

Econometric Theory, 20, 2004, 671–689. Printed in the United States of America.
DOI: 10.1017/S0266466604204030

A SIMPLE TEST OF NORMALITY FOR TIME SERIES

IGNACIO N. LOBATO

Instituto Tecnológico Autónomo de México (ITAM)

CARLOS VELASCO

*Institució Catalana de Recerca i Estudis Avançats
and*

Universitat Autònoma de Barcelona

This paper considers testing for normality for correlated data. The proposed test procedure employs the skewness-kurtosis test statistic, but studentized by standard error estimators that are consistent under serial dependence of the observations. The standard error estimators are sample versions of the asymptotic quantities that do not incorporate any downweighting, and, hence, no smoothing parameter is needed. Therefore, the main feature of our proposed test is its simplicity, because it does not require the selection of any user-chosen parameter such as a smoothing number or the order of an approximating model.

1. INTRODUCTION

There has been 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, whereas 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 reasons for its widespread use are its

We are very grateful to Don Andrews and two referees for useful comments and suggestions. We are especially thankful to a referee who provided a FORTRAN code. Lobato acknowledges financial support from Asociación Mexicana de Cultura and from Consejo Nacional de Ciencia y Tecnología (CONACYT) under project grant 41893-S. Velasco acknowledges financial support from Spanish Dirección General de Enseñanza Superior, BEC 2001-1270. Address correspondence to: Ignacio N. Lobato, 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; and Carlos Velasco, Dep. d'Economia i d'Història Econòmica, Universitat Autònoma de Barcelona, 08193 Cerdanyola del Valles, Spain; e-mail: Carlos.Velasco@uab.es.

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 because 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 (see, e.g., Lütkepohl, 1991, Sect. 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, on many occasions either the researcher might specify the model incorrectly or might not even be interested in modeling the serial correlation. In both cases, when the considered data are 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 because it does not control asymptotically the type I error.

In this paper we propose to employ the standard test statistic based on the sample skewness and sample kurtosis, but studentized by standard error estimators that are consistent under serial dependence of the observations. The standard error estimators are sample versions of the asymptotic quantities that do not incorporate any downweighting, and, hence, no smoothing parameter is needed. These standard error estimators are consistent even though the asymptotic standard errors involve infinite sums of terms that depend on all autocorrelations. The reason is that in the expression of the asymptotic standard errors, the autocorrelations enter raised to the cubic or fourth powers. Hence, the powers of the sample autocorrelations provide stochastic dampening factors, similar to the nonstochastic dampening factors that appear in the standard nonparametric approach. By contrast, Bai and Ng (2001) and Bontemps and Meddahi (2002) rely on smoothing with kernel methods.

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 of the paper in the Appendixes we stress a frequency domain estimator because it is relatively easier to handle theoretically. In addition, for conciseness of exposition, we only analyze the univariate case.

The plan of the paper 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 Appendixes.

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 $\hat{\mu}_k = n^{-1} \sum_{t=1}^n (x_t - \bar{x})^k$ being the corresponding sample moments where \bar{x} is the sample mean and n is the sample size. In addition, $\gamma(j)$ denotes the population autocovariance of order j , $\gamma(j) = E[(x_1 - \mu)(x_{1+j} - \mu)]$, and $\hat{\gamma}(j)$ is the corresponding sample autocovariance, $\hat{\gamma}(j) = n^{-1} \sum_{t=1}^{n-|j|} (x_t - \bar{x})(x_{t+|j|} - \bar{x})$. Notice that $\mu_2 = \gamma(0)$. Let $f(\lambda)$ be the spectral density function of x_t , defined by

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

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, $\kappa_q(j_1, \dots, j_{q-1})$ denotes the q th-order cumulant of $x_1, x_{1+j_1}, \dots, x_{1+j_{q-1}}$, and the marginal cumulant of order q is $\kappa_q = \kappa_q(0, \dots, 0)$.

Null and alternative hypotheses. 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 test (Shapiro and 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, the Cramér–von Mises, or the Anderson–Darling test have been proposed. A test based on L_2 distance between Gaussian and empirical characteristic functions has been introduced by Epps and Pulley (1983) and developed by Henze and others. For more details see Mardia (1980), Henze (1997), Epps (1999), and references therein. For the independent case, the omnibus tests are consistent, but it has been shown that their finite sample performance can be very poor (see, e.g., Shapiro, Wilk, and Chen, 1968). For the weak dependent case, no such analysis exists because inference with these omnibus test statistics is problematic as a result of the fact that 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) that is based on the characteristic function. However, Epps’s procedure is hard to implement because sensing functions $g(x_t, v)$ have to be selected, a joint spectral density of sensing functions has to be found, a matrix has to be inverted, and a quadratic form has to be minimized to estimate the marginal mean and variance. In addition, there is the disadvantage of having to choose the parameters v that enter $g(\cdot, v)$.

In practice, instead of the previous omnibus tests, the common procedure just 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 we test that the marginal distribution is normal by testing that $\mu_3 = 0$ and $\mu_4 = 3\mu_2^2$. Of course, the derived tests are not consistent because they cannot detect deviations from normality that are not reflected in the third or fourth moments.

The skewness-kurtosis test. This test compares the skewness-kurtosis test statistic

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

against upper critical values of a χ^2_2 distribution (see Bowman and Shenton, 1975). 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 previously. In fact, nowadays most econometrics packages customarily report the *SK* test, which is called the Jarque–Bera test.

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

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

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

$$\sum_{j=0}^{\infty} |\gamma(j)| < \infty, \tag{3}$$

the result (2) is replaced by

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

where

$$F^{(k)} = \sum_{i=-\infty}^{\infty} \gamma(i)^k, \tag{5}$$

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

Hence, when the series exhibits serial correlation, the *SK* test is invalid because the denominators of its components do not estimate consistently the true asymp-

otic variances in (4), implying that asymptotically its rejection probabilities do not coincide with the desired nominal levels under the null hypothesis.

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. One strategy to overcome this problem is to carry out a two-step test where the *SK* procedure is applied after testing that the considered series is uncorrelated. However, this solution is not simple because there is an obvious pretest problem in such a sequential procedure and, furthermore, testing for uncorrelatedness for non-Gaussian series is rather challenging (see Lobato, Nankervis, and Savin, 2002).

Looking at (4) two natural solutions appear. The first one consists of 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 by Gasser (1975, Sect. 6), who suggested truncating the infinite sums that 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 because it does not require the selection of any truncation number. The second solution estimates the unknown asymptotic variances with the bootstrap; that is, it employs the *SK* test statistic with bootstrap-based critical values. Implementing the bootstrap in a time series context is problematic because generally valid bootstrap procedures require the introduction of an arbitrary user-chosen number, typically a block length (see, e.g., Davison and Hinkley, 1997, Ch. 8). Therefore in this paper we follow the first approach. Furthermore, in our case the bootstrap does not present a clear theoretical advantage because the *SK* statistic is not asymptotically pivotal.

Before introducing our test statistic, let us consider the following estimator of $F^{(k)}$, which is the sample analog of (5):

$$\hat{F}^{(k)} = \sum_{j=1-n}^{n-1} \hat{\gamma}(j)^k. \quad (6)$$

In the next section we consider alternative versions of this estimator and study their large sample properties; in particular, Lemma 1 establishes the consistency of $\hat{F}^{(k)}$ for $F^{(k)}$ for Gaussian processes that satisfy condition (3). Then, our proposed test statistic, the generalized *SK* statistic, 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, in view of (4) and Lemma 1 in the next section, the proposed test consists of 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 whenever $\mu_3 \neq 0$ or $\mu_4 \neq 3\mu_2^2$, the G test rejects with probability tending to 1 as n tends to infinity. Notice that the conditions of Gasser (1975) that 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 $q = 2, 3, \dots, 16$,

$$\sum_{j_1=-\infty}^{\infty} \dots \sum_{j_{q-1}=-\infty}^{\infty} |\kappa_q(j_1, \dots, j_{q-1})| < \infty, \tag{7}$$

and, for $k = 3, 4$,

$$\sum_{j=1}^{\infty} [E|(E(x_0 - \mu)^k | \mathfrak{S}_{-j}) - \mu_k|^2]^{1/2} < \infty, \tag{8}$$

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

$$E[(x_0 - \mu)^k - \mu_k]^2 + 2 \sum_{j=1}^{\infty} E([(x_0 - \mu)^k - \mu_k][(x_j - \mu)^k - \mu_k]) > 0. \tag{9}$$

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 (7) implies that condition (3) 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 (9) assures that the asymptotic variances of estimates are positive.

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

THEOREM 1.

- (i) Under the null hypothesis and for Gaussian processes that satisfy condition (3), $G \rightarrow_d \chi_2^2$.
- (ii) Under Assumption A, the test statistic G diverges to infinity whenever $\mu_3 \neq 0$ or $\mu_4 \neq 3\mu_2^2$.

The asymptotic null distribution is straightforward to derive given the consistency of $\hat{F}^{(k)}$ for $F^{(k)}$ that is proved in Lemma 1 in the next section. The proof of (ii) is omitted because it follows easily using that under the alternative hypothesis $\hat{F}^{(k)}$ converges to a bounded positive constant (by (7) and (9)), whereas the numerator of G diverges.

4. CONSISTENT VARIANCE ESTIMATORS

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

$$\sum_{j=1-n}^{n-1} w_j \hat{\gamma}(j)^k \tag{10}$$

In (10) 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 that 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 $\hat{\gamma}(j)^k$ for large j . Because of this dampening, the estimator in (10) is consistent for (5) as it happens in the case $k = 1$, where $f(0)$ is consistently estimated by autocorrelation robust estimators (see, e.g., Robinson and Velasco, 1997).

As mentioned 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 that do not require any smoothing.

Our first estimator $\hat{F}^{(k)}$, introduced in equation (6), 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):

$$\begin{aligned} F^{(k)} &= \sum_{j=-\infty}^{\infty} \gamma(j)^k = \sum_{j=-\infty}^{\infty} \prod_{h=1}^k \left\{ \int_{\Pi} f(v_h) \exp(ijv_h) dv_h \right\} \\ &= 2\pi \int_{\Pi^{k-1}} f(v_1 + \dots + v_{k-1}) \prod_{h=1}^{k-1} \{f(v_h) dv_h\}. \end{aligned} \tag{11}$$

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

$$\tilde{F}^{(k)} = \frac{(2\pi)^k}{n^{k-1}} \sum_{j_1=1}^{n-1} \dots \sum_{j_{k-1}=1}^{n-1} I(\lambda_{j_1}) \dots I(\lambda_{j_{k-1}}) I(\lambda_{j_1} + \dots + \lambda_{j_{k-1}}), \tag{12}$$

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

$$I(\lambda_j) = \frac{1}{2\pi} \sum_{t=1-n}^{n-1} \exp(it\lambda_j) \hat{\gamma}(t), \quad j \neq 0, \text{ mod } n, \tag{13}$$

into equation (12). After some algebra, in Appendix A it is shown that

$$\tilde{F}^{(k)} = \sum_{t=1-n}^{n-1} \hat{\gamma}(t) \{ \hat{\gamma}(t) + \hat{\gamma}(n - |t|) \}^{k-1}. \quad (14)$$

Notice that both expressions for $\tilde{F}^{(k)}$ are numerically identical, but in the Appendixes, for technical reasons, we stress the frequency domain version (12). Expression (12) guarantees that $\tilde{F}^{(k)}$ is positive in finite samples.

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

LEMMA 1. *Under the null hypothesis, for Gaussian time series that satisfy condition (3),*

- (i) $\tilde{F}^{(k)} = F^{(k)} + o_p(1)$ and
- (ii) $\hat{F}^{(k)} - \tilde{F}^{(k)} = o_p(1)$ for $k = 3, 4$.

At first look, consistency of $\hat{F}^{(k)}$ and $\tilde{F}^{(k)}$ could be surprising because no smoothing parameter has been introduced. Robinson (1998) analyzes a special regression model where smoothing is not necessary for establishing consistency of asymptotic covariance matrix estimators. The reason is that the specific form of the covariance matrix that he considers (see his equation (1.2)) allows for a stochastic dampening of some sample autocovariances by other sample autocovariances. The time domain versions (6) and (14) provide a similar intuition where the powers of the sample autocovariances provide the stochastic dampening factors.

In the frequency domain, (11) 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 (11) 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 mentioned 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 mentioned previously, bootstrap procedures valid for 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 denotes the residuals of the regression or time series model, and x_t denotes the true disturbances.

Assumption B. Let the Gaussian process x_t satisfy (3) and let $e_t = x_t - \hat{x}_t$ satisfy

$$\sum_{t=1}^n e_t^2 = O_p(1) \quad \text{and} \quad \sum_{t=1}^n e_t^4 = o_p(n^{-1/4}). \tag{15}$$

The first condition in (15) guarantees the consistency of the estimates of $F^{(k)}$ based on residuals, whereas the second guarantees that the residual SK test has the same asymptotic distribution as the original SK test. Assumption B is very general and covers many interesting cases such as linear regressions with possible trending stochastic and deterministic regressors that 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 (15) 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 $Z_t = (y_{t-1}, \dots, y_{t-p})'$, 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 lemma, which shows that the use of residuals does not affect the consistent studentization that we propose in this paper.

LEMMA 2. *Under the null hypothesis 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 lemma and Hölder’s inequality, it is straightforward to prove the next theorem, 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 whenever $\mu_3 \neq 0$ or $\mu_4 \neq 3\mu_2^2$ the G test rejects with probability tending to 1 as n tends to infinity.

THEOREM 2. *Let \hat{G} be the test statistic G calculated from residuals \hat{x}_t .*

- (i) *Under the null hypothesis and Assumption B, $\hat{G} \rightarrow_d \chi_2^2$.*
- (ii) *If $\mu_3 \neq 0$ or $\mu_4 \neq 3\mu_2^2$ and Assumptions A and B hold, then the test statistic \hat{G} diverges to infinity.*

6. FINITE SAMPLE PERFORMANCE

This section compares briefly the finite sample behavior of the previous tests with the Epps (1987) test. 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 eight values: $-0.9, -0.5, 0, 0.5, 0.6, 0.7, 0.8,$ and 0.9 . We report the results for a detailed grid of positive values of ϕ because positive autocorrelation is particularly relevant for many empirical applications.

Along with the null hypothesis, 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 $S = n\hat{\mu}_3^2/6\hat{\mu}_2^3$ and the generalized skewness test statistic $GS = n\hat{\mu}_3^2/6\hat{F}^{(3)}$ and compare them with upper critical values from a χ_2^2 . 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 (see, e.g., Bowman and Shenton, 1975, p. 243). In Tables 1A and 1B we report the empirical rejection probabilities for the tests for three sample sizes, $n = 100, 500,$ and $1,000$, and three nominal levels, $\alpha = 0.10, 0.05,$ and 0.01 . In these experiments 5,000 replications are carried out.

The main conclusions derived from Table 1A are the following. For the case of testing symmetry, the S test is not reliable since it severely underrejects for the cases when $\phi < 0$ and substantially overrejects for the cases when $\phi > 0$. 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 underrejection of the S test, whereas when ϕ is positive the opposite effect occurs. The most interesting evidence is the magnitude of these distortions, which are very large for negative values of ϕ and all sample sizes, whereas for positive ϕ 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 (the only exception is when $n = 100$ and $\phi = 0.9$).

Table 1B reports the results for testing normality for the three tests, SK , G , and Epps test. The SK test, which is the sum of the skewness test and the kur-

TABLE 1. Empirical rejection probabilities for 3 sample sizes and 3 nominal levels

n	ϕ	100			500			1,000		
		0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01
A. <i>S</i> and <i>GS</i> tests										
-0.9	<i>S</i>	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	<i>GS</i>	0.083	0.038	0.008	0.097	0.047	0.012	0.091	0.045	0.010
-0.5	<i>S</i>	0.056	0.025	0.005	0.063	0.027	0.005	0.064	0.026	0.005
	<i>GS</i>	0.093	0.047	0.011	0.103	0.052	0.010	0.101	0.051	0.009
0	<i>S</i>	0.092	0.047	0.011	0.105	0.056	0.009	0.104	0.053	0.012
	<i>GS</i>	0.097	0.051	0.012	0.105	0.056	0.010	0.104	0.054	0.013
0.5	<i>S</i>	0.117	0.064	0.019	0.146	0.080	0.023	0.152	0.090	0.025
	<i>GS</i>	0.095	0.048	0.012	0.100	0.052	0.011	0.105	0.054	0.012
0.6	<i>S</i>	0.140	0.081	0.026	0.175	0.107	0.037	0.181	0.109	0.039
	<i>GS</i>	0.094	0.046	0.011	0.097	0.050	0.009	0.098	0.052	0.009
0.7	<i>S</i>	0.176	0.109	0.039	0.233	0.158	0.061	0.231	0.157	0.069
	<i>GS</i>	0.091	0.046	0.010	0.099	0.049	0.011	0.097	0.049	0.009
0.8	<i>S</i>	0.228	0.153	0.065	0.315	0.231	0.116	0.321	0.238	0.123
	<i>GS</i>	0.089	0.045	0.009	0.098	0.049	0.011	0.092	0.049	0.012
0.9	<i>S</i>	0.278	0.196	0.090	0.442	0.361	0.238	0.467	0.389	0.261
	<i>GS</i>	0.067	0.029	0.005	0.090	0.047	0.013	0.096	0.049	0.010
B. <i>SK</i> , <i>G</i> , and <i>Epps</i> tests										
-0.9	<i>SK</i>	0.078	0.034	0.009	0.229	0.157	0.072	0.261	0.195	0.106
	<i>G</i>	0.031	0.014	0.004	0.062	0.038	0.013	0.067	0.038	0.012
	<i>E</i>	0.172	0.109	0.046	0.121	0.072	0.018	0.126	0.068	0.015
-0.5	<i>SK</i>	0.051	0.032	0.015	0.079	0.043	0.016	0.082	0.044	0.010
	<i>G</i>	0.065	0.039	0.014	0.090	0.047	0.014	0.095	0.047	0.011
	<i>E</i>	0.118	0.063	0.020	0.109	0.059	0.013	0.104	0.055	0.011
0	<i>SK</i>	0.069	0.045	0.021	0.094	0.048	0.014	0.095	0.047	0.014
	<i>G</i>	0.070	0.045	0.021	0.094	0.048	0.014	0.095	0.048	0.014
	<i>E</i>	0.123	0.067	0.020	0.099	0.055	0.013	0.106	0.055	0.010
0.5	<i>SK</i>	0.080	0.050	0.023	0.120	0.071	0.025	0.138	0.082	0.026
	<i>G</i>	0.063	0.040	0.015	0.084	0.045	0.014	0.094	0.053	0.014
	<i>E</i>	0.130	0.069	0.021	0.113	0.061	0.017	0.101	0.054	0.014
0.6	<i>SK</i>	0.093	0.058	0.025	0.157	0.095	0.036	0.170	0.104	0.039
	<i>G</i>	0.056	0.035	0.015	0.079	0.045	0.017	0.088	0.047	0.013
	<i>E</i>	0.139	0.079	0.022	0.128	0.070	0.017	0.110	0.060	0.014
0.7	<i>SK</i>	0.114	0.075	0.036	0.221	0.142	0.063	0.238	0.158	0.067
	<i>G</i>	0.054	0.033	0.012	0.076	0.043	0.017	0.081	0.046	0.015
	<i>E</i>	0.158	0.087	0.026	0.134	0.078	0.019	0.121	0.064	0.011
0.8	<i>SK</i>	0.168	0.103	0.047	0.329	0.236	0.114	0.356	0.266	0.132
	<i>G</i>	0.045	0.026	0.009	0.064	0.042	0.018	0.075	0.041	0.016
	<i>E</i>	0.185	0.115	0.036	0.141	0.081	0.024	0.120	0.060	0.016
0.9	<i>SK</i>	0.267	0.154	0.062	0.549	0.440	0.265	0.585	0.489	0.323
	<i>G</i>	0.027	0.015	0.006	0.056	0.036	0.013	0.067	0.043	0.013
	<i>E</i>	0.236	0.155	0.064	0.180	0.110	0.033	0.149	0.084	0.026

Note: Data follow a Gaussian AR(1) process with parameter ϕ . Sample size is denoted by n .

tosis test, inherits their characteristics. Notice that for the cases where $\phi < 0$, there is a fair amount of compensation between the skewness and kurtosis, making the distortions of the *SK* test much smaller than 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 moderate serial correlation. When $|\phi| = 0.9$, the *G* test is severely affected by its kurtosis component. In fact, even for $n = 1,000$ the *G* test appears to be very conservative. For the cases $\phi = 0.7$ and $\phi = 0.8$, a similar pattern can be observed. Similar to the *G* test, the Epps test is also insensitive to moderate serial correlation. However, for the case $\phi = -0.9$, and also for the most interesting cases where $\phi \geq 0.7$, the Epps test appears to be too liberal.

We also conducted power experiments for data generated by the previous AR(1) model for six different distributions: standard log-normal, student's t with 10 degrees of freedom, χ_1^2 , χ_{10}^2 , beta with parameters (1,1), and beta with parameters (2,1). Although distributions with bounded support are not that popular in econometrics, it is well known that in the independent and identically distributed (i.i.d) setting the *SK* test performs very poorly against such alternatives. Hence, it is of interest to examine the performance of the *G* test in these difficult cases. Table 2 reports the power results for the *G* and the Epps tests for three sample sizes, $n = 100, 500$, and $1,000$, respectively, and for a 5% nominal level. In these experiments 2,000 replications are carried out. The main conclusions from these tables are the following. For both tests it appears that the sign of the autocorrelation has little relevance in terms of power (although generally the empirical power is slightly greater for positive ϕ). Using the white noise as the reference case, higher values for $|\phi|$ lead to a decrease in the empirical power that in some cases is very exacerbated. The empirical rejection probabilities for the *G* test are particularly high for heavily skewed distributions such as the lognormal or the χ_1^2 . For these cases the *G* test is clearly preferable to the Epps test. When the distribution is symmetric or slightly skewed, both tests are comparable. For the t_{10} and the χ_{10}^2 distributions, the *G* test presents higher empirical power, especially for a moderate degree of serial correlation. For these cases and when $|\phi| = 0.9$, the tests present very low empirical power even for $n = 1,000$. Notice that when $n = 1,000$ and $\phi = 0.7$ or 0.8 , for the χ_{10}^2 case, the empirical powers of both the Epps test and, especially, the *G* test are moderately high, but the power deteriorates suddenly for $\phi = 0.9$. For the beta distributions, both tests (and especially the *G* test) appear to be very sensitive to a high degree of serial correlation. In fact, when $|\phi| = 0.9$, the power of both tests is very low even when $n = 1,000$. Here again, there is a sudden decrease in the empirical power when ϕ increases from 0.6 to 0.7 for $n = 500$ and when ϕ increases from 0.7 to 0.8 for $n = 1,000$.

We end with a suggestion on further research. In this section we have seen that for small sample sizes, because of 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 perfor-

mance is by using the bootstrap. Because 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.

TABLE 2. Empirical rejection probabilities at the 0.05 nominal levels for the G and Epps (E) tests for 3 sample sizes

ϕ		Log N	t_{10}	χ^2_1	χ^2_{10}	Beta(1,1)	Beta(2,1)
$n = 100$							
-0.9	G	0.291	0.045	0.175	0.048	0.006	0.035
	E	0.041	0.115	0.052	0.115	0.120	0.128
-0.5	G	0.999	0.187	0.998	0.437	0.002	0.121
	E	0.673	0.043	0.843	0.200	0.557	0.484
0	G	1	0.299	1	0.798	0.511	0.740
	E	0.971	0.079	0.996	0.537	0.993	0.978
0.5	G	1	0.177	1	0.435	0.001	0.114
	E	0.865	0.059	0.953	0.215	0.532	0.544
0.6	G	0.992	0.122	0.985	0.312	0.001	0.064
	E	0.612	0.057	0.782	0.121	0.257	0.270
0.7	G	0.936	0.080	0.899	0.188	0.004	0.040
	E	0.318	0.062	0.392	0.060	0.099	0.080
0.8	G	0.742	0.045	0.591	0.104	0.007	0.034
	E	0.146	0.083	0.120	0.017	0.025	0.008
0.9	G	0.371	0.042	0.187	0.055	0.013	0.028
	E	0.111	0.144	0.054	0.006	0.010	0.003
$n = 500$							
-0.9	G	0.959	0.080	0.734	0.126	0.023	0.078
	E	0.400	0.064	0.226	0.075	0.101	0.097
-0.5	G	1	0.484	1	0.995	0.971	0.995
	E	1	0.132	1	0.756	0.992	0.987
0	G	1	0.773	1	1	1	1
	E	1	0.320	1	0.996	1	1
0.5	G	1	0.471	1	0.998	0.964	1
	E	1	0.139	1	0.857	0.992	0.998
0.6	G	1	0.323	1	0.960	0.465	0.914
	E	0.999	0.113	1	0.701	0.785	0.901
0.7	G	1	0.194	1	0.773	0.059	0.430
	E	0.999	0.085	1	0.473	0.326	0.551
0.8	G	1	0.105	1	0.403	0.037	0.116
	E	0.980	0.077	0.950	0.249	0.105	0.217
0.9	G	0.947	0.075	0.737	0.133	0.030	0.062
	E	0.572	0.090	0.382	0.059	0.032	0.027

continued

TABLE 2. Continued

ϕ		Log N	t_{10}	χ^2_1	χ^2_{10}	Beta(1,1)	Beta(2,1)
$n = 1,000$							
-0.9	<i>G</i>	1	0.091	0.978	0.245	0.058	0.116
	<i>E</i>	0.881	0.064	0.484	0.086	0.097	0.101
-0.5	<i>G</i>	1	0.700	1	1	1	1
	<i>E</i>	1	0.243	1	0.971	1	1
0	<i>G</i>	1	0.953	1	1	1	1
	<i>E</i>	1	0.583	1	1	1	1
0.5	<i>G</i>	1	0.710	1	1	1	1
	<i>E</i>	1	0.250	1	0.991	1	1
0.6	<i>G</i>	1	0.489	1	1	0.959	1
	<i>E</i>	1	0.176	1	0.946	0.974	0.997
0.7	<i>G</i>	1	0.298	1	0.980	0.352	0.907
	<i>E</i>	1	0.113	1	0.766	0.570	0.845
0.8	<i>G</i>	1	0.148	1	0.688	0.068	0.331
	<i>E</i>	1	0.076	1	0.439	0.164	0.403
0.9	<i>G</i>	1	0.094	0.980	0.219	0.061	0.111
	<i>E</i>	0.962	0.086	0.724	0.168	0.075	0.118

Note: Data follow an AR(1) process with parameter ϕ .

REFERENCES

Bai, J. & S. Ng (2001) Tests for Skewness, Kurtosis, and Normality for Time Series Data. Preprint, Boston College.

Bontemps, C. & N. Meddahi (2002) Testing Normality: A GMM Approach. Preprint, Université de Montréal.

Bowman, K.O. & 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*. Holden Day.

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

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

Epps, T.W. (1999) Limiting behavior of the integrated characteristic function test for normality under Gram-Charlier alternatives. *Statistics and Probability Letters* 42, 175–184.

Epps, T.W. & L.B. Pulley (1983) A test for normality based on the empirical characteristic function. *Biometrika* 70, 723–726.

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

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

Henze, N. (1997) A new approach to the BEHP tests for multivariate normality. *Journal of Multivariate Analysis* 62, 1–23.

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

Kilian, L. & 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, & 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*. Springer Verlag.

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

Robinson, P.M. (1998) Inference-without-smoothing in the presence of nonparametric autocorrelation. *Econometrica* 66, 1163–1182.

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

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

Shapiro, S.S. & M.B. Wilk, (1965) An analysis of variance test for normality (complete samples). *Biometrika* 52, 591–611.

Shapiro, S.S., M.B. Wilk, & 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 University Press.

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:

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

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

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

Equation (12) is the natural discrete approximation of (A.1).

Second, to obtain the time domain expression of $\tilde{F}^{(k)}$ we just plug (13) into equation (12) to get

$$\begin{aligned} \tilde{F}^{(k)} &= \frac{1}{n^{k-1}} \sum_{t_1=1-n}^{n-1} \hat{\gamma}(t_1) \dots \sum_{t_{k-1}=1-n}^{n-1} \hat{\gamma}(t_{k-1}) \sum_{t_k=1-n}^{n-1} \hat{\gamma}(t_k) \\ &\quad \times \sum_{j_1=1}^n \dots \sum_{j_{k-1}=1}^n \exp\{i(t_1\lambda_{j_1} + \dots + t_{k-1}\lambda_{j_{k-1}} + t_k(\lambda_{j_1} + \dots + \lambda_{j_{k-1}}))\} \\ &= \frac{1}{n^{k-1}} \sum_{t_1=1-n}^{n-1} \hat{\gamma}(t_1) \dots \sum_{t_{k-1}=1-n}^{n-1} \hat{\gamma}(t_{k-1}) \sum_{t_k=1-n}^{n-1} \hat{\gamma}(t_k) \phi_n(\lambda_{t_1} + \lambda_{t_k}) \dots \phi_n(\lambda_{t_{k-1}} + \lambda_{t_k}), \end{aligned}$$

where $\phi_n(\lambda) = \sum_{t=1}^n \exp(it\lambda)$. Finally, using that $\phi_n(\lambda_j) = 0$ if $\lambda_j = 2\pi j/n, j \neq 0 \pmod n$, and $\phi_n(0) = n$, and denoting the indicator function by 1, we obtain for $j = 1, \dots, k - 1$,

$$\begin{aligned} \frac{1}{n} \sum_{t_j=1-n}^{n-1} \hat{\gamma}(t_j) \phi_n(\lambda_{t_j} + \lambda_k) &= \hat{\gamma}(-t_k) + \hat{\gamma}(n - t_k) 1_{\{t_k > 0\}} + \hat{\gamma}(-n - t_k) 1_{\{t_k < 0\}} \\ &= \hat{\gamma}(t_k) + \hat{\gamma}(n - |t_k|), \end{aligned}$$

where we have used that $\hat{\gamma}$ is even. Then (14) follows immediately.

APPENDIX B

Proof of Lemma 1(i). We just report the analysis for $\tilde{F}^{(3)}$ because the analysis 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 $n \rightarrow \infty$.

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

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

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

$$\begin{aligned} E[I(\lambda_{j_1})I(\lambda_{j_2})I(\lambda_{j_1} + \lambda_{j_2})] &= E[w(\lambda_{j_1})w(-\lambda_{j_1})w(\lambda_{j_2})w(-\lambda_{j_2})w(\lambda_{j_1} + \lambda_{j_2})w(-\lambda_{j_1} - \lambda_{j_2})] \tag{B.1} \\ &= \sum_{\nu} \text{cum}(\nu_1) \cdots \text{cum}(\nu_q), \end{aligned}$$

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

$$\{j_1, -j_1, j_2, -j_2, j_1 + j_2, -j_1 - j_2\} \tag{B.2}$$

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

To evaluate the expectation (B.1), by Gaussianity the only cumulants different from zero are second-order cumulants, κ_2 , with $q = 3$. Hence

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

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 (B.2). In fact, following Brillinger (1981, Theorem 4.3.1), the only relevant combinations in the sum in κ_2^3 are those for which $\nu_i(1) + \nu_i(2) = 0 \pmod n, i = 1, 2, 3$. 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 (3), we obtain that (B.3) is

$$\begin{aligned} E[\tilde{F}^{(3)}] &= \frac{(2\pi)^3}{n^2} \sum_{j_1=1}^{n-1} \sum_{j_2=1}^{n-1} \left\{ \sum_{\kappa_3^3} \int_{\Pi^3} \prod_{i=1,2,3} [f(\mu_i) \Phi_n^{(2)}(\mu_i - \lambda_{\nu_i}) d\mu_i] \right\} + o(1) \\ &= \frac{(2\pi)^3}{n^2} \sum_{j_1=1}^{n-1} \sum_{j_2=1}^{n-1} f(\lambda_{j_1}) f(\lambda_{j_2}) f(\lambda_{j_1+j_2}) + o(1) \\ &= 2\pi \int_{\Pi^2} f(\lambda) f(\lambda) f(\lambda + \mu) d\lambda d\mu + o(1) \\ &= F^{(3)} + o(1), \quad \text{as } n \rightarrow \infty, \end{aligned} \tag{B.4}$$

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 $\tilde{F}^{(3)}$,

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

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:

$$\begin{array}{cccccc} j_1 & -j_1 & j_2 & -j_2 & j_1 + j_2 & -j_1 - j_2, \\ j'_1 & -j'_1 & j'_2 & -j'_2 & j'_1 + j'_2 & -j'_1 - j'_2. \end{array} \tag{B.5}$$

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

$$\begin{aligned} \text{Var}[\tilde{F}^{(3)}] &= \frac{1}{n^{10}} \sum_{j_1=1}^{n-1} \sum_{j_2=1}^{n-1} \sum_{j'_1=1}^{n-1} \sum_{j'_2=1}^{n-1} \left\{ \sum_{\kappa_2^6} \int_{\Pi^6} \prod_{i=1}^6 \{f(\mu_i) \phi_n(\mu_i + \lambda_{\nu_i(1)}) \right. \\ &\quad \left. \times \phi_n(\lambda_{\nu_i(2)} - \mu_i) d\mu_i \right\} \end{aligned} \tag{B.6}$$

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 (B.5) 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 (B.6) characterized by a restriction $\nu_i(1) + \nu_i(2) = 0 \pmod n$, for just one $i \in \{1, \dots, 6\}$ (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

$$\begin{aligned} \text{Var}[\tilde{F}^{(3)}] &= \frac{(2\pi)^6}{n^4} 18 \sum_{j_1=1}^{n-1} \sum_{j_2=1}^{n-1} \sum_{j_3=1}^{n-1} f^2(\lambda_{j_1}) f(\lambda_{j_2}) f(\lambda_{j_3}) f(\lambda_{j_1} + \lambda_{j_2}) f(\lambda_{j_1} + \lambda_{j_3}) + o(n^{-1}) \\ &= O(n^{-1}) = o(1) \end{aligned} \tag{B.7}$$

as $n \rightarrow \infty$. Hence, from (B.4) and (B.7) we conclude that $\tilde{F}^{(3)} = F^{(3)} + o_p(1)$. ■

Proof of Lemma 1(ii). Notice that

$$\begin{aligned} \hat{F}^{(k)} - \tilde{F}^{(k)} &= \sum_{t=1}^{n-1} \hat{\gamma}(t)^{k-1} \hat{\gamma}(n-|t|) + \dots + \sum_{t=1}^{n-1} \hat{\gamma}(t) \hat{\gamma}(n-|t|)^{k-1} \\ &= 2 \sum_{t=1}^{n-1} \hat{\gamma}(t)^{k-1} \hat{\gamma}(n-|t|) + \dots + 2 \sum_{t=1}^{n-1} \hat{\gamma}(t) \hat{\gamma}(n-|t|)^{k-1}, \end{aligned} \tag{B.8}$$

because $\hat{\gamma}(n) = 0$. Then, setting $M = n^{1/2}$, the first element in (B.8) is equal to

$$2 \sum_{t=1}^M \hat{\gamma}(t)^{k-1} \hat{\gamma}(n-t) + 2 \sum_{t=M+1}^{n-1} \hat{\gamma}(t)^{k-1} \hat{\gamma}(n-t). \tag{B.9}$$

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 Lemma 1(i), 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 $k = 3, 4$,

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

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

Hence, both terms on the right-hand side of (B.9) are $o_p(1)$. Similar reasoning can be used to show that the remaining terms in (B.8) are also asymptotically negligible and conclude that $\hat{F}^{(k)} - \tilde{F}^{(k)} = o_p(1)$. ■

APPENDIX C

Proof of Lemma 2. 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,

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

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 $\sum_{1-n}^{n-1} B^4(j)$. ■