\lambda_{j_1+j_2}) + o(1) \\ &= 2\pi \int_{\Pi^2} f(\lambda) f(\lambda) f(\lambda+\mu) d\lambda d\mu + o(1) \\ &= F^{(3)} + o(1), \qquad as\; n \to \infty, \end{aligned} \tag{22}$$

where $\Phi_n^{(2)}(\mu) = (2\pi n)^{-1} |\phi_n(\mu)|^2$ and $\int_\Pi \Phi_n^{(2)}(\mu) d\mu = 1$.

Second, we study the variance of $\tilde{F}^{(3)}$,

$$\mathrm{Var}[\tilde{F}^{(3)}] = cum(\tilde{F}^{(3)}, \tilde{F}^{(3)}) = \sum_\nu cum(\nu_1) \cdots cum(\nu_q).$$

Now, we need to consider all the indecomposable partitions $\nu = \nu_1 \cup \cdots \cup \nu_q$, $q = 1, \ldots, 6$ of the following array with 12 elements,

$$\begin{array}{cccccc} j_1 & -j_1 & j_2 & -j_2 & j_1+j_2 & -j_1-j_2 \\ j_1' & -j_1' & j_2' & -j_2' & j_1'+j_2' & -j_1'-j_2'. \end{array} \tag{23}$$

By Gaussianity, the relevant partitions only involve six second order cumulants, that is,

$$Var[\tilde{F}^{(3)}] = \frac{1}{n^{10}} \sum_{j_1=1}^{n-1} \sum_{j_2=1}^{n-1} \sum_{j_1'=1}^{n-1} \sum_{j_2'=1}^{n-1} \left\{ \sum_{\kappa_2^6} \int_{\Pi^6} \prod_{i=1}^{6} \left\{ f(\mu_i) \phi_n(\mu_i + \lambda_{\nu_i(1)}) \phi_n(\lambda_{\nu_i(2)} - \mu_i) d\mu_i \right\} \right\} \tag{24}$$

where the sum in κ_2^6 is for all the different 6-tuples $\nu = \nu_1 \cup \cdots \cup \nu_6$ of pairs $\nu_i = (\nu_i(1), \nu_i(2))$ constructed in such a way that at least one ν_i in ν has elements in each of the rows of the array (23) to guarantee an indecomposable partition. Following the same arguments, the only terms that contribute to the leading term of the variance of $\tilde{F}^{(3)}$ are those in (24) characterized by a restriction $\nu_i(1) + \nu_i(2) = 0 \bmod n$, for just one $i \in \{1, \ldots, 6\}$ (e.g. $j_1 = -j_1'$). Then, taking into account all the possible partitions (6×3) and using the continuity of f, the variance of $\tilde{F}^{(3)}$ is

$$\begin{aligned} \mathrm{Var}[\tilde{F}^{(3)}] &= \frac{(2\pi)^6}{n^4} 18 \sum_{j_1=1}^{n-1} \sum_{j_2=1}^{n-1} \sum_{j_3=1}^{n-1} f^2(\lambda_{j_1}) f(\lambda_{j_2}) f(\lambda_{j_3}) f(\lambda_{j_1} + \lambda_{j_2}) f(\lambda_{j_1} + \lambda_{j_3}) + o(n^{-1}) \\ &= O(n^{-1}) = o(1) \end{aligned} \tag{25}$$

as $n \to \infty$. Hence, from (22) and (25) we conclude that $\tilde{F}^{(3)} = F^{(3)} + o_p(1)$.

Proof of Theorem 1(b)

Notice that

$$\begin{aligned}\hat{F}^{(k)} - \tilde{F}^{(k)} &= \sum_{t=1-n}^{n-1} \hat{\gamma}(t)^{k-1}\hat{\gamma}(n-|t|) + \cdots + \sum_{t=1-n}^{n-1} \hat{\gamma}(t)\hat{\gamma}(n-|t|)^{k-1} \\ &= 2\sum_{t=1}^{n-1} \hat{\gamma}(t)^{k-1}\hat{\gamma}(n-|t|) + \cdots + 2\sum_{t=1}^{n-1} \hat{\gamma}(t)\hat{\gamma}(n-|t|)^{k-1}, \qquad (26)\end{aligned}$$

because $\hat{\gamma}(n) = 0$. Then, setting $M = n^{1/2}$, the first element in (26) is equal to

$$2\sum_{t=1}^{M} \hat{\gamma}(t)^{k-1}\hat{\gamma}(n-t) + 2\sum_{t=M+1}^{n-1} \hat{\gamma}(t)^{k-1}\hat{\gamma}(n-t). \qquad (27)$$

Now, $E\hat{\gamma}(n-t)^2 = O(M^2 n^{-2})$ for $0 < t \leq M$, and using the same methods of the proof of Theorem 1(a), it is easy to see that for $p = 2, 4, 6$,

$$E\hat{\gamma}(t)^p = O(\gamma(t)^p + n^{-p/2}).$$

Hence, we obtain that for $k = 3, 4$

$$\begin{aligned}E\left|\sum_{t=1}^{M} \hat{\gamma}(t)^{k-1}\hat{\gamma}(n-t)\right| &\leq \left(\sum_{t=1}^{M} E\hat{\gamma}(t)^{2(k-1)} \sum_{t=1}^{M} E\hat{\gamma}(n-t)^2\right)^{1/2} \\ &= O\left(\left(\sum_{t=1}^{n}\left\{\gamma(t)^{2(k-1)} + n^{1-2(k-1)}\right\} M^3 n^{-2}\right)^{1/2}\right) \\ &= O\left(M^{3/2} n^{-1}\right) = o(1).\end{aligned}$$

Next,

$$E\left|\sum_{t=M+1}^{n-1} \hat{\gamma}(t)^{k-1}\hat{\gamma}(n-t)\right| \leq \left(\sum_{t=M+1}^{n-1} E\hat{\gamma}(t)^{2(k-1)} \sum_{t=M+1}^{n-1} E\hat{\gamma}(n-t)^2\right)^{1/2}$$

where $\sum_{t=M+1}^{n-1} E\hat{\gamma}(t)^{2(k-1)} = O\left(\sum_{t=M+1}^{n-1}\left\{\gamma(t)^{2(k-1)} + n^{1-k}\right\}\right) = o(1)$ as $n \to \infty$ for $k = 3, 4$, and $\sum_{t=M+1}^{n-1} E\hat{\gamma}(n-t)^2 = O\left(\sum_{t=1}^{n-1}\{\gamma(t)^2 + n^{-1}\}\right) = O\left(1 + \sum_{t=0}^{\infty}|\gamma(t)|\right) = O(1)$.

Hence, both terms on the right hand side of (27) are $o_p(1)$. Similar reasoning can be used to show that the remaining terms in (26) also asymptotically negligible, and conclude that $\hat{F}^{(k)} - \tilde{F}^{(k)} = o_p(1)$.

Appendix C

Write

$$\begin{aligned}\hat{\gamma}_{\hat{x}}(j) - \hat{\gamma}_x(j) &= \frac{1}{n}\sum_{t=1}^{n-|j|} e_t e_{t-|j|} + \frac{1}{n}\sum_{t=1}^{n-|j|} e_t x_{t-|j|} + \frac{1}{n}\sum_{t=1}^{n-|j|} e_{t-|j|} x_t \\ &= A(j) + B(j) + C(j),\end{aligned}$$

say. Thus,

$$\sum_{j=1-n}^{n-1} \hat{\gamma}_{\hat{x}}(j)^4 = \sum_{j=1-n}^{n-1} \left\{ \hat{\gamma}_x(j)^4 + 4\hat{\gamma}_x(j)^3 A(j) + \cdots + A^4(j) + B^4(j) + C^4(j) \right\}.$$

Hence, using from Appendix B that $\sum_{1-n}^{n-1} \hat{\gamma}_x(j)^4 = O_p(1)$ and the Cauchy-Schwartz inequality, we only need to show that

$$\sum_{j=1-n}^{n-1} A^4(j) + \sum_{j=1-n}^{n-1} B^4(j) + \sum_{j=1-n}^{n-1} C^4(j) = o_p(1).$$

First,

$$\begin{aligned}\sum_{j=1-n}^{n-1} A^4(j) &= \frac{1}{n^4}\sum_{j=1-n}^{n-1}\left(\sum_{t=1}^{n-|j|} e_t e_{t-|j|}\right)^4 \\ &\leq 2n^{-3}\left(\sum_{t=1}^{n} e_t^2\right)^4 = O_p(n^{-3}) = o_p(1),\end{aligned}$$

where we have used *Assumption B*.

Second,

$$\begin{aligned}\sum_{1-n}^{n-1} B^4(j) &= \frac{1}{n^4}\sum_{j=1-n}^{n-1}\left[\sum_{t=1}^{n-|j|} e_t x_{t-|j|}\right]^4 \leq \frac{1}{n^4}\sum_{j=1-n}^{n-1}\left[\sum_{t=1}^{n-|j|} e_t^2 \sum_{t=1}^{n-|j|} x_{t-|j|}^2\right]^2 \\ &\leq 2n^{-1}\left[\hat{\gamma}_x(0)\sum_{t=1}^{n} e_t^2\right]^2 = O_p(n^{-1}) = o_p(1),\end{aligned}$$

where we have employed the Cauchy-Schwartz inequality. The analysis of $\sum_{1-n}^{n-1} C^4(j)$ is omitted because it is similar to that of $\sum_{1-n}^{n-1} B^4(j)$.

Tables

Table I. Empirical rejection probabilities at the 0.10, 0.05 and 0.01 nominal levels for (a) the S test and the GS test and for (b) the SK test and the G test. Data follow a Gaussian AR(1) process with parameter ϕ. Sample size is denoted by n.

n		100			500			1000		
ϕ		0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01
(a) Skewness										
-0.5	S	.056	.025	.005	.063	.027	.005	.064	.026	.005
	GS	.093	.047	.011	.103	.052	.010	.101	.051	.009
0	S	.092	.047	.011	.105	.056	.009	.104	.053	.012
	GS	.097	.051	.012	.105	.056	.010	.104	.054	.013
0.5	S	.117	.064	.019	.146	.080	.023	.152	.090	.025
	GS	.095	.048	.012	.100	.052	.011	.105	.054	.012
(b) Skewness and Kurtosis										
-0.5	SK	.051	.032	.015	.079	.043	.016	.082	.044	.010
	G	.065	.039	.014	.090	.047	.014	.095	.047	.011
0	SK	.069	.045	.021	.094	.048	.014	.095	.047	.014
	G	.070	.045	.021	.094	.048	.014	.095	.048	.014
0.5	SK	.080	.050	.023	.120	.071	.025	.138	.082	.026
	G	.063	.040	.015	.084	.045	.014	.094	.053	.014

Table II. Empirical rejection probabilities at the 0.05 nominal levels for the G test. Data follow an AR(1) process with parameter ϕ and various distributions.

n	50			100			200		
ϕ	-0.5	0	0.5	-0.5	0	0.5	-0.5	0	0.5
Lognormal	.915	.996	.918	1.00	1.00	1.00	1.00	1.00	1.00
t_{10}	.104	.179	.108	.159	.294	.163	.243	.421	.234
t_{20}	.057	.097	.058	.092	.133	.073	.125	.201	.108
χ^2_1	.893	.997	.913	1.00	1.00	1.00	1.00	1.00	1.00
χ^2_{10}	.224	.412	.212	.437	.785	.431	.769	.991	.765

References

Bai, J. and S. Ng (2001). Tests for skewness, kurtosis, and normality for time series data. Boston College, preprint.

Bontemps, C. and N. Meddahi (2002). Testing normality: a GMM approach. Université de Montréal, preprint.

Bowman, K.O. and L.R. Shenton (1975). Omnibus test contours for departures from normality based on $\sqrt{b_1}$ and b_2. *Biometrika*, **62**, 243-250.

Brillinger, D.R. (1981). *Time Series: Data Analysis and Theory.* San Francisco: Holden Day.

Davison, A.C. and D.V. Hinkley (1997). *Bootstrap Methods and their Application.* Cambridge: Cambridge University Press.

Epps, T.W. (1987). Testing that a stationary time series is Gaussian. *Annals of Statistics,* **15**, 1683-1698.

Gasser, T. (1975). Goodness-of-fit tests for correlated data. *Biometrika*, **62**, 563-570.

Gleser, L.J. and D.S. Moore (1983). The effect of dependence on chi-squared and empiric distribution tests of fit. *Annals of Statistics*, **11**, 1100-1108.

Jarque, C.M. and A.K. Bera (1987). A test for normality of observations and regression residuals. *International Statistical Review*, **55**, 163-172.

Kilian, L. and U. Demiroglu (2000). Residual-based tests for normality in autoregressions: asymptotic theory and simulation evidence. *Journal of Business and Economic Statistics*, **18**, 40-50.

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

Lomnicki, Z.A. (1961). Tests for departure from normality in the case of linear stochastic processes. *Metrika*, **4**, 37-62.

Lütkepohl, H. (1991). *Introduction to Multiple Time Series Analysis.* New York: Springer Verlag.

Mardia, K.V. (1980). Tests of univariate and multivariate normality. In *Handbook of Statistics: Robust Inference* (ed. P.R. Krishnaiah), vol.1, 279-320. Amsterdam: North-Holland.

Robinson, P.M. (1998). Inference-Without-Smoothing in the Presence of Nonparametric Autocorrelation. *Econometrica*, **66**, 1163-1182.

Robinson, P.M., and C. Velasco (1997). Autocorrelation Robust Inference. In *Handbook of Statistics: Robust Inference* (eds. G. S. Maddala and C.R. Rao), vol. 15, 267-298, Amsterdam: North-Holland.

Rosenblatt, M. (1985). *Stationary Sequences and Random Fields.* Boston: Birkhäuser.

Shapiro, S.S., and M.B. Wilk (1965). An Analysis of Variance Test for Normality (Complete Samples). *Biometrika*, **52**, 591-611.

Shapiro, S.S., M.B. Wilk and H.J. Chen (1968). A Comparative Study of Various Tests for Normality. *Journal of the American Statistical Association*, **63**, 1343-1372.

Zygmund, A. (1977). *Trigonometric Series.* Cambridge: Cambridge University Press.