A simple test for normality for time series

Ignacio N. Lobato and Carlos Velasco
Centro de Investigación Económica
Instituto Tecnológico Autónomo de México
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 paper considers testing for normality for time series data. In econometrics the typical testing procedure employs the Jarque-Bera 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 speci…ed (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. In these cases the Jarque-Bera test is invalid because it does not take the serial correlation into account. In this paper we propose a simple nonparametric modi…cation of the Jarque-Bera test that is robust to the presence of serial correlation of a general form. Besides its simplicity, the remarkable feature of our test is that it does not require the selection of any user-chosen parameter such as a smoothing number or the order of an approximating model.

JEL nos: C12, C14, C22 and C52.

Keywords: Gaussianity, nonparametric, autocorrelation.

1 Introduction

In econometrics testing for normality is customarily performed by means of the skewness-kurtosis test, widely known as the Jarque-Bera (JB) test. The main reason for its widespread use is its straightforward interpretation and implementation. The JB test statistic is the sum of the square of the sample skewness coefficient divided by its asymptotic variance in the white noise case, 6, and the sample excess kurtosis coefficient divided by its asymptotic variance in the white noise case, 24. Implementing the JB test is very simple since it compares the JB statistic against upper critical values of a chi-squared distribution with two degrees of freedom (χ^2_2). This test has been typically applied to the residual series of econometric or time series models (see, for instance, Lütkepohl (1991, Section 4.5) and Kilian and Demiroglu (2000)).

In many empirical studies with time series data the application of the JB test is questionable, though. The reason is that the previous asymptotic variances are correct under the assumption that the model is correctly specified, what implies that the time series under examination is uncorrelated. However, many times either the researcher might specify incorrectly the model or she might not even been interested in modeling the serial correlation. In both cases, when the considered time series is correlated, the asymptotic variances are no longer 6 and 24 but some functions of all the autocovariances. In this situation the JB test is invalid since it does not control asymptotically the type I error. The potential solutions belong to one of the following two strategies. The first is to perform a two step test where the JB is applied after testing that the considered series is uncorrelated. The second solution is to modify the JB test to account for the possibility of serial correlation. The first approach is problematic because there is an obvious pre-test problem in such a sequential procedure and, furthermore, testing for uncorrelatedness is rather challenging itself; see Lobato, Nankervis and Savin (2001).

In this paper we follow the second approach and propose a modification of the JB test statistic that is valid for serially correlated data. The proposed test statistic is a very simple modification of the JB 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 remarkable feature of our procedure is that, as opposed to most of the literature concerning consistent variance estimation (see, for instance, Robinson and

Velasco (1997)), we are able to provide consistent estimators without introducing any user-chosen object such as smoothing numbers, kernel functions or approximating parametric models. 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 meaningless in a testing framework (see the discussion in Robinson (1998, p.1165)).

Our test could be based on 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 we stress an asymptotically equivalent frequency domain estimator since it is relatively easier to handle theoretically. In addition, for conciseness of/in?? exposition, we only analyze the univariate case.

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 analyzes the proposed variance estimators. Section 5 examines the case where the considered series are the residuals of econometric models. Section 6 considers the finite sample performance of the proposed test in a Monte Carlo exercise and Section 7 concludes. The technical material is relegated to the Appendices.

2 Framework

Notation. Let x_t be an ergodic stationary process with mean μ and centered moments denoted by $\mu_k = E(x_t - \mu)^k$ for k natural (hence, μ_2 is the variance), with $\hat{\mu}_k = n^{-1}\sum_{t=1}^{n}(x_t - \bar{x})^k$ being the corresponding sample moments where $\bar{x}$ is the sample mean and n is the sample size. In addition, $\gamma(i)$ denotes the population autocovariance of order i, $\gamma(i) = E[(x_1 - \mu)(x_{1+i} - \mu)]$; and $\hat{\gamma}(i)$ is the corresponding sample autocovariance, $\hat{\gamma}(i) = n^{-1}\sum_{t=1}^{n-|i|}(x_t - \bar{x})(x_{t+|i|} - \bar{x})$. Notice that $\mu_2 = \gamma(0)$. Let $f(\lambda)$ be the spectral density function (sdf, hereafter) of x_t, defined by

$$\gamma(j) = \int_{-\pi}^{\pi} f(\lambda)\cos(j\lambda)d\lambda \qquad j=0,1,2,...$$

and $I(\lambda)$ denotes the periodogram $I(\lambda) = |w(\lambda)|^2$ where $w(\lambda) = (2\pi n)^{-1/2}(\sum_{t=1}^{n} x_t e^{-it\lambda})$: In addition, we denote the q-th order cumulant of $x_1; x_{1+j_1}; \dots; x_{1+j_{q-1}}$ as $\kappa_q(j_1; \dots; j_{q-1})$

and the marginal cumulant of order q as $\kappa_q = \kappa_q(0; \ldots; 0)$. Finally, the q-th order cumulant spectral density is denoted by $f^{(q)}(\lambda)$; where $\lambda \in \Pi^{q-1}$ and where $\Pi = [-\pi; \pi]$, see expression (2.6.2) in Brillinger (1981, p.25).

Null and alternative hypotheses. In principle, the null hypothesis of interest is that the marginal distribution of x_t is normal and the alternative is that it is not. For the independent case, omnibus tests for normality such as the Shapiro-Wilk (1965) that is based on order statistics, or tests based on the distance between the empirical distribution function and the normal cumulative distribution function (cdf) 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 behavior can be very poor (see Shapiro, Wilk and Chen (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 process is invalid, see Gleser and Moore (1983). Employing the bootstrap with these omnibus tests is an area that merits further research.

Instead of testing that the marginal cdf is normal, in econometrics the common procedure is testing that 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 the econometrics practice, and in this paper the considered null hypothesis is

$$H_0 : \mu_3 = 0 \text{ and } \mu_4 = 3\mu_2^2; \tag{1}$$

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

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

During the last two decades many macroeconomists, concerned with the basic properties of the data, have repeatedly considered as different null hypotheses that the skewness and excess kurtosis are zero (see, for instance, Cecchetti, Lam and Mark (1990, p.402)). Hence, we also consider the null hypotheses

$$H_S : \mu_3 = 0 \tag{3}$$

and

$$H_K : \mu_4 = 3\mu_2^2; \tag{4}$$

against, respectively, the alternatives

$$\overline{H}_S : \mu_3 \neq 0 \tag{5}$$

and

$$\overline{H}_K : \mu_4 \neq 3\mu_2^2. \tag{6}$$

The Jarque-Bera test statistic. The null hypothesis (1) is commonly tested using the skewness-kurtosis test statistic (see, for instance, Bowman and Shenton (1975)), in econometrics known as the Jarque-Bera test statistic,

$$JB = \frac{n\hat{\mu}_3^2}{6\hat{\mu}_2^3} + \frac{n(\hat{\mu}_4 - 3\hat{\mu}_2)^2}{24\hat{\mu}_2^4};$$

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 explained in the introduction. In fact, nowadays most econometrics packages report customarily the JB test.

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

$$\sqrt{n}\begin{pmatrix} \hat{\mu}_3 \\ \hat{\mu}_4 - 3\hat{\mu}_2^2 \end{pmatrix} \to_d N\begin{pmatrix} 6\mu_2^3 & 0 \\ 0 & 24\mu_2^4 \end{pmatrix}; \tag{7}$$

where the symbol $\to_d$ denotes convergence in distribution. However, when x_t presents serial correlation, (7) does not hold. In fact, under the null hypothesis (1) and additional weak dependent assumptions which do not require Gaussianity (such as assumptions A that we state below), the following result holds

$$\sqrt{n}\begin{pmatrix} \hat{\mu}_3 \\ \hat{\mu}_4 - 3\hat{\mu}_2^2 \end{pmatrix} \to_d N\begin{pmatrix} 6F^{(3)} & 0 \\ 0 & 24F^{(4)} \end{pmatrix}; \tag{8}$$

where

$$F^{(k)} = \sum_{i=-\infty}^{\infty} \gamma(i)^k; \tag{9}$$

for $k = 3, 4$.

Hence, when the series exhibit serial correlation, the JB test is invalid since asymptotically its rejection probabilities do not coincide with the desired nominal levels under the null hypothesis.

Central Limit Theorem (CLT) for the sample moments. Next, we give a formal justification for the CLT (8). First, introduce

Assumptions A:

A1: The process is strictly stationary of 8-th order and satisfies

$$\sum_{j=0}^{\infty} |\gamma(j)| < \infty; \tag{10}$$

for $q = 2, \ldots, 8$;

$$\sum_{j_1=-\infty}^{\infty} \cdots \sum_{j_{q-1}=-\infty}^{\infty} |\kappa_q(j_1, \ldots, j_{q-1})| < \infty; \tag{11}$$

for $k = 3, 4$;

$$\sum_{j=1}^{\infty} \left[E \left| E\left(x_0^k \mid \mathcal{F}_{-j}\right) - \mu_k \right|^2 \right]^{1/2} < \infty; \tag{12}$$

where $\mathcal{F}_{-j}$ denotes the σ-field generated by x_t, $t \le -j$; and, for $k = 3, 4$;

$$E(x_0^k - \mu_k)^2 + 2 \sum_{j=1}^{\infty} E\left[\left(x_0^k - \mu_k\right)\left(x_j^k - \mu_k\right)\right] > 0. \tag{13}$$

A2:

$$f^{(k)}(\lambda) = 0, \quad \forall \lambda \in \Pi^{k-1}; \text{ for } k = 3, 4, 6 \text{ and } 8;$$

where $\Pi = [-\pi, \pi]$.

Assumption A1 is a weak dependent assumption that guarantees asymptotic normality. It is the natural extension of the condition assumed in Theorem 3 in Rosenblatt (1985) where a CLT for the sample second moment is established. Notice that condition (10) guarantees that all $F^{(k)}$ are well defined since it entails that $\sum |\gamma(j)|^r < \infty$, for all natural r.

Although A1 is the only condition necessary in order to have asymptotic normality, we have imposed A2 because it simplifies technical arguments and because it is necessary to achieve the simple expression for the asymptotic covariance matrix in (8). Note that it implies that all cumulants of orders 3, 4, 6 and 8 are zero, so the left hand side of (13) is only a function of $F^{(k)}$. Gaussian processes not only verify assumption A2 but also that all their higher order cumulant spectral densities, $q > 2$; are zero.

The previous CLT (8) has been established previously in the literature under stronger conditions. Lomnicki (1961) established it for Gaussian autoregressive process while Gasser (1975, p.569) imposed summability conditions on the cumulants of any order. Incidentally, under Gasser's conditions the asymptotic variances in (8) would be di¤erent since the cumulants terms would not vanish (see Lomnicki (1961)).

The CLT (8) can be derived under alternative sets of weak dependent assumptions. For instance, it can be assumed that the considered process is linear with martingale di¤erence innovations (see, for instance, Phillips and Solo (1992)) or that it satis…es some strong mixing or near epoch dependent conditions (see, for instance, De Jong and Davidson (2000)). Since these assumptions do not imply each other (for instance, Assumptions A cover some linear processes with non-martingale di¤erence innovations, see also Brillinger (1980) for a discussion) we have chosen the cumulant-summability condition for its simplicity and natural appeal in the present context.

Finally, we state the following lemma whose proof is a straightforward application of Theorem 2 in Rosenblatt (1985, p.56).

Lemma 1. Under assumptions A1 and A2, the CLT (8) holds.

3 The generalized JB test

In the previous section we have seen that the JB test is invalid when the considered process x_t exhibits serial correlation. In view of (8) two solutions appear naturally. First, to modify the JB test statistic by including consistent estimators of $F^{(3)}$ and $F^{(4)}$ in the denominators of its components. Second, to estimate the unknown asymptotic variances with the bootstrap, that is, to employ the JB test statistic with bootstrap based critical values. In this paper we follow the …rst approach for two reasons. The …rst reason is that in our case the bootstrap does not present a clear theoretical advantage because the asymptotic distribution of the JB statistic is not pivotal. The second reason is that implementing the bootstrap in a time series context is somewhat arbitrary because valid bootstrap procedures require the introduction of a user chosen number what complicates statistical inference (see Horowitz (1999) and

Lobato (2001)).

Before introducing our test statistic, let consider the following estimator of $F^{(k)}$ which is the sample analog of (9)

$$\hat{F}^{(k)} = \sum_{i=1-n}^{n-1} \hat{\gamma}(i)^k . \tag{14}$$

In the next section we consider alternative versions of this estimator that are asymptotically equivalent and study their large sample properties. Lemma 3 in the next section establish the consistency of $\hat{F}^{(k)}$ for $F^{(k)}$ under very general conditions. In particular, in order to guarantee this consistency under H_0, we need to strengthen (11) in assumption A2 to

Assumption A3. The process is strictly stationary of 16-th order and satisfies (11) for $q = 10, 12, 14$ and 16.

Assumption A3 implies that the higher order spectral densities up to the 16-th order are bounded and continuous. We require bounded moments up to the 16-th order because we need to evaluate the variance of the 4-th power of the sample autocovariances.

Then, our proposed test statistic (the Generalized JB) is

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

The G statistic does not require the introduction of any user chosen number and, given the consistency of $\hat{F}^{(k)}$ for $F^{(k)}$, it is straightforward to establish its asymptotic properties.

Lemma 2. Under the null hypothesis (1) and assumptions A1, A2 and A3,

$$G \to_d \chi_2^2 .$$

Under the alternative hypothesis (2) and assumptions A1, A2 and A3, G is consistent, that is, as $n \to \infty$

$$P(G > \chi_2^2(\alpha)) \to 1,$$

where $\chi_2^2(\alpha)$ represents the α upper quantile of the χ_2^2 distribution.

Hence, the proposed test for (1) consists on comparing the G test statistic against upper critical values from a χ_2^2 distribution. For testing the null hypotheses (3) and (4), the obvious tests consist on comparing the components of G versus upper critical values from a χ^2_1 distribution.

4 Consistent variance estimators

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

$$\sum_{j=1-n}^{n-1} w_j \hat{\gamma}(j)^k. \tag{15}$$

In (15) the weights $\{w_j\}$ are usually obtained through a lag window $\left\{w_j = w(\frac{j}{M})\right\}$ such that the weight function $w(\cdot)$ verifies some regularity properties and M is a smoothing number that has to increase with n (usually $w(x) = 0$, $|x| > 1$). In this approach the weights $\{w_j\}$ would provide a nonstochastic dampening on the $\hat{\gamma}(j)^k$, what would lead to the consistency of (15) for (9). 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 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 $\hat{F}^{(k)}$ introduced in equation (14) 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

$$\begin{aligned} F^{(k)} &= \sum_{j=-\infty}^{\infty} \gamma(j)^k = \sum_{j=-\infty}^{\infty} \prod_{i=1}^{k} \left\{ \int_{\Pi} f(v_i) e^{i v_i j} dv_i \right\} \\ &= 2\pi \int_{\Pi^{k-1}} f(v_1 + \cdots + v_{k-1}) \prod_{i=1}^{k-1} \{ f(v_i) dv_i \}. \end{aligned} \tag{16}$$

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

$$\tilde{F}^{(k)} = \frac{(2\pi)^k}{n^{k-1}} \sum_{j_1=1}^{n-1} \cdots \sum_{j_{k-1}=1}^{n-1} I(\lambda_{j_1}) \cdots I(\lambda_{j_{k-1}}) I(\lambda_{j_1} + \cdots + \lambda_{j_{k-1}}), \tag{17}$$

where $\lambda_j = 2\pi j/n$. Notice that $\tilde{F}^{(k)}$ does not depend on the value of the mean μ, since it does not include the periodogram evaluated at the zero frequency. The estimator $\tilde{F}^{(k)}$ can also be written in the time domain by plugging

$$I(\lambda_j) = \frac{1}{2\pi} \sum_{t=1-n}^{n-1} e^{it\lambda_j} \hat{\gamma}(t); \quad j \neq 0; \mathrm{mod}\, n; \tag{18}$$

into equation (17). After some algebra, in Appendix A it is shown that

$$\tilde{F}^{(k)} = \sum_{t=1-n}^{n-1} \hat{\gamma}(t) \{\hat{\gamma}(t) + \hat{\gamma}(n-|t|)\}^{k-1}. \tag{19}$$

Notice that both expressions for $\tilde{F}^{(k)}$ are numerically identical, but in the appendices for technical reasons we stress the frequency domain version (17).

The next lemma whose proof is in Appendix B, state the consistency of $\tilde{F}^{(k)}$ for $F^{(k)}$, and that both estimators, $\hat{F}^{(k)}$ and $\tilde{F}^{(k)}$, are asymptotically equivalent.

Lemma 3. Under assumptions A1, A2 and A3, (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 $\hat{F}^{(k)}$ or $\tilde{F}^{(k)}$ can 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 covariance matrix estimates. 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 (14) and (19) provide a similar intuition. In (14) the powers of the sample autocovariances provide the stochastic dampening factors, whereas in (19) the term $\hat{\gamma}(n-|t|)$ does not contribute asymptotically for small values of $|t|$, for which $\hat{\gamma}(t)$ is large.

In the frequency domain, (16) provides a complementary explanation. Recall that the standard problem is that the relevant asymptotic variance depends on the sdf evaluated at a unique point, the zero frequency. However, in our case (16) shows that the asymptotic variance, $F^{(k)}$, is a convolution of $f(\lambda)$, instead of being a single value of $f(\lambda)$. Therefore (17) just estimates this convolution similarly as the integrated periodogram estimates the population variance (that can be seen as the integral of the sdf).

5 Residual testing

The previous analysis covers the case when raw data are under examination. However, in econometrics the most interesting case is when the test is applied to the residuals of econometric models. Again, two approaches can be used: first, the G test that we propose and second, employing the JB statistic with bootstrap based critical values. The bootstrap has been employed by Kilian and Demiroglou (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 can be justified if the model is correctly specified (notice that in this case, the JB test is also asymptotically valid, though). However, in the absence of the knowledge of the true DGP a parametric bootstrap is invalid, that is, there is no guarantee that the type I error is controlled properly asymptotically. When the model is not correctly specified, valid bootstrap procedures require the introduction of a user chosen number (typically a blocking number) what complicates statistical inference in finite samples, see Lobato (2001).

Next, we introduce a general assumption that validates the use of the G statistic applied to the residuals of general econometric models where the correlation structure is not correctly specified or it is not specified at all. Let call $\hat{x}_t$ to the residuals of the econometric model, with x_t being the true disturbances.

Assumption B. Let x_t have zero mean and satisfy Assumptions A and defining $e_t = \hat{x}_t - x_t$; for $t = 1, \ldots, n$; let e_t satisfy

$$\sum_{t=1} e_t^2 = O_p(1). \tag{20}$$

Assumption B is very general since it covers many interesting cases:

a) Linear regressions with possibly trending stochastic and deterministic regressors (see Grenander's conditions in Grenander and Rosenblatt (1957)) and weakly dependent errors. Here

$$e_t = -(\hat{\beta} - \beta)' Z_t,$$

where Z_t is a p-dimensional sequence of regressors and (20) requires that $(\hat{\beta} - \beta)' Z' Z (\hat{\beta} - \beta) = O_p(1)$; which is commonly satisfied.

b) The case where $\hat{x}_t$ 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 $1 - \beta(w) = 1 - \sum_{j=1}^{p} \beta_j w^j$ has no roots on or inside the unit circle. Notice that in this case the limit process $x_t = y_t - \beta' Z_t$ inherits the weak dependence properties of y_t when Assumption A holds for y_t.

c) Nonlinear dynamic models for which $x_t = x_t(\beta_o)$ satisfies Assumptions A and $\hat{x}_t = x_t(\hat{\beta})$ with $\sqrt{n}(\hat{\beta} - \beta_o) = O_p(1)$ and

$$E \sup_{\beta \in B} \left\| \frac{\partial}{\partial \beta} x_t(\beta) \right\|^2 < \infty;$$

where $B \subset R^p$ is some neighbourhood of β_o (cf. Assumption B of Andrews and Monahan (1992)).

Then, we can establish the following lemma whose proof is in Appendix C.

Lemma 4. Under assumption B, for $k = 3; 4;$

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

This lemma validates the use of the G test applied to the residuals of general econometric models whose correlation structure is ignored or misspecified, since it establishes the asymptotic equivalence between estimators of $F^{(k)}$ based on the true disturbances, x_t, or based on the residuals, $\hat{x}_t$:

6 Finite sample behavior

This section compares briefly the finite sample behavior of the previous tests. We just consider experiments under the null hypothesis. We generate data from an AR(2) process $x_t = \phi x_{t-2} + \varepsilon_t$ where ε_t is independent and identically distributed (IID) N(0,1) and the autoregressive parameter ϕ takes five values: $-0.6; -0.3; 0; 0.3;$ and 0.6: We consider the situation where the researcher specifies incorrectly the DGP. Hence, instead of an AR(2) she fits an AR(1) to the data and then tests for normality in the residuals. This study contrasts with the Monte Carlo study in Lütkepohl and Schneider (1989) who mainly assume that the researcher specifies correctly the DGP. In addition, this study complements Kilian and Demiroglu (2000) who have shown with Monte Carlo and with empirical evidence that the JB test can be completely unreliable with time series data.

We consider the three null hypotheses (1), (3) and (4). For (1) we calculate the JB and the G tests statistics; for (3) we compute the skewness test statistic $n\hat{\mu}_3^2/6\hat{\mu}_2^3$ and the generalized skewness test statistic $n\hat{\mu}_3^2/6\hat{F}^{(3)}$; and, for (4) we compute the kurtosis test statistic $n(\hat{\mu}_4 - 3\hat{\mu}_2)^2/24\hat{\mu}_2^4$ and the generalized kurtosis test statistic $n(\hat{\mu}_4 - 3\hat{\mu}_2)^2/24\hat{F}^{(4)}$. Initially, we compare these test statistics against upper critical values from their asymptotic null distributions. However, notice that since the kurtosis coefficient converges very slowly to its asymptotic normal distribution (see, for instance, Bowman and Stanton (1975, p. 243)), the asymptotic critical values are not ideal in practice, especially if the sample size is small or moderate. Hence, in the second part of the Monte Carlo experiments the previous statistics are compared against some simulated critical values.

In Tables I, II and III we report the empirical rejection probabilities (RP's) for the six tests when asymptotic critical values are employed. We consider three sample sizes (n=100, 500 and 1000), three nominal levels (α=0.10, 0.05 and 0.01), and carried out 10,000 replications of the experiments. One asterisk denotes acceptance of the nominal rejection probability by a 0.05 symmetric asymptotic test, and two asterisks denote acceptance by a 0.01 symmetric asymptotic test.

The main conclusions derived from these tables are the following. First, the skewness test is not reliable since it severely under-rejects when $\phi = -0.6$ and substantially over-rejects when $\phi = 0.6$. On top of that, for these two cases the distortions increase with the sample size as could be expected. On the contrary, for the generalized skewness test the empirical RP's are very close to the nominal levels for all the parameter values and all sample sizes. Second, for n=100, both the kurtosis and the generalized kurtosis tests are completely unreliable even in the white noise case. For the null hypothesis H_K it should be kept in mind the slow convergence of the sample kurtosis to the normal asymptotic distribution even in the white noise case. For n=500 and n=1000, using the white noise case as a benchmark, we see that the generalized kurtosis test slightly under-rejects when $\phi = \pm 0.6$, especially for n=500. On the contrary, for these two cases the kurtosis test over rejects and the magnitude of the the distortion increases with the sample size. Finally, the JB test, which is the sum of the skewness test and the kurtosis test, inherits their characteristics. Notice that for $\phi = -0.6$ there is some compensation between the skewness and kurtosis that makes the distortions of the JB test smaller that those of its components. The G test inherits the slow convergence from the kurtosis test, but, using the white noise case as a reference, it appears to be robust to the presence of serial correlation (at least for n=500 and 1000).

Tables II and III are somewhat worrying since they show that statistical inference is not accurate (even in the white noise case) unless moderate and big samples are available. This fact has been noticed previously in the literature for the white noise case; the proposed solution was to replace the asymptotic critical values by some …nite sample critical values which were calculated either analytically or by simulations (see for instance, Pearson and Hartley (1966), D'Agostino and Pearson (1973) and Jarque and Bera (1987)). The typically employed simulated critical values have been the upper quantiles of the empirical distributions of the relevant tests statistics when data are IID N(0,1) sequences. Unfortunately, tables for these …nite sample critical values are usually incomplete (for instance, Deb and Sefton (1996, p.125)) or wrong (for instance, Jarque and Bera (1987, p.169)). Since we have not been able to …nd a reference with the needed critical values, we estimated them by generating sequences of IID N(0,1) variates using the IMSL DRNNOR random number generator. Table IV reports the estimated 10%, 5% and 1% upper quantiles of the distributions of the skewness, kurtosis and JB statistics for our sample sizes. These numbers are based on 500,000 replications. Table IV is especially important for the kurtosis and the JB tests. Notice that for the JB case, the …gures in Table IV basically coincide with the corresponding …gures in Table 1 of Deb and Sefton (1996).

In Tables V, VI and VII we report the empirical RP's when the simulated critical values of Table IV are used instead of the asymptotic critical values. Again, these numbers are based on 10,000 replications of the experiments. As expected, Tables V, VI and VII indicate that the empirical RP's of the skewness, kurtosis and JB tests are very close to the nominal levels for the white noise case. However, when correlation is present, the three tests present severe distortions similar to those in Tables I, II and III. The main di¤erence is that now the kurtosis test over-rejects for n=100. On the contrary, the generalized skewness, generalized kurtosis and the G tests are essentially correct for all sample sizes (the main exception is the generalized kurtosis test for n=100). These experiments suggest that for the considered sample sizes the …nite sample distributions of the JB test statistic in the white noise case are very close to the …nite sample distributions of the G statistic for the correlated case. This is an interesting empirical observation but there is not an immediate theoretical reason behind it.

7 Conclusions

This paper has proposed a very simple modification of the Jarque Bera normality test that is robust to the presence of serial correlation. The modification consists on studentizing the sample skewness coefficient and the sample excess kurtosis coefficient with consistent estimators of their asymptotic variances. The main feature of the proposed test statistic is its simplicity since these consistent estimators do not require the selection of any smoothing number or further modelization. The test can be carried out both in the time and in the frequency domain but we have stressed the time domain version for simplicity. The finite sample behavior of the test reveals that the asymptotic approximations are not very accurate for small sample sizes even for the white noise case. However, using this case as a benchmark, we have observed that the proposed test appears to be very robust to the presence of serial correlation for moderate sample sizes.

This paper can be extended in different directions. First, residuals in other regression frameworks and the multivariate case could also be analyzed. Second, it would be of interest to study the implementation of the G statistic with bootstrap based critical values. Since the G test statistic is asymptotically pivotal, it can be expected that application of the bootstrap will deliver an asymptotic refinement. In fact, employing bootstrap based critical values appears to improve substantially the accuracy of the JB test in finite samples (see the evidence in in Kilian and Demiroglu, 2000). However, notice that implementing the bootstrap requires the introduction of a user chosen number (typically a blocking number) which complicates statistical inference. This is under current investigation.

8 Tables

n		100			500			1000	
Á	10%	5%	1%	10%	5%	1%	10%	5%	1%
					(a)				
-0.6	.038	.015	.002	.040	.015	.002	.041	.014	.001
-0.3	.080	.039	.010*	.090	.046*	.010*	.085	.043	.008*
0	.089	.044**	.012*	.099*	.051*	.011*	.096*	.050*	.009*
0.3	.093**	.048*	.013**	.104*	.052*	.011*	.103*	.053*	.011*
0.6	.132	.076	.024	.168	.103	.033	.178	.105	.034
					(b)				
-0.6	.089	.047*	.009*	.098*	.048*	.012*	.097*	.049*	.010*
-0.3	.094*	.047*	.012*	.101*	.054*	.012*	.097*	.049*	.011*
0	.093**	.047*	.012*	.100*	.051*	.011*	.097*	.050*	.009*
0.3	.095*	.048*	.012*	.096*	.048*	.009*	.094*	.049*	.009*
0.6	.095*	.047*	.013**	.094*	.049*	.010*	.095*	.047*	.008*

Table I. Empirical RP's at the 10%, 5% and 1% nominal levels for the skewness test (a) and the generalized skewness test (b) when asymptotic critical values are employed. Data follow a Gaussian AR(1) process with parameter Á. Sample size is denoted by n. Number of replications is 10,000. One asterisk denotes acceptance of the nominal rejection probability by a 0.05 symmetric asymptotic test, and two asterisks denote acceptance by a 0.01 symmetric asymptotic test.

n	100			500			1000		
Á	10%	5%	1%	10%	5%	1%	10%	5%	1%
				(a)					
-0.6	.061	.031	.014	.114	.061	.021	.130	.072	.023
-0.3	.054	.027	.012*	.082	.042	.012*	.092**	.045**	.012*
0	.059	.032	.014	.083	.041	.011*	.087	.041	.010*
0.3	.057	.031	.015	.088	.044**	.011*	.091	.047*	.011*
0.6	.060	.033	.015	.114	.060	.019	.129	.065	.020
				(b)					
-0.6	.033	.020	.008*	.072	.038	.012*	.086	.042	.012*
-0.3	.048	.026	.011*	.079	.040	.011*	.088	.043**	.011*
0	.056	.031	.014	.082	.040	.011*	.087	.041	.010*
0.3	.052	.028	.014	.083	.042	.010*	.088	.045**	.011*
0.6	.038	.021	.011*	.074	.038	.012*	.084	.042	.011*

Table II. Empirical RP's at the 10%, 5% and 1% nominal levels for the kurtosis test (a) and the generalized kurtosis test (b) when asymptotic critical values are employed. Same design as Table I.

n		100			500			1000	
Á	10%	5%	1%	10%	5%	1%	10%	5%	1%
					(a)				
-0.6	.044	.028	.012*	.073	.039	.015	.083	.044**	.014
-0.3	.057	.035	.017	.084	.044**	.014	.086	.045**	.013**
0	.064	.041	.020	.084	.045**	.015	.088	.048*	.012*
0.3	.068	.043	.020	.094*	.052*	.015	.097*	.050*	.012*
0.6	.088	.058	.026	.151	.089	.035	.169	.101	.033
					(b)				
-0.6	.056	.032	.012*	.083	.044**	.013**	.088	.047*	.013**
-0.3	.061	.036	.018	.088	.046*	.015	.089	.048*	.014
0	.065	.042	.020	.084	.046*	.015	.088	.048*	.012*
0.3	.066	.042	.018	.087	.047*	.014	.089	.045**	.011*
0.6	.059	.039	.017	.080	.045**	.015	.087	.044**	.012*

Table III. Eempirical RP's at the 10%, 5% and 1% nominal levels for the JB test (a) and the G test (b) when asymptotic critical values are employed. Same design as Table I.

n	(a)			(b)			(c)		
	10%	5%	1%	10%	5%	1%	10%	5%	1%
100	2.53	3.67	6.75	1.97	2.89	7.92	3.66	5.42	12.58
500	2.68	3.83	6.71	2.46	3.57	7.33	4.34	5.86	10.78
1000	2.69	3.83	6.63	2.58	3.71	7.09	4.47	5.93	10.11
1	2.706	3.841	6.635	2.706	3.841	6.635	4.605	5.991	9.210

Table IV. Simulated …nite sample critical values for the skewness (a), kurtosis (b) and JB (c) test statistics for the white noise case.

n	100			500			1000		
Á	10%	5%	1%	10%	5%	1%	10%	5%	1%
					(a)				
-0.6	.045	.017	.002	.039	.016	.002	.039	.014	.001
-0.3	.088	.043	.006	.088	.043	.007	.086	.040	.009*
0	.096*	.048*	.008*	.101*	.055**	.009*	.097*	.047*	.010*
0.3	.106*	.052*	.010*	.111	.061	.011*	.111	.057	.013**
0.6	.140	0.82	.022	.177	.113	.036	.183	.112	.039
					(b)				
-0.6	.106*	.052*	.008*	.097*	.052*	.009*	.097*	.048*	.009*
-0.3	.103*	.052*	.009*	.099*	.052*	.009*	.095*	.046*	.010*
0	.101*	.051*	.010*	.101*	.053*	.009*	.098*	.048*	.010*
0.3	.106*	.053*	.010*	.100*	.053*	.009*	.101*	.052*	.011*
0.6	.103*	.055**	.010*	.100*	.053*	.011*	.101*	.053*	.012*

Table V. Empirical RP's at the 10%, 5% and 1% nominal levels for the skewness test (a) and the generalized skewness test (b) when the simulated critical values from Table IV are employed. Same design as Table I.

n		100			500			1000	
Á	10%	5%	1%	10%	5%	1%	10%	5%	1%
					(a)				
-0.6	.116	.057	.010*	.139	.072	.016	.141	.076	.017
-0.3	.098*	.048*	.010*	.104*	.051*	.010*	.097*	.051*	.009*
0	.100*	.051*	.009*	.097*	.050*	.009*	.099*	.047*	.007**
0.3	.096*	.050*	.009*	.097*	.048*	.010*	.100*	.050*	.007**
0.6	.111	.051*	.010*	.134	.069	.016	.144	.079	.017
					(b)				
-0.6	.062	.030	.006	.089	.044**	.009*	.092**	.045**	.009*
-0.3	.087	.042	.009*	.099*	.048*	.009*	.094*	.050*	.009*
0	.095*	.049*	.009*	.095*	.050*	.009*	.098*	.047*	.008*
0.3	.087	.046*	.009*	.092**	.046*	.010*	.097*	.048*	.007*
0.6	.068	.030	.006	.088	.042	.009*	.095*	.045**	.009*

Table VI. Empirical RP's at the 10%, 5% and 1% nominal levels for the kurtosis test (a) and the generalized kurtosis test (b) when the simulated critical values from Table IV are employed. Same design as Table I.

n	100			500			1000		
Á	10%	5%	1%	10%	5%	1%	10%	5%	1%
(a)									
-0.6	.066	.032	.006	.081	.039	.009*	.087	.044**	.011*
-0.3	.092**	.047*	.008*	.087	.045**	.007**	.093**	.045**	.008*
0	.099*	.050*	.009*	.100*	.051*	.008*	.094*	.044**	.008*
0.3	.102*	.049*	.009*	.109**	.056**	.010*	.108**	.053*	.012*
0.6	.129	.068	.013**	.174	.100	.026	.185	.110	.032
(b)									
-0.6	.091**	.040	.006	.092**	.045**	.008*	.094*	.046*	.008*
-0.3	.097*	.050*	.008*	.096*	.048*	.008*	.096*	.047*	.009*
0	.100*	.051*	.009*	.100*	.052*	.008*	.097*	.047*	.008*
0.3	.100*	.048*	.009*	.101*	.051*	.009*	.100*	.050*	.010*
0.6	.087	.045**	.008*	.095*	.046*	.012*	.099*	.052*	.012*

Table VII. Empirical RP's at the 10%, 5% and 1% nominal levels for the JB test (a) and the G test (b) when the simulated critical values from Table IV are employed. Same design as Table I.

9 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 $\int_{-\pi}^{\pi} D_n(v)dv = 2\pi$ 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

$$\hat F^{(k)} \approx 2\pi\int_{\Pi^{k-1}} I_{x-\bar x}(\lambda_1)\cdots I_{x-\bar x}(\lambda_{k-1})I_{x-\bar x}(\lambda_1+\cdots+\lambda_{k-1})d\lambda_1\cdots d\lambda_{k-1}. \tag{21}$$

Equation (17) is the natural discrete approximation of (21). Notice that (17) does not depend on the sample mean, because it does not use the periodogram evaluated at the zero frequency.

Second, in order to obtain the time domain expression of $\tilde F^{(k)}$ we just plug in (18) into equation (17) to get

$$\begin{aligned}\tilde F^{(k)} &= \frac{1}{n^{k-1}}\sum_{t_1=1-n}^{n-1}\hat\gamma(t_1)\cdots\sum_{t_{k-1}=1-n}^{n-1}\hat\gamma(t_{k-1})\sum_{t_k=1-n}^{n-1}\hat\gamma(t_k) \\ &\quad\times\sum_{j_1=1}\cdots\sum_{j_{k-1}=1}\exp\left\{-i\left[t_1\lambda_{j_1}+\cdots+t_{k-1}\lambda_{j_{k-1}}+t_k(\lambda_{j_1}+\cdots+\lambda_{j_{k-1}})\right]\right\} \\ &= \frac{1}{n^{k-1}}\sum_{t_1=1-n}^{n-1}\hat\gamma(t_1)\cdots\sum_{t_{k-1}=1-n}^{n-1}\hat\gamma(t_{k-1})\sum_{t_k=1-n}^{n-1}\hat\gamma(t_k)\psi_n(\lambda_{t_1}+\lambda_{t_k})\cdots\psi_n(\lambda_{t_{k-1}}+\lambda_{t_k});\end{aligned}$$

where $\psi_n(\lambda) = \sum_{t=1}^{n}\exp(it\lambda)$. Finally, using that $\psi_n(\lambda_j) = 0$ if $\lambda_j = \frac{2\pi j}{n}$; $j \neq 0 \bmod n$ and $\psi_n(0) = n$; we obtain (19).

10 Appendix B

10.1 Proof of Lemma 3(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 $I(\lambda)$, we have that

$$\begin{aligned} & E\left[I(\lambda_{j_1}) I(\lambda_{j_2}) I(\lambda_{j_1} + \lambda_{j_2})\right] \\ = \; & E\left[w(\lambda_{j_1}) w(-\lambda_{j_1}) w(\lambda_{j_2}) w(-\lambda_{j_2}) w(\lambda_{j_1} + \lambda_{j_2}) w(-\lambda_{j_1} - \lambda_{j_2})\right] \\ = \; & \sum_{\nu} \mathrm{cum}(\nu_1) \cdots \mathrm{cum}(\nu_q), \end{aligned} \tag{22}$$

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

$$\{j_1, -j_1, j_2, -j_2, j_1 + j_2, -j_1 - j_2\} \tag{23}$$

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

In order to evaluate the expectation (22), using Assumptions A the only relevant combinations of cumulants only involve second order cumulants, Υ_2 with $q = k = 3$. Hence, writing $f = f^{(2)}$,

$$E[\tilde{F}^{(3)}] = \frac{1}{n^5} \sum_{j_1=1}^{n-1} \sum_{j_2=1}^{n-1} \left\{ \sum_{\Upsilon_2^3} \int_{\Pi^3} \prod_{i=1,2,3} \left[f(\mu_i) \varphi_n(\mu_i + \lambda_{\nu_i(1)}) \varphi_n(\lambda_{\nu_i(2)} - \mu_i) d\mu_i \right] \right\}, \tag{24}$$

where the sum in Υ_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 the all permutations of the coefficients in (23). In fact, following Brillinger (1981, Theorem 4.3.1), the only relevant combinations in the sum in Υ_2^3 are those for which $\nu_i(1) + \nu_i(2) = 0 \bmod n$, $i = 1, 2, 3$. Therefore, using that $|\varphi_n(\mu)| \leq 2 \min\{|\mu|^{-1}, n\}$, see Zygmund (1977, pp.49-51), and the continuity of $f(\lambda)$ implied by (10), we obtain that (24) is

$$
\begin{aligned}
E[\hat F^{(3)}] &= \frac{(2\pi)^3}{n^2}\sum_{j_1=1}^{n-1}\sum_{j_2=1}^{n-1}\left\{\sum_{\kappa_2^3}\int_{\Pi^3}\prod_{i=1,2,3}\left[f(\mu_i)\Phi_n^{(2)}(\mu_i-\lambda_{\nu_i})d\mu_i\right]\right\}+o(1)\\
&= \frac{(2\pi)^3}{n^2}\sum_{j_1=1}^{n-1}\sum_{j_2=1}^{n-1} f(\lambda_{j_1})f(\lambda_{j_2})f(\lambda_{j_1+j_2})+o(1)\\
&= 2\pi\int_{\Pi^2} f(\lambda)f(\mu)f(\lambda+\mu)d\lambda d\mu+o(1)\\
&= F^{(3)}+o(1);\qquad \text{as } n\to\infty;
\end{aligned}
\tag{25}
$$

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

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

$$
Var[\hat F^{(3)}]=cum(\hat F^{(3)};\hat F^{(3)})=\sum_\nu cum(\nu_1)\cdots cum(\nu_q):
$$

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

$$
\begin{array}{cccccc}
j_1 & -j_1 & j_2 & -j_2 & j_1+j_2 & -j_1-j_2\\
j_1' & -j_1' & j_2' & -j_2' & j_1'+j_2' & -j_1'-j_2'.
\end{array}
\tag{26}
$$

Using assumptions A, the relevant combinations ν of cumulants are only κ_2^6; κ_{12}; and $\kappa_{10}\kappa_2$; since all the remaining higher order cumulants are zero by A2. Next, we examine these three terms.

The term in κ_{12} has $\nu=\nu_1$ equal to the whole array (26), and is equal to

$$
\frac{1}{n^{10}}\sum_{j_1=1}^{n-1}\sum_{j_2=1}^{n-1}\sum_{j_1'=1}^{n-1}\sum_{j_2'=1}^{n-1}\left\{\int_{\Pi^{11}}f^{(12)}(\mu_1;\dots;\mu_{11})\prod_{i=1}^{11}\left[\varphi_n(\mu_i+\lambda_{\nu(i)})d\mu_i\right]\varphi_n\Big(\lambda_{\nu(12)}-\sum_{i=1}^{11}\mu_i\Big)\right\};
$$

which is $O(n^{-5}\log^{11}n)$ using the properties of $\varphi_n(\lambda)$ and the boundedness of $f^{(12)}$ by Assumption A3.

Similarly, the terms with a factorization in 2 cumulants ($\nu=\nu_1\cup\nu_2$; that is, $\kappa_{10}\kappa_2$) have a maximal contribution only if there is a restriction between the indices in the rows and columns of (26). Hence, they are at most of order $O(n^{-5}\log^{10}n)$:

The contribution of the terms with only second order cumulants, κ_2^6; is

$$
\frac{1}{n^{10}}\sum_{j_1=1}^{n-1}\sum_{j_2=1}^{n-1}\sum_{j_1'=1}^{n-1}\sum_{j_2'=1}^{n-1}\left\{\sum_{\kappa_2^6}\int_{\Pi^6}\prod_{i=1}^{6}\left(f(\mu_i)\varphi_n(\mu_i+\lambda_{\nu_i(1)})\varphi_n(\lambda_{\nu_i(2)}-\mu_i)d\mu_i\right)\right\}
\tag{27}
$$

where the sum in $\sum_2^6$ is for all the different 6-tuples $\nu = \nu_1 \cup \cdots \cup v_6$ 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 (26) 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 (27) 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

$$\begin{aligned} Var[\tilde{F}^{(3)}] &= \frac{(2\pi)^6}{n^4} 18 \sum_{j_1=1}^{n-1} \sum_{j_2=1}^{n-1} \sum_{j_3=1}^{n-1} f^2(\lambda_{j_1}) f(\lambda_{j_2}) f(\lambda_{j_3}) f(\lambda_{j_1} + \lambda_{j_2}) f(\lambda_{j_1} + \lambda_{j_3}) + o(n^{-1}) \\ &= O(n^{-1}) = o(1) \end{aligned} \tag{28}$$

as $n \to \infty$: Hence, from (25) and (28) we conclude that $\tilde{F}^{(3)} = F^{(3)} + o_p(1)$:

10.2 Proof of Lemma 3(b)

Notice that

$$\begin{aligned} \hat{F}^{(k)} - \tilde{F}^{(k)} &= \sum_{t=1-n}^{n-1} \hat{\gamma}(t)^{k-1} \hat{\gamma}(n - |t|) + \cdots + \sum_{t=1-n}^{n-1} \hat{\gamma}(t) \hat{\gamma}(n - |t|)^{k-1} \\ &= 2 \sum_{t=1}^{n-1} \hat{\gamma}(t)^{k-1} \hat{\gamma}(n - |t|) + \cdots + 2 \sum_{t=1}^{n-1} \hat{\gamma}(t) \hat{\gamma}(n - |t|)^{k-1}; \end{aligned} \tag{29}$$

because $\hat{\gamma}(n) = 0$: Then, setting $M = n^{1/2}$; the first element in (29) is equal to

$$2 \sum_{t=1}^{n-1} \hat{\gamma}(t)^{k-1} \hat{\gamma}(n - t) = 2 \sum_{t=1}^{M} \hat{\gamma}(t)^{k-1} \hat{\gamma}(n - t) + 2 \sum_{t=M+1}^{n-1} \hat{\gamma}(t)^{k-1} \hat{\gamma}(n - t): \tag{30}$$

Now, since $E\hat{\gamma}(n - t)^2 = O(M^2 n^{-2})$ for $0 < t \leq M$; and using Assumptions A, it is easy to see that for $p = 2; \ldots; 6$;

$$E\hat{\gamma}(t)^p = O(\gamma(t)^p + \gamma(t) n^{-1} + n^{1-p}):$$

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

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

Next,

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

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

Then both terms on the right hand side of (30) are $o_p(1)$. Similar reasoning can be used to show that the remaining terms in (29) also asymptotically negligible, and conclude that $\hat{F}^{(k)}-\tilde{F}^{(k)}=o_p(1)$.

11 Appendix C

Using assumption B,

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

say. Thus,

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

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

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

First,

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

where we have used that $\sum_{t=1}^{n} e_t^2 = O_p(1)$ by Assumption B.

Second,

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

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 $\sum_{1-n}^{n-1} B^4(j)$.

12 References

Andrews, D.W. K. and J.C. Monahan (1992), "An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator", Econometrica 60, 953-966.

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. (1980), "The comparison of least squares and third-order periodogram procedures in the estimation of bifrequency", Journal of Time Series Analysis, 2, 95-102.

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

Cecchetti, S.G., P.-S. Lam and N.C. Mark (1990), "Mean reversion in equilibrium asset prices", American Economic Review, 80, 398-418.

D'Agostino, R. and E.S. Pearson (1973) "Tests for departure from normality. Empirical results for the distributions of b_2 and $\sqrt{b_1}$", Biometrika, 60, 613-622.

De Jong, R.M. and J. Davidson (2000), "The functional central limit theorem and weak convergence to stochastic integrals I - Weakly dependent processes", Econometric Theory, 16, 621-642.

Deb, P. and M. Sefton (1996), "The distribution of a Lagrange multiplier test for normality", Economics Letters, 51, 123-130.

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.

Grenander, U. and M. Rosenblatt (1957), Statistical Analysis of Stationary Time Series. Wiley, New York.

Horowitz, J.L. (1999) "The Bootstrap," Handbook of Econometrics, Vol. 5, forthcoming.

Jarque, C.M. and Bera, A.K. (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. (2001), "Testing that a dependent process is uncorrelated", Journal of the American Statistical Association, forthcoming.

Lobato, I.N., Nankervis, J.C., and Savin, N.E. (2001), "Testing for zero autocorrelation in the presence of statistical dependence", Econometric Theory, forthcoming.

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.

Lütkepohl, H. and Schneider, W. (1989) "Testing for nonnormality of autoregressive time series", Computational Statistics Quarterly, 2, 151-168.

Mardia, K.V. (1980) "Tests of univariate and multivariate normality", in Handbook of Statistics: Robust Inference (Vol. 1), eds. P.R. Krishnaiah, Amsterdam, North-Holland, 279-320.

Pearson, E.A. and H.O. Hartley (1966), Biometrika Tables for Statisticians, vol.2 Cambridge University Press.

Phillips, P.C.B. and V. Solo (1992), "Asymptotics for linear processes", Annals of Statistics, 20, 971-1001.

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

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

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

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.