Stat Methods Appl (2010) 19:193–215
DOI 10.1007/s10260-009-0124-1

ORIGINAL ARTICLE

Behaviour of skewness, kurtosis and normality tests in long memory data

Mohamed Boutahar

Accepted: 15 June 2009 / Published online: 7 July 2009

Abstract We establish the limiting distributions for empirical estimators of the coefficient of skewness, kurtosis, and the Jarque–Bera normality test statistic for long memory linear processes. We show that these estimators, contrary to the case of short memory, are neither $\sqrt{n}$-consistent nor asymptotically normal. The normalizations needed to obtain the limiting distributions depend on the long memory parameter d. A direct consequence is that if data are long memory then testing normality with the Jarque–Bera test by using the chi-squared critical values is not valid. Therefore, statistical inference based on skewness, kurtosis, and the Jarque–Bera normality test, needs a rescaling of the corresponding statistics and computing new critical values of their nonstandard limiting distributions.

Keywords Hermite polynomials · Jarque–Bera normality test · Kurtosis · Long memory data · Skewness

1 Introduction

Let (x_t) be a covariance stationary process with mean $E(x_t) = \mu$ and autocovariance function $\gamma_x(k) = E(x_{t+k} - \mu)(x_t - \mu)$, $\sigma^2 = \gamma_x(0)$. We say that (x_t) is short or long memory according whether the sum $\sum_{k\in Z} |\gamma_x(k)|$ is finite or infinite. The sum is infinite if the process (x_t) satisfies one of the following:

- There exist $d \in (0, 1/2)$ and a constant $c_1 > 0$ such that

$$\gamma_x(k)\, k^{-2d+1} \to c_1 \quad \text{as } k \to \infty, \tag{1}$$

M. Boutahar (✉)
GREQAM, University of Méditerranée, Marseille, France
e-mail: mohammed.boutahar@univmed.fr

or

- There exist $d \in (0, 1/2)$ and a constant $c_2 > 0$ such that

$$|\lambda|^{2d} f_x(\lambda) \to c_2 \quad \text{as } \lambda \to 0, \tag{2}$$

  where $f_x(\lambda)$ is the spectral density function of (x_t), i.e.

$$f_x(\lambda) = \sum_{k \in Z} \gamma_x(k) e^{-ik\lambda}/2\pi.$$

For short memory processes, the covariance decays quickly (exponential decay), and the spectral density is at least bounded. The stationary and invertible *ARMA* is a short memory process. For long memory processes, the covariance decays slowly (hyperbolic decay) and the spectral density is unbounded at frequency 0. A well known long memory process is the *ARFIMA* (p, d, q) defined by

$$\phi(L)(1-L)^d x_t = \theta(L) u_t,$$

where $\phi(L) = 1 - \phi_1 L - \cdots - \phi_p L^p, \theta(L) = 1 + \theta_1 L + \cdots + \theta_q L^q$, $d \in R$ is the memory parameter, L is the backward shift operator $Lx_t = x_{t-1}$, u_t is a white noise with mean 0 and variance σ^2, $(1-L)^d$ is the fractional difference operator defined by the binomial series

$$(1-L)^d = \sum_{j=0}^{\infty} \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)} L^j,$$

where Γ is the gamma function, see Beran (1994), Doukhan et al. (2003) and Robinson (2003) among others.

The coefficient of skewness and kurtosis are defined as

$$S = \frac{\mu_3}{\sigma^3} = \frac{E(x_t-\mu)^3}{\left(E(x_t-\mu)^2\right)^{3/2}} \quad \text{and} \quad K = \frac{\mu_4}{\sigma^4} = \frac{E(x_t-\mu)^4}{\left(E(x_t-\mu)^2\right)^2}, \quad \mu_2 = \sigma^2.$$

Sample estimates of S and K are obtained by replacing population moments $\mu_j = E(x_t - \mu)^j$ by the sample moments $\widehat{\mu}_j = \sum_{k=1}^n (x_k - \overline{x}_n)^j/n$, $\overline{x}_n = \sum_{k=1}^n x_k/n$, i.e.

$$\widehat{S} = \frac{\widehat{\mu}_3}{\widehat{\mu}_2^{3/2}} \quad \text{and} \quad \widehat{K} = \frac{\widehat{\mu}_4}{\widehat{\mu}_2^2}. \tag{3}$$

The statistic $\widehat{S}$ is useful for testing symmetry of data around the sample mean, see Delong and Summers (1985). The statistic $\widehat{K}$ is informative about the tail behaviour of data in many empirical studies, see Boumahdi (1996) and Heinz (2001). There are many tests of normality in the literature. Almost all these tests can be gathered into four classes. The first class measures the distance between the theoretical distribution function and the empirical distribution function (Kolmogorov 1933; Anderson and Darling 1954). The second class of statistics is derived by combining skewness and kurtosis (Bowman and Shenton 1975; Jarque and Bera 1987). The third class is based on generalization of the classical chi-square distance (Pearson 1900). The last class relies on linear regression procedures (Shapiro and Wilk 1965 and D'Agostino 1972 which are based on order statistics). See Yazici and Yolacan (2007) for comparison of various tests of normality.

In this paper we will be interested to the second class by considering the Jarque–Bera statistic (1987) given by $JB = (\widehat{S}^2/6 + (\widehat{K} - 3)^2/24)$. It has been extensively used to find out whether a sample is drawn from a normal distribution or not. It has become very popular since it is very easy to compute. See, for example, Hassler and Wolters (1995), Caporin (2003), Forsberg and Ghysels (2007) and Ajmi et al. (2008) among others. If (x_t) are independent and identically distributed (i.i.d.) $N(\mu, \sigma^2)$ then it is well known that

$$\sqrt{n}\begin{pmatrix}\widehat{\mu}_3\\ \widehat{\mu}_4 - 3\widehat{\mu}_2^2\end{pmatrix} \xrightarrow{\mathcal{L}} N(0, \Sigma), \quad \Sigma = \begin{pmatrix}6\mu_2^3 & 0\\ 0 & 24\mu_2^4\end{pmatrix}, \tag{4}$$

which implies that $JB \xrightarrow{\mathcal{L}} \chi_2^2$, a chi-square distribution with 2 degrees of freedom.

As the skewness and kurtosis measures are based on moments of the data, these tests can have a large size distortion in many situations: heteroskedastic data, presence of outliers and correlated data. Moreover, without taking into account such situations can make these tests worthless.

Concerning heteroskedastic data, Fiorentini et al. (2004) show that the Jarque–Bera test can still be applied to a broad class of *GARCH-M* processes,

$$x_t = \mu_t(\theta) + \sigma_t(\theta)\varepsilon_t, \tag{5}$$

but it can have a size distortions for processes such that the following condition is violated

$$\sqrt{\frac{3n}{2}}\left\{\frac{1}{n}\sum_{t=1}^{n}\varepsilon_t^2(\widetilde{\theta}_n) - 1\right\} = o_p(1),$$

where $\varepsilon_t(\widetilde{\theta}_n)$ are the estimated standardized innovations and $\widetilde{\theta}_n$ is the pseudomaximum likelihood estimator of θ, (see Fiorentini et al. 2004, p. 309).

Some robust versions of the Jarque–Bera test have been suggested by Brys et al. (2004) and Gel and Gastwirth (2008) to handle outliers.

For correlated data, such as *ARMA* processes, if (x_t) is a Gaussian short memory then (see Lomnicki 1961; Gasser 1975) the convergence (4) becomes

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

where

$$F^{(r)} = \sum_{k \in Z} (\gamma_x(k))^r, \quad r = 3, 4. \tag{7}$$

Gasser (1975) suggests consistent estimators of $F^{(3)}$and $F^{(4)}$ by truncating the infinite sums in (7). Lobato and Velasco (2004) consider an estimator of $F^{(r)}$ which is the sample analog of (7), i.e.

$$\widehat{F}^{(r)} = \sum_{|k|<n} (\widehat{\gamma}_x(k))^r, \quad \widehat{\gamma}_x(k) = \frac{1}{n} \sum_{t=1}^{n-|k|} (x_{t+k} - \overline{x}_n)(x_t - \overline{x}_n). \tag{8}$$

They show that if (x_t) is a Gaussian short memory then $\widehat{F}^{(r)}$ is a consistent estimator of $F^{(r)}$, and hence a statistic to test normality under serially correlated data is given by

$$G = n\left(\frac{\widehat{\mu}_3^2}{6\widehat{F}^{(3)}} + \frac{\left(\widehat{\mu}_4 - 3\widehat{\mu}_2^2\right)^2}{24\widehat{F}^{(4)}}\right), \tag{9}$$

which has an asymptotic χ_2^2 distribution.

Bai and Ng (2005) consider short memory processes that satisfy the following Central Limit Theorem (CLT)

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} Z_t \xrightarrow{\mathcal{L}} N(0, \Phi), \tag{10}$$

where $Z_t = (x_t - \mu, (x_t - \mu)^2 - \sigma^2, (x_t - \mu)^3, (x_t - \mu)^4 - 3\sigma^4)'$ and Φ is the spectral density matrix at frequency 0 of Z_t. By considering a consistent estimator of Φ, they suggest a robust version of JB to test normality under serial correlation. Like Lobato and Velasco's (2004) results, the normalization needed is equal to n and the limiting distribution is a χ^2.

However, Lobato and Velasco's (2004) and Bai and Ng's (2005) results no longer hold if the data are long memory. Lobato and Velasco's (2004) results are based on building consistent estimators of $F^{(3)}$and $F^{(4)}$, but if (x_t) satisfies (1) with $3/8 < d < 1/2$ then $F^{(3)} = F^{(4)} = \infty$. In Bai and Ng (2005), serial dependence in (x_t) was explicitly taken into account through the spectral density at frequency 0 of (x_t). If (x_t) is long memory then the spectral density is unbounded at frequency 0 and the CLT (10) no longer holds. In this paper we obtain a very different picture. In Theorems 1 and 2, we will show that the normalizations needed for the sample skewness and kurtosis depend on the long memory parameter d. Moreover, the limiting distributions of which are not χ^2.

We shall consider covariance stationary processes satisfying (1) with the following $MA(\infty)$ representation

$$x_t = \mu + \sum_{j=0}^{\infty} \psi_j u_{t-j}, \tag{11}$$

where (u_j) is a sequence of i.i.d. random variables with zero mean and variance σ_u^2, (ψ_j) is a sequence which decays hyperbolically, i.e.

$$\psi_j \sim \delta j^{-\beta}, \quad \delta > 0, \quad \beta = 1 - d, \quad \text{as } j \to \infty, \tag{12}$$

$$E(u_t^4) = \sigma_u^4(3 + \kappa) \quad \text{for some constant } \kappa \geq 0. \tag{13}$$

It can be shown that the *ARFIMA* process satisfies the representation (11) with $\psi_j \sim \frac{|\theta(1)|}{|\phi(1)|\Gamma(d)} j^{d-1}$ as $j \to \infty$ (see Hosking 1996, p. 272).

Our paper excludes processes with long memory in volatility such as fractionally integrated generalized autoregressive conditional heteroskedasticity (*FIGARCH*) and long memory stochastic volatility (*LMSV*) processes. Recall that a *FIGARCH* is defined by

$$x_t = \sigma_t \varepsilon_t, \tag{14}$$

(ε_t) is a white noise with mean 0 and variance σ^2 and x_t^2 is an *ARFIMA*(p, d, q):

$$\phi(L)(1 - L)^d x_t^2 = \theta(L) v_t, \quad v_t = x_t^2 - \sigma_t^2.$$

It is well known that the unconditional variance of x_t is infinite (see Baillie et al. 1996, p. 3) and neither x_t nor x_t^2 are covariance stationary, consequently our result can not to be applied to test normality of x_t or x_t^2 even if the process x_t^2 admits the $MA(\infty)$ representation (11)–(13).

The long memory stochastic volatility process, introduced by Breidt et al. (1998), is defined by (14) where (ε_t) is a sequence of i.i.d. random variables with zero mean and variance 1, $\sigma_t = \sigma \exp(v_t/2)$ and (v_t) is an *ARFIMA* independent of (ε_t). If both ε_t and v_t are Gaussian it can be shown that the coefficient of kurtosis of (x_t) is given by $K = 3(\exp(\gamma_v(0)) - 1) > 3$, consequently the process (x_t) displays an excess of kurtosis and hence is non Gaussian.

Throughout this paper we shall use the following notations:

$\xrightarrow{P}$ denotes the convergence in probability.

$\xrightarrow{\mathcal{L}}$ denotes the convergence in distribution.

$||X||_p = \left\{E(|X|^p\right\}^{1/p}$ denotes the Euclidian norm of the random variable $X \in L^p(\Omega)$.

$\xrightarrow{L^p(\Omega)}$ denotes the convergence in the $L^p(\Omega)$ space.

$a_j \sim b_j$ as $j \to \infty$ means that $a_j/b_j \to 1$ as $j \to \infty$.

$u_n = O_p(v_n)$ means that there exists a positive constant C such that $P(|u_n/v_n| < C) = 1$.

$u_n = o_p(v_n)$ means that $u_n/v_n \overset{P}{\to} 0$.

$\mathbf{H}_0$: data have a normal distribution.

$\mathbf{H}_1$: data do not have a normal distribution.

2 Preliminary results

Assume that the process (x_t) satisfies (1) and let $y_t = (x_t - \mu)/\sigma$ be the process with zero mean and variance 1,

$$S_{i,n} = \sum_{k=1}^{n} y_k^i,\ H_{i,n} = \sum_{k=1}^{n} H_i(y_k), \quad 1 \le i \le 4, \tag{15}$$

where $H_n(x)$ is the nth Hermite polynomial which is given by

$$H_n(x) = (-1)^n e^{\frac{x^2}{2}} \frac{d^n}{dx^n}\left(e^{-\frac{x^2}{2}}\right), \quad n = 0, 1, \ldots$$

It can also be computed recursively by

$$H_n(x) = xH_{n-1}(x) - (n-1)H_{n-2}(x), \quad n = 2, \ldots, \quad H_0(x) = 1, \quad H_1(x) = x.$$

The first five Hermite polynomials are

$$H_0(x) = 1, \quad H_1(x) = x, \quad H_2(x) = x^2 - 1, \quad H_3(x) = x^3 - 3x,$$
$$H_4(x) = x^4 - 6x^2 + 3.$$

The limiting distribution of the sample skewness and kurtosis coefficients will be established by using the following two lemmas.

Lemma 1 *Assume that the process* (x_t) *is covariance stationary satisfying* (1) *and define* $c = c_1/\sigma^2$,

$$Z_m = \left(\frac{4m!}{[2-2m(1-2d)]\,[2-m(1-2d)]}\right)^{1/2} \overline{Z}_m(1),$$
$$\overline{Z}_m(t) = K(m,d) \int_{R^m}'' \frac{e^{i(\lambda_1+\cdots+\lambda_m)t} - 1}{i(\lambda_1+\cdots+\lambda_m)} |\lambda_1|^{-d} \cdots |\lambda_m|^{-d} \tag{16}$$
$$\times W(d\lambda_1) \cdots W(d\lambda_m),$$
$$K(m,d) = \left(\frac{[2-2m(1-2d)]\,[2-m(1-2d)]}{4m!\,(2\Gamma(1-2d)\sin d\pi)^m}\right)^{1/2},$$

where $W(.)$ is the Gaussian "white noise" complex–valued random measure satisfying:

$$W(d\lambda) = \overline{W(-d\lambda)}, \quad EW(d\lambda) = 0, \quad E\left(W(d\lambda)\overline{W(d\mu)}\right) = \begin{cases} d\lambda & \text{if } \lambda = \mu \\ 0 & \text{if } \lambda \neq \mu \end{cases},$$

$\int_{R^m}''$ *is the multiple Wiener-Itô integral defined in* Major (1981). *Then under* $\boldsymbol{H}_0$ *we have*

1. *If* $0 < d < 1/2$ *then*

$$\frac{H_{1,n}}{c^{1/2}n^{d+1/2}} \xrightarrow{\mathcal{L}} Z_1. \tag{17}$$

2. *If* $1/4 < d < 1/2$ *then*

$$\left(\frac{H_{1,n}}{c^{1/2}n^{d+1/2}}, \frac{H_{2,n}}{cn^{2d}}\right) \xrightarrow{\mathcal{L}} (Z_1, Z_2)\,. \tag{18}$$

3. *If* $1/3 < d < 1/2$ *then*

$$\mathcal{Z}_{1,n} = \left(\frac{H_{m,n}}{c^{m/2}n^{1-(1/2-d)m}}, 1 \leq m \leq 3\right) \xrightarrow{\mathcal{L}} \mathcal{Z}_1 = (Z_m, 1 \leq m \leq 3)\,. \tag{19}$$

4. *If* $3/8 < d < 1/2$ *then*

$$\mathcal{Z}_{2,n} = \left(\frac{H_{m,n}}{c^{m/2}n^{1-(1/2-d)m}}, 1 \leq m \leq 4\right) \xrightarrow{\mathcal{L}} \mathcal{Z}_2 = (Z_m, 1 \leq m \leq 4)\,. \tag{20}$$

Proof The proof can be obtained from (Taqqu 1975; Dobrushin and Major 1979) and hence is omitted.

Lemma 2 *Assume that the process* (x_t) *is covariance stationary satisfying* (1) *and* (11)–(13) *and let* $\widehat{\gamma}_x(k) = \sum_{t=1}^{n-|k|}(x_{t+k} - \overline{x}_n)(x_t - \overline{x}_n)/n$ *be the empirical estimator of* $\gamma_x(k)$. *Then for all fixed* $k \geq 0$, $\widehat{\gamma}_x(k)$ *is a consistent estimator of* $\gamma_x(k)$, *i.e.*

$$\widehat{\gamma}_x(k) \xrightarrow{P} \gamma_x(k). \tag{21}$$

Proof From Theorem 3 of Hosking (1996), it is not difficult to show that for all k,

$$||\widehat{\gamma}_x(k) - \gamma_x(k)||_2 \sim \begin{cases} c_1^2 n^{2(2d-1)} & \text{if } \frac{1}{4} < d < \frac{1}{2} \\ c_2^2 \frac{\log n}{n} & \text{if } d = \frac{1}{4} \\ c_3^2 n^{-1} & \text{if } 0 < d < \frac{1}{4}, \end{cases} \tag{22}$$

for some positive constants c_1, c_2 and c_3, which implies that for all $d \in (0, 1/2)$,

$$\widehat{\gamma}_x(k) \overset{L^2(\Omega)}{\rightarrow} \gamma_x(k) \quad \text{for all } k \geq 0,$$

and then $\widehat{\gamma}_x(k) \overset{P}{\rightarrow} \gamma_x(k)$ for all $k \geq 0$. Hence the desired conclusion holds.

3 Behaviour of the sample skewness

Let

$$Z_S = Z_3 - 3Z_1Z_2 + 2Z_1^3, \tag{23}$$

where the random variables Z_m, $1 \leq m \leq 3$, are given by (16), $\sigma_{H_3}^2 = \sum_{k \in Z} \gamma_{H_3}(k)$, $\gamma_{H_3}(k)$ is the covariance function of the process $H_3(y_k) = y_k^3 - 3y_k$, $y_t = (x_t - \mu)/\sigma$.

Theorem 1 *Assume that the process (x_t) is covariance stationary satisfying (1) and (11)–(13) and define $c = c_1/\sigma^2$. Then under $\boldsymbol{H}_0$ we have*

1. *If $1/3 < d < 1/2$ then*

$$n^{3(1/2-d)}\widehat{S} \overset{\mathcal{L}}{\rightarrow} c^{3/2} Z_S. \tag{24}$$

2. *If $d = 1/3$ then*

$$\frac{n^{1/2}}{(\log n)^{1/2}}\widehat{S} \overset{\mathcal{L}}{\rightarrow} \left(12c^3\right)^{1/2} N(0, 1). \tag{25}$$

3. *If $0 < d < 1/3$ then*

$$n^{1/2}\widehat{S} \overset{\mathcal{L}}{\rightarrow} \sigma_{H_3} N(0, 1). \tag{26}$$

Proof See the Appendix.

Usually, the statistic $b_1 = n\widehat{S}^2/6$ is used to test symmetric behaviour of time series. Theorem 1 implies that

- If $1/3 < d < 1/2$ then $b_1 \sim Cn^{2(3d-1)}$ and hence $b_1 \overset{P}{\rightarrow} +\infty$,
- If $d = 1/3$ then $b_1 \sim C \log n$ which implies that $b_1 \overset{P}{\rightarrow} +\infty$,
- If $0 < d < 1/3$ then $b_1 \overset{\mathcal{L}}{\rightarrow} (\sigma_{H_3}^2/6)\chi_1^2$.

Consequently, for all $1/3 \leq d < 1/2$, a test based on b_1 will reject more often the null of symmetric behaviour even data are normally distributed. The rejection frequencies increase with d and the sample size n.

The normalization needed to establish the limiting distribution of the sample skewness coefficient $\widehat{S}$ can be written as $n^{\alpha_n(d)}$ where

$$\alpha_n(d) = \begin{cases} \frac{1}{2} & \text{if } 0 < d < \frac{1}{3} \\ \frac{1}{2} - \frac{1}{2}\frac{\log\log n}{\log n} & \text{if } d = \frac{1}{3} \\ \frac{3}{2} - 3d & \text{if } \frac{1}{3} < d < \frac{1}{2}. \end{cases}$$

The function α_n is a nondecreasing and continuous for all $d \in (0, 1/2)$, $d \neq 1/3$.

4 Behaviour of the sample kurtosis

Let

$$Z_K = Z_4 - 4Z_1Z_3 + 12Z_1^2Z_2 - 3Z_2^2 - 6Z_1^4, \tag{27}$$

where the random variables Z_m, $1 \le m \le 4$, are given by (16); $\sigma_{H_4}^2 = \sum_{k\in Z} \gamma_{H_4}(k)$, $\gamma_{H_4}(k)$ is the covariance function of the process $H_4(y_k) = y_k^4 - 6y_k^2 + 3$, $y_t = (x_t - \mu)/\sigma$.

Theorem 2 *Assume that the process (x_t) is covariance stationary satisfying (1) and (11)–(13) and define $c = c_1/\sigma^2$. Then under $\boldsymbol{H}_0$ we have*

1. *If $3/8 < d < 1/2$ then*

$$n^{4(1/2-d)}(\widehat{K} - 3) \xrightarrow{\mathcal{L}} c^2 Z_K. \tag{28}$$

2. *If $d = 3/8$ then*

$$\frac{n^{1/2}}{(\log n\,)^{1/2}}(\widehat{K} - 3) \xrightarrow{\mathcal{L}} \left(48c^2\right)^{1/2} N(0, 1). \tag{29}$$

3. *If $0 < d < 3/8$ then*

$$n^{1/2}(\widehat{K} - 3) \xrightarrow{\mathcal{L}} \sigma_{H_4} N(0, 1). \tag{30}$$

Proof See the Appendix.

The leptokurtic behaviour of data is tested by the statistic $b_2 = n(\widehat{K} - 3)^2/24$. If data are long memory, then Theorem 2 implies that

- If $3/8 < d < 1/2$ then $b_2 \sim Cn^{2d}$ and hence $b_2 \xrightarrow{P} +\infty$,
- If $d = 3/8$ then $b_2 \sim C \log n$ which implies that $b_2 \xrightarrow{P} +\infty$,
- If $0 < d < 3/8$ then

$$b_2 \xrightarrow{\mathcal{L}} b_{2,\infty} = (\sigma_{H_4}^2/24)\chi_1^2 = \left(1 + \sum_{k \in Z^*} (\rho_x(k))^4 /24\right)\chi_1^2,$$

where $\rho_x(k) = \gamma_x(k)/\sigma^2$ is the autocorrelation function of (x_t). Note that $b_{2,\infty} \geq \chi_1^2$. Therefore, for all $0 < d < 1/2$, a test based on b_2 will reject more often the null even data are normally distributed. The rejection frequencies increase with d and the sample size n if $3/8 \leq d < 1/2$.

The normalization needed to establish the limiting distribution of the sample excess of kurtosis coefficient $\widehat{K} - 3$ can be written as $n^{\beta_n(d)}$ where

$$\beta_n(d) = \begin{cases} \frac{1}{2} & \text{if } 0 < d < \frac{3}{3} \\ \frac{1}{2} - \frac{1}{2}\frac{\log\log n}{\log n} & \text{if } d = \frac{3}{8} \\ 2 - 4d & \text{if } \frac{3}{8} < d < \frac{1}{2}. \end{cases}$$

Note that the function β_n is discontinuous at $d = 3/8$.

5 Behaviour of the Jarque–Bera test statistic

The Jarque–Bera test statistic *JB* combines the skewness and the kurtosis to test normality of data. If data are long memory then the behaviour of *JB* can be described by the following

Theorem 3 *Assume that the process (x_t) is covariance stationary satisfying (1) and (11)–(13) and define $c = c_1/\sigma^2$, $\sigma_{H_i}^2 = \sum_{k \in Z} \gamma_{H_i}(k)$, $\gamma_{H_i}(k)$ is the covariance function of the process $H_i(y_k)$, $i = 3, 4$, $y_t = (x_t - \mu)/\sigma$. Then under $\boldsymbol{H}_0$ we have*

1. *If $3/8 < d < 1/2$ then*

$$\frac{n^{6(1/2-d)}}{c^3}\widehat{S}^2 + \frac{n^{8(1/2-d)}}{c^4}(\widehat{K} - 3)^2 \xrightarrow{\mathcal{L}} Z_S^2 + Z_K^2$$

and

$$n^{2(1-3d)} JB \xrightarrow{\mathcal{L}} \frac{c^3}{6} Z_S^2,$$

where Z_S and Z_K are given by (23) and (27) respectively. Therefore $JB \sim Cn^{2(3d-1)}$ and hence $JB \xrightarrow{P} +\infty$.

2. *If $d = 3/8$ then*

$$\frac{n^{3/4}}{c^3}\widehat{S}^2 + \frac{n}{48c^2 \log n}(\widehat{K} - 3)^2 \xrightarrow{\mathcal{L}} Z_S^2 + \chi_1^2 \quad (31)$$

and

$$\frac{JB}{n^{1/4}} \xrightarrow{\mathcal{L}} \frac{c^3}{6} Z_S^2,$$

which implies that $JB \sim Cn^{1/4}$ *and* $JB \xrightarrow{P} +\infty$.

3. *If* $1/3 < d < 3/8$ *then*

$$\frac{n^{6(1/2-d)}}{c^3} \widehat{S}^2 + \frac{n}{\sigma_{H_4}^2} (\widehat{K} - 3)^2 \xrightarrow{\mathcal{L}} Z_S^2 + \chi_1^2 \tag{32}$$

and

$$n^{2(1-3d)} JB \xrightarrow{\mathcal{L}} \frac{c^3}{6} Z_S^2,$$

therefore $JB \sim Cn^{2(3d-1)}$ *and hence* $JB \xrightarrow{P} +\infty$.

4. *If* $d = 1/3$ *then*

$$\frac{n}{12c^3 \log n} \widehat{S}^2 + \frac{n}{\sigma_{H_4}^2} (\widehat{K} - 3)^2 \xrightarrow{\mathcal{L}} \chi_2^2$$

and

$$\frac{1}{\log n} JB \xrightarrow{\mathcal{L}} 2c^3 \chi_1^2,$$

which implies that $JB \sim C \log n$ *and* $JB \xrightarrow{P} +\infty$.

5. *If* $0 < d < 1/3$ *then*

$$n \left(\frac{\widehat{S}^2}{\sigma_{H_3}^2} + \frac{(\widehat{K} - 3)^2}{\sigma_{H_4}^2} \right) \xrightarrow{\mathcal{L}} \chi_2^2.$$

Proof The proof follows by combining Theorems 1 and 2.

5.1 Example

Assume that (x_t) is a Fractional Gaussian Noise, i.e. the stationary Gaussian process with mean 0 and covariance

$$\gamma_x(k) = \frac{\sigma^2}{2} \left\{ |k+1|^{2d+1} - 2\,|k|^{2d+1} + |k-1|^{2d+1} \right\}. \tag{33}$$

and let

$$\mathcal{JB} = n\left(\frac{\widehat{S}^2}{6} + \frac{(\widehat{K}-3)^2}{24}\right) \quad \text{and} \quad \mathcal{JB}_s = n\left(\frac{\widehat{S}^2}{\sigma_{H_3}^2} + \frac{(\widehat{K}-3)^2}{\sigma_{H_4}^2}\right). \tag{34}$$

- If $d = 0$, i.e. (y_t) is an i.i.d. $N(0, 1)$, then $\gamma_y(k) = 0$ for all $k \neq 0$ which implies that $\sigma_{H_3}^2 = 3! \sum_{k\in Z} (\gamma_y(k))^3 = 6$ and $\sigma_{H_4}^2 = 4! \sum_{k\in Z} (\gamma_y(k))^4 = 24$ since $\gamma_{H_3}(k) = 3!(\gamma_y(k))^3$ and $\gamma_{H_4}(k) = 4!(\gamma_y(k))^4$, hence $JB_s = JB$.
- If $0 < d < 1/3$, then the covariance $\gamma_x(k)$ is positive, which implies that $\sigma_{H_3}^2 > 6$ and $\sigma_{H_3}^2 > 24$, consequently $\mathcal{JB} > JB_s$. Let p-value$_{\mathcal{JB}} = 1 - F_{\chi_2^2}(\widehat{\mathcal{JB}})$ and p-value$_{\mathcal{JB}_s}$ $=1 - F_{\chi_2^2}(\widehat{\mathcal{JB}_s})$ be the p-value corresponding to the observed value of the Jarque–Bera statistic $\widehat{\mathcal{JB}}$ and the observed statistic $\widehat{\mathcal{JB}_s}$ respectively, where $F_{\chi_2^2}(.)$ is the cumulative χ_2^2 distribution. We have p-value$_{\mathcal{JB}} <$ p-value$_{\mathcal{JB}_s}$, hence if the Jarque–Bera test is applied without taking into accounts the correlation of data, then the test will reject the null of normality more often.
- If $1/3 \leq d < 1/2$, then the rejection frequencies increase with d and the sample size n.

As a conclusion, in long memory environment, a rescaled versions of the skewness, kurtosis, and the Jarque–Bera normality test statistics are needed. Moreover, a simulation/bootstrap method for computing critical values or p-values of the nonstandard distributions given in Theorems 1–3 is required to make a correct statistical inference.

6 Monte Carlo simulations

In this section, we study the size of the Jarque–Bera test. We generate two Gaussian long memory processes. The first one is an *ARFIMA* generated by using the function arima.fracdiff.sim in SPLUS 6.0. The second is a fractional Gaussian noise (FGN) by using Beran's code in Beran (1994).

We carry out an experiment of 1,000 samples for three long memory processes: an *ARFIMA*$(0, d, 0)$, an *ARFIMA*$(1, d, 1)$ and a fractional Gaussian noise. We consider five values for d, $d = 0$, $d = 3/10$, $d = 1/3$, $d = 3/8$ and $d = 7/16$; and we use four different sample sizes, $n = 500$, $n = 1{,}000$, $n = 2{,}000$ and $n = 5{,}000$.

6.1 The analysis of the simulation results for an *ARFIMA (0,d,0)* and an *ARFIMA (1,d,1)*

The data generating processes are

$$(1-L)^d y_t = u_t, \quad \text{where } (u_t) \sim i.i.d.N(0, 1), \tag{35}$$

and

$$(1-\phi_1 L)(1-L)^d y_t = (1+\theta_1 L)u_t, \quad \text{where } (u_t) \sim i.i.d.N(0, 1), \tag{36}$$

θ_1 is fixed to 0.6, whereas the parameter ϕ_1 takes three values $\phi_1 = 0$, $\phi_1 = 0.5$ and $\phi_1 = 0.8$.

Table 1 Empirical test sizes (in %)

α (%)	$d = 0$	$d = \frac{3}{10}$	$d = \frac{1}{3}$	$d = \frac{3}{8}$	$d = \frac{7}{16}$
$n = 500$					
1	2.4	1.5	1.7	4	5.7
5	5.9	6.7	8.1	9	13.7
10	8	12.2	12.3	12.7	22.3
$n = 1{,}000$					
1	0.8	2.3	3.2	5	10.4
5	5.6	7.7	8.5	10.7	19.3
10	10	13.7	15	17.7	33.3
$n = 2{,}000$					
1	1	1.9	4.1	6.9	17
5	4.7	6.9	10.5	15.7	31.6
10	9.2	13	18.5	21.6	40.5
$n = 5{,}000$					
1	1	3.1	4.1	9.4	26.4
5	5.9	8.3	12.6	19.7	42.4
10	10.6	15.2	19.1	26	52.8

Table 1 contains rejection frequencies of the null hypothesis of normality. Rejection frequencies are based on 1,000 replications generated from the DGP: $(1-L)^d y_t = u_t$, $u_t \sim i.i.d.N(0, 1)$, the nominal significance levels are $\alpha = 1\%$, $\alpha = 5\%$ and $\alpha = 10\%$, the sample sizes are $n = 500$, $n = 1{,}000$, $n = 2{,}000$ and $n = 5{,}000$

6.1.1 The analysis of the simulation results for an ARFIMA$(0, d, 0)$

From Table 1, we observe that if $d = 0$, i.e. the data are i.i.d. then the Jarque–Bera test has a good performance. If $d > 0$, then the Jarque–Bera test suffers from a size distortion; for example, if $d = 7/16$ and $\alpha = 5\%$, then the empirical size of Jarque–Bera test is 42.4% if $n = 5{,}000$. If $d \geq 1/3$, we observe that for α fixed, the rejection frequencies of the null increase with the sample size n. The rejection frequencies increase also with d. If $\alpha = 5\%$, the empirical size is equal to 8.5% if $(d, n) = (1/3, 1{,}000)$ and 10.5% if $(d, n) = (1/3, 2{,}000)$, i.e. if $d = 1/3$ then the empirical size increases by 2% when the sample size n increases from 1,000 to 2,000. However, if $d = 3/8$ and $d = 7/16$ then the empirical size increases by 5% and by 11.3% respectively when the sample size n increases from 1,000 to 2,000. This of course is in accordance with our theoretical finding in Theorem 3 which implies that $JB \sim C \log n$ if $d = 1/3$, $JB \sim Cn^{1/4}$ if $d = 3/8$ and $JB \sim Cn^{5/8}$ if $d = 7/16$.

6.1.2 The analysis of the simulation results for an $ARFIMA(1, d, 1)$

From Table 2, we observe that if $d = 0$, i.e. the data are short memory (the DGP is an *MA(1)*) then the Jarque–Bera test has a small size distortion. If $d > 0$, then the Jarque–Bera test suffers from a size distortion. We observe also that the presence of short memory component increases the size distortion, for example, if $d = 7/16$, $\alpha = 5\%$ and $n = 5{,}000$ then the empirical size of Jarque–Bera test is 42.4% if $\theta_1 = 0$ (Table 1) and increases to 58.9% if $\theta_1 = 0.6$ (Table 2).

From Table 3, we observe that if $d = 0$, i.e. the data are short memory (the DGP is an *ARMA(1,1)*) then the Jarque–Bera test has a size distortion. The effect of the

Table 2 Empirical test sizes (in %), $\phi_1 = 0$

α (%)	$d = 0$	$d = \frac{3}{10}$	$d = \frac{1}{3}$	$d = \frac{3}{8}$	$d = \frac{7}{16}$
$n = 500$					
1	1.9	5.5	6.2	8.9	13.6
5	5.7	13.9	16.4	19.6	37.9
10	8.8	19.6	25.4	28.5	38
$n = 1{,}000$					
1	1.6	6.8	9.9	11.9	21
5	5.3	17.3	20.7	28.4	38.4
10	12.6	24.5	29.6	37	44.7
$n = 2{,}000$					
1	1.4	9.9	11.3	18.5	31.8
5	6.9	20.8	22.1	30	47.8
10	13.1	28.1	31.8	40.3	59.2
$n = 5{,}000$					
1	2.3	10.8	17	23.8	46.2
5	6.5	24.1	28.7	37.9	58.9
10	13.9	32.5	37.3	48.6	66.7

Table 2 contains rejection frequencies of the null hypothesis of normality. Rejection frequencies are based on 1,000 replications generated from the DGP: $(1 - L)^d y_t = u_t + 0.6u_{t-1}$, $u_t \sim i.i.d.N(0, 1)$

Table 3 Empirical test sizes (in %), $\phi_1 = 0.5$

α (%)	$d = 0$	$d = \frac{3}{10}$	$d = \frac{1}{3}$	$d = \frac{3}{8}$	$d = \frac{7}{16}$
$n = 500$					
1	7.3	16.9	17.8	22.3	31.8
5	14.5	31	34.3	40.1	50.2
10	22.6	43.7	44.6	52.9	59.8
$n = 1{,}000$					
1	4.3	18.8	25	31.1	40.7
5	16.9	36.9	41.8	50.6	56.9
10	25.1	48.1	51.2	59.1	66.9
$n = 2{,}000$					
1	6.6	27.9	32	36.8	53.7
5	16.1	42.7	47.3	55.5	65.6
10	23.5	53.8	58.8	63.5	73.7
$n = 5{,}000$					
1	7	31.2	37.6	47.2	64.1
5	16	46.3	52.3	59.9	77.3
10	24.6	55.5	60.5	67.5	80.2

Table 3 contains rejection frequencies of the null hypothesis of normality. Rejection frequencies are based on 1,000 replications generated from the DGP: $(1 - 0.5L)(1 - L)^d y_t = u_t + 0.6u_{t-1}$, $u_t \sim i.i.d.N(0, 1)$

autoregressive component is more marked; for example, if $\alpha = 5\%$ and $n = 5{,}000$ then the empirical size of Jarque–Bera test is 16% if $\phi_1 = 0.5$ (Table 3) and increases to 39.7% if $\phi_1 = 0.8$ (Table 3). If $d > 0$, then the Jarque–Bera test suffers from a

Table 4 Empirical test sizes (in %), $\phi_1 = 0.8$

α (%)	$d = 0$	$d = \frac{3}{10}$	$d = \frac{1}{3}$	$d = \frac{3}{8}$	$d = \frac{7}{16}$
$n = 500$					
1	15.9	38.7	43.2	48.7	55.7
5	29.4	57.3	63.9	66.8	70.7
10	43	67.9	68.5	71.9	78.2
$n = 1{,}000$					
1	19	49.6	51.4	57.6	66.4
5	37	62.1	66.2	70.1	78.8
10	44.5	73	75.3	76.9	82.1
$n = 2{,}000$					
1	22.5	55.2	61	66.5	72.3
5	38.9	66.6	72.2	77.8	82.5
10	46.5	74.6	77	81.5	84.9
$n = 5{,}000$					
1	24	60.9	66.2	72.2	78.9
5	39.7	73.5	77.7	80.5	88.1
10	51.6	77.1	79.5	85	88.8

Table 4 contains rejection frequencies of the null hypothesis of normality. Rejection frequencies are based on 1,000 replications generated from the DGP: $(1 - 0.8L)(1 - L)^d y_t = u_t + 0.6u_{t-1}$, $u_t \sim i.i.d.N(0, 1)$

Table 5 Empirical test sizes (in %)

α (%)	$d = 0$	$d = \frac{3}{10}$	$d = \frac{1}{3}$	$d = \frac{3}{8}$	$d = \frac{7}{16}$
$n = 500$					
1	2.4	2.2	3	4.6	7.1
5	5.9	6.2	8.7	13.5	18.8
10	8	12.6	15.2	18.5	27.5
$n = 1{,}000$					
1	0.8	2.6	4	5.9	14.7
5	5.6	8.5	11.9	15.4	27.2
10	10	15.4	17.7	24.2	37.1
$n = 2{,}000$					
1	1	3.3	5.1	9.8	24
5	4.7	8.4	14	19.7	36
10	9.2	14.9	21.3	27.7	45.9
$n = 5{,}000$					
1	1	3.6	6	14.9	33.9
5	5.9	11	16.5	26.2	47.2
10	10.6	18.7	25.2	34.7	59.5

Table 5 contains rejection frequencies of the null hypothesis of normality. Rejection frequencies are based on 1,000 replications generated from the Fractional. Gaussian Noise (33)

large size distortion. For example, if $d = 7/16$, $\alpha = 5\%$ and $n = 5{,}000$ then the empirical size of Jarque–Bera test is 77.3% if $\phi_1 = 0.5$ (Table 3) and increases to 88.1% if $\phi_1 = 0.8$ (Table 4).

6.2 The analysis of the simulation results for FGN

The DGP is the FGN given by (33) with $\sigma^2 = 1$. From Table 5, we observe that the results obtained for the Fractional Gaussian Noise are fairly similar to those obtained in Table 1 for an $ARFIMA(0, d, 0)$. The rejection frequencies are slightly greater in the former case. For example, if $d = 7/16$, $\alpha = 5\%$ and $n = 5{,}000$ then the empirical size of Jarque–Bera test is 42.4% if the DGP is an $ARFIMA(0, d, 0)$ (Table 1) and increases to 47.2% if the DGP is a Fractional Gaussian Noise (Table 5).

Acknowledgments I would like to thank the referees for their constructive comments.

Appendix

Proof of Theorem 1

1. By applying Lemma 2, it follows that $\widehat{S}$ has the same limiting distribution as $\widetilde{S}$ given by

$$\begin{aligned}\widetilde{S} &= \frac{\widehat{\mu}_3}{\left(\sigma^2\right)^{3/2}}\\ &= \frac{1}{n}\sum_{k=1}^{n}\left(\frac{x_k - \overline{x}_n}{\sigma}\right)^3\\ &= \frac{1}{n}\sum_{k=1}^{n}\left(y_k - \overline{y}_n\right)^3, \end{aligned}\tag{37}$$

since $y_k = (x_k - \mu)/\sigma$ implies that $\overline{y}_n = (\overline{x}_n - \mu)/\sigma$ and that

$$\begin{aligned} y_k - \overline{y}_n &= \frac{x_k - \mu}{\sigma} - \frac{1}{\sigma}(\overline{x}_n - \mu)\\ &= \frac{x_k - \overline{x}_n}{\sigma}. \end{aligned}$$

Moreover, by using (15), $\widetilde{S}$ can be rewritten as

$$\begin{aligned}\widetilde{S} &= \frac{1}{n}\sum_{k=1}^{n} y_k^3 - 3\overline{y}_n\frac{1}{n}\sum_{k=1}^{n} y_k^2 + 2\left(\overline{y}_n\right)^3\\ &= \frac{S_{3,n}}{n} - \frac{3}{n^2}S_{1,n}S_{2,n} + 2\left(\frac{S_{1,n}}{n}\right)^3\\ &= \frac{H_{3,n} + 3H_{1,n}}{n} - \frac{3}{n^2}H_{1,n}(H_{2,n} + n) + 2\left(\frac{H_{1,n}}{n}\right)^3\\ &= \frac{H_{3,n}}{n} - \frac{3}{n^2}H_{1,n}H_{2,n} + 2\left(\frac{H_{1,n}}{n}\right)^3. \end{aligned}\tag{38}$$

Hence

$$\frac{n^{3(1/2-d)}}{c^{3/2}}\widetilde{S} = \frac{n^{3(1/2-d)}}{c^{3/2}}\left\{\frac{H_{3,n}}{n} - \frac{3}{n^2}H_{1,n}H_{2,n} + 2\left(\frac{H_{1,n}}{n}\right)^3\right\}$$
$$= \frac{H_{3,n}}{c^{3/2}n^{3d-1/2}} - 3\left(\frac{H_{1,n}}{c^{1/2}n^{d+1/2}}\right)\left(\frac{H_{2,n}}{cn^{2d}}\right) + 2\left(\frac{H_{1,n}}{c^{1/2}n^{d+1/2}}\right)^3.$$

Note that $n^{3(1/2-d)}\ \widetilde{S}\ /c^{3/2}$ is a continuous functional from R^3 into R of the vector $\mathcal{Z}_{1,n}$ given by the left hand side of (19), hence Lemma 1 and the continuous mapping theorem imply that

$$\frac{n^{3(1/2-d)}}{c^{3/2}}\widetilde{S} \xrightarrow{\mathcal{L}} Z_S,$$

therefore (24) follows.

2. From Lemma 1, it follows that for all $d > 1/4$ we have $H_{1,n} = O_p(n^{d+1/2})$ and $H_{2,n} = O_p(n^{2d})$. Hence if $d = 1/3$ then $H_{1,n} = O_p(n^{5/6})$ and $H_{2,n} = O_p(n^{2/3})$. Let $L(|t|) = \gamma_{H_3}(t)\,|t|$ and $L_3(n) = \sum_{k=-n}^{n}\gamma_{H_3}(k)$. By using the fact that the Hermite polynomials satisfy

$$E\left(H_j(\eta)H_k(\zeta)\right) = \begin{cases} k!\,(E(\eta\zeta))^k & \text{if } j = k,\\ 0 & \text{if } j \neq k,\end{cases} \tag{39}$$

we have

$$\begin{aligned}\gamma_{H_3}(k) &= E\left(H_3(y_{t+k})H_3(y_t)\right)\\ &= 3!(\gamma_y(k))^3\\ &\sim 3!\left(ck^{2d-1}\right)^3\\ &\sim \frac{3!c^3}{k},\end{aligned}$$

since from (1), $\gamma_y\,(k) \sim ck^{2d-1}$.

Hence $L(|t|) \sim 3!c^3$ and $L_3(n) \sim 12c^3\log(n)$. Since L and L_3 are slowly varying functions (i.e. for all $a > 0$, $L(at)/L(t) \to 1$ and $L_3(at)/L_3(t) \to 1$ if $t \to +\infty$), and $L_3(n) \to \infty$ as $n \to \infty$, application of Theorem 6 of Giraitis and Surgailis (1985) leads to

$$\frac{1}{(nL_3(n)\,)^{1/2}}\sum_{k=1}^{n}H_3(y_k) \xrightarrow{\mathcal{L}} N(0,1)$$

or

$$\frac{H_{3,n}}{(n\log n\,)^{1/2}} = \frac{1}{(n\log n\,)^{1/2}}\sum_{k=1}^{n}H_3(y_k) \xrightarrow{\mathcal{L}} \left(12c^3\right)^{1/2}N(0,1).$$

From (38), it follows that

$$\begin{aligned}\frac{n^{1/2}}{(\log n)^{1/2}}\widetilde{S} &= \frac{H_{3,n}}{(n\log n)^{1/2}} - \frac{3n^{1/2}}{n^2(\log n)^{1/2}}H_{1,n}H_{2,n} + 2\frac{n^{1/2}}{(\log n)^{1/2}}\left(\frac{H_{1,n}}{n}\right)^3\\ &= \frac{H_{3,n}}{(n\log n)^{1/2}} - O_p\left(\frac{n^{1/2}}{n^2(\log n)^{1/2}}n^{5/6}n^{2/3}\right)\\ &\quad + O_p\left(\frac{n^{1/2}}{n^3(\log n)^{1/2}}n^{5/2}\right)\\ &= \frac{H_{3,n}}{(n\log n)^{1/2}} + O_p\left(\frac{1}{(\log n)^{1/2}}\right).\end{aligned}$$

Consequently

$$\frac{n^{1/2}}{(\log n\ )^{1/2}}\widetilde{S} \xrightarrow{\mathcal{L}} \left(12c^3\right)^{1/2} N(0,1).$$

Therefore (25) follows.

3. If $0 < d < 1/3$ then $\sigma^2_{H_3} = \sum_{k\in Z}\gamma_{H_3}(k) \sim 3!c^3\sum_{k\in Z}\left(k^{2d-1}\right)^3 < \infty$, hence from Theorem 6 of Giraitis and Surgailis (1985) we have

$$\frac{H_{3,n}}{n^{1/2}} \xrightarrow{\mathcal{L}} \sigma_{H_3}N(0,1), \tag{40}$$

moreover (17) implies that

$$\begin{aligned}n^{1/2}\left(\frac{H_{1,n}}{n}\right)^3 &= O_p\left(n^{1/2}\left(\frac{n^{d+1/2}}{n}\right)^3\right)\\ &= O_p\left(n^{1/2}\left(\frac{n^{d+1/2}}{n}\right)^3\right)\\ &= O_p\left(n^{3d-1}\right)\\ &= o_p(1) \text{ since } d < \frac{1}{3}.\end{aligned}$$

Hence

$$\begin{aligned}n^{1/2}\widetilde{S} &= \frac{H_{3,n}}{n^{1/2}} - \frac{3n^{1/2}}{n^2}H_{1,n}H_{2,n} \ + 2n^{1/2}\left(\frac{H_{1,n}}{n}\right)^3\\ &= \frac{H_{3,n}}{n^{1/2}} - \frac{3n^{1/2}}{n^2}H_{1,n}H_{2,n} \ + o_p(1).\end{aligned} \tag{41}$$

Now we will examine the behaviour of second term in the right hand side of (41).

3.1. If $1/4 < d < 1/3$ then from (18) we obtain $H_{1,n} = O_p(n^{d+1/2})$ and $H_{2,n} = O_p(n^{2d})$ which implies that

$$\begin{aligned} \frac{3n^{1/2}}{n^2} H_{1,n} H_{2,n} &= O_p\left(\frac{n^{1/2}}{n^2} n^{d+1/2} n^{2d}\right) \\ &= O_p\left(n^{3d-1}\right) \\ &= o_p(1) \text{ since } d < \frac{1}{3}. \end{aligned} \tag{42}$$

3.2. If $d = 1/4$ then let $L_2(n) = \sum_{k=-n}^{n} \gamma_{H_2}(k)$, from (39) it follows that

$$\begin{aligned} \gamma_{H_2}(k) &= E\ (H_2(y_{t+k}) H_2(y_t)) \\ &= 2!(\gamma_y(k))^2 \\ &\sim 2\left(ck^{2d-1}\right)^2 \\ &\sim \frac{2c^2}{k}. \end{aligned}$$

This implies that $L_2(n) \sim 4c^2 \log(n)$. Since $L_2(n)$ is a slowly varying function and $L_2(n) \to \infty$ as $n \to \infty$, application of Theorem 6 of Giraitis and Surgailis (1985) leads to

$$\frac{1}{(n \log n\ )^{1/2}} \sum_{k=1}^{n} H_2(y_k) \xrightarrow{\mathcal{L}} 2cN(0, 1),$$

which implies that $H_{2,n} = O_p\left((n \log n\ )^{1/2}\right)$. Consequently

$$\begin{aligned} \frac{3n^{1/2}}{n^2} H_{1,n} H_{2,n} &= O_p\left(\frac{n^{1/2}}{n^2} n^{d+1/2} (n \log n\ )^{1/2}\right) \\ &= O_p\left(\frac{(\log n\ )^{1/2}}{n^{1/4}}\right) \\ &= o_p(1). \end{aligned} \tag{43}$$

3.3. If $0 < d < 1/4$ then $\sigma^2_{H_2} = \sum_{k \in Z} \gamma_{H_2}(k) \sim 2c^2 \sum_{k \in Z} \left(k^{2d-1}\right)^2 < \infty$, hence from Theorem 5 of Giraitis and Surgailis (1985) we have

$$\frac{H_{2,n}}{n^{1/2}} \xrightarrow{\mathcal{L}} \sigma_{H_2} N(0, 1), \tag{44}$$

which implies, together with (17), that

$$\begin{aligned}\frac{3n^{1/2}}{n^2}H_{1,n}H_{2,n} &= O_p\left(\frac{n^{1/2}}{n^2}n^{d+1/2}n^{1/2}\right)\\ &= O_p\left(n^{d-1/2}\right)\\ &= o_p(1) \text{ since } d < \frac{1}{2}. \end{aligned} \tag{45}$$

From (41), (42), (43) and (45), it follows that for all d, $0 < d < 1/3$,

$$n^{1/2}\widetilde{S} = \frac{H_{3,n}}{n^{1/2}} + o_p(1). \tag{46}$$

The desired conclusion (26) follows from (40) and (46).

Proof of Theorem 2 1. By writing $\widehat{K} - 3 = \widehat{\mu}_4/\widehat{\mu}_2^2 - 3 = (\widehat{\mu}_4 - 3\widehat{\mu}_2^2)/\widehat{\mu}_2^2$, and applying Lemma 2, it follows that $\widehat{K} - 3$ has the same limiting distribution as $\widetilde{K}$ given by

$$\begin{aligned}\widetilde{K} &= \frac{\widehat{\mu}_4 - 3\widehat{\mu}_2^2}{\sigma^4}\\ &= \frac{1}{n}\sum_{k=1}^{n}\left(\frac{x_k - \overline{x}_n}{\sigma}\right)^4 - 3\left(\frac{1}{n}\sum_{k=1}^{n}\left(\frac{x_k - \overline{x}_n}{\sigma}\right)^2\right)^2\\ &= \frac{1}{n}\sum_{k=1}^{n}\left(y_k - \overline{y}_n\right)^4 - 3\left(\frac{1}{n}\sum_{k=1}^{n}\left(y_k - \overline{y}_n\right)^2\right)^2.\end{aligned}$$

Straightforward computations lead to

$$\widetilde{K} = \frac{H_{4,n}}{n} - \frac{4}{n^2}H_{1,n}H_{3,n} + \frac{12}{n^3}H_{1,n}^2H_{2,n} - \frac{3}{n^2}H_{2,n}^2 - \frac{6}{n^4}H_{1,n}^4,$$

consequently

$$\begin{aligned}\frac{n^{4(1/2-d)}}{c^2}\widetilde{K} = &\frac{H_{4,n}}{2n^{1-4(1/2-d)}} - 4\left(\frac{H_{1,n}}{c^{1/2}n^{d+1/2}}\right)\left(\frac{H_{3,n}}{c^{3/2}n^{3d-1/2}}\right)\\ &+12\left(\frac{H_{1,n}}{c^{1/2}n^{d+1/2}}\right)\left(\frac{H_{2,n}}{cn^{2d}}\right) - 3\left(\frac{H_{2,n}}{cn^{2d}}\right)^2\\ &-6\left(\frac{H_{1,n}}{c^{1/2}n^{d+1/2}}\right)^4,\end{aligned}$$

since $3/8 < d < 1/2$, the convergence (20) and the continuous mapping theorem imply (28).

2. If $d = 3/8$ then from (19) we deduce that $H_{i,n} = O_p\left(n^{1-(1/2-d)i}\right)$, $i = 1, 2, 3$, and then $H_{1,n} = O_p\left(n^{7/8}\right)$, $H_{2,n} = O_p\left(n^{3/4}\right)$ and $H_{3,n} = O_p\left(n^{5/8}\right)$. Therefore

$$\begin{aligned}
\frac{n^{1/2}}{(\log n)^{1/2}}\widetilde{K} &= \frac{n^{1/2}}{(\log n)^{1/2}} \\
&\quad \times \left(\frac{H_{4,n}}{n} - \frac{4}{n^2}H_{1,n}H_{3,n} + \frac{12}{n^3}H_{1,n}^2H_{2,n} - \frac{3}{n^2}H_{2,n}^2 - \frac{6}{n^4}H_{1,n}^4\right) \\
&= \frac{H_{4,n}}{(n\log n)^{1/2}} + O_p\left(\frac{n^{1/2}}{n^2(\log n)^{1/2}}n^{7/8}n^{5/8}\right) \\
&\quad + O_p\left(\frac{n^{1/2}}{n^3(\log n)^{1/2}}n^{7/4}n^{3/4}\right) + O_p\left(\frac{n^{1/2}}{n^2(\log n)^{1/2}}n^{3/2}\right) \\
&\quad + O_p\left(\frac{n^{1/2}}{n^4(\log n)^{1/2}}n^{7/2}\right) \\
&= \frac{H_{4,n}}{(n\log n)^{1/2}} + O_p\left(\frac{1}{(\log n)^{1/2}}\right) \\
&= \frac{H_{4,n}}{(n\log n)^{1/2}} + o_p(1).
\end{aligned} \tag{47}$$

Let $L_4(n) = \sum_{k=-n}^{n}\gamma_{H_4}(k)$. From (39), it follows that

$$\begin{aligned}
\gamma_{H_4}(k) &= E\left(H_4(y_{t+k})H_4(y_t)\right) \\
&= 4!(\gamma_y(k))^4 \\
&\sim 24\left(ck^{2d-1}\right)^4 \\
&\sim \frac{24c^4}{k}.
\end{aligned}$$

Hence $L_4(n) \sim 48c^4\log(n)$. Since $L_4(n)$ is a slowly varying function and $L_4(n) \to \infty$ as $n \to \infty$, application of Theorem 6 of Giraitis and Surgailis (1985) leads to

$$\frac{1}{(n\log n)^{1/2}}H_{4,n} \xrightarrow{\mathcal{L}} \left(48c^2\right)^{1/2} N(0,1). \tag{48}$$

Combining the last convergence with (47) yields (29).

3. If $0 < d < 3/8$ then from previous results we deduce the following:

- If $0 < d < 1/4$ then $H_{1,n} = O_p\left(n^{d+1/2}\right)$, $H_{2,n} = O_p\left(n^{1/2}\right)$ and $H_{3,n} = O_p\left(n^{1/2}\right)$.
- If $d = 1/4$ then $H_{1,n} = O_p\left(n^{3/4}\right)$, $H_{2,n} = O_p\left((n\log n)^{1/2}\right)$ and $H_{3,n} = O_p\left(n^{1/2}\right)$.
- If $1/4 < d < 1/3$ then $H_{1,n} = O_p\left(n^{d+1/2}\right)$, $H_{2,n} = O_p\left(n^{2d}\right)$ and $H_{3,n} = O_p\left(n^{1/2}\right)$.

- If $d = \frac{1}{3}$ then $H_{1,n} = O_p\left(n^{5/6}\right)$, $H_{2,n} = O_p\left(n^{2/3}\right)$ and $H_{3,n} = O_p\left((n\log n)^{1/2}\right)$.
- If $1/3 < d < 3/8$ then $H_{1,n} = O_p\left(n^{d+1/2}\right)$, $H_{2,n} = O_p\left(n^{2d}\right)$ and $H_{3,n} = O_p\left(n^{3d-1/2}\right)$.

This implies that for all d such that $0 < d < 3/8$,

$$n^{1/2}\widetilde{K} = n^{1/2}\left(\frac{H_{4,n}}{n} - \frac{4}{n^2}H_{1,n}H_{3,n} + \frac{12}{n^3}H_{1,n}^2H_{2,n} - \frac{3}{n^2}H_{2,n}^2 - \frac{6}{n^4}H_{1,n}^4\right)$$
$$= \frac{H_{4,n}}{n^{1/2}} + o_p(1). \qquad (49)$$

Moreover, if $0 < d < 3/8$ then $\sigma_{H_4}^2 = \sum_{k\in Z}\gamma_{H_4}(k) \sim 4!c^4\sum_{k\in Z}\left(k^{2d-1}\right)^4 < \infty$, hence from Theorem 5 of Giraitis and Surgailis (1985) we have

$$\frac{H_{4,n}}{n^{1/2}} \overset{\mathcal{L}}{\to} \sigma_{H_4}N(0,1). \qquad (50)$$

Therefore, the convergence (30) follows from (49) and (50).

References

Ajmi AN, Ben Nasr A, Boutahar M (2008) Seasonal nonlinear long memory model for the US inflation rates. Comput Econ 31(3):243–254

Anderson TW, Darling DA (1954) A test of goodness of fit criteria based on stochastic processes. J Stat Assoc 49:765–769

Bai J, Ng S (2005) Tests for skewness, kurtosis, and normality for time series data. J Bus Econ Stat 23(1): 49–60

Baillie RT, Bollerslev T, Mikkelsen HO (1996) Fractionally integrated generalized autoregressive conditional heteroskedasticity. J Econ 74:3–30

Beran J (1994) Statistics for long-memory processes. Chapman & Hall, London

Boumahdi M (1996) Blind identification using the kurtosis with applications to field data. Signal Process 48(3):205–216

Boutahar M, Marimoutou V, Nouira L (2007) Estimation methods of the long memory parameter: Monte Carlo analysis and application. J Appl Stat 34(3):261–301

Bowman KO, Shenton LR (1975) Omnibus test contours for departures from normality based on $\sqrt{b_1}$ and b_2. Biometrika 62(2):243–250

Breidt GJ, Crato N, Lima PD (1998) The detection and estimation of long memory in stochastic volatility. J Econom 83:325–348

Brys G, Hubert M, Struyf A (2004) A robustification of the Jarque–Bera test of normality. COMPSTAT 2004.s

Caporin M (2003) Identification of long memory in GARCH models. Stat Methods Appl 12(2):133–151

D'Agostino RB (1972) Small sample probability points for the D test of normality (complete samples). Biometrika 59:219–221

Delong JB, Summers LH (1985) Are business cycle symmetrical. In: Americain business cycle: continuity and change. University of Chicago Press, Chicago, pp 166–178

Dobrushin RL, Major P (1979) Non-central limits theorem for non-linear functionals of Gaussian fields. Z Wahrsch verw Gebiete 50:27–52

Doukhan P, Oppenheim G, Taqqu MS (2003) Theory and applications of long-range dependence. Birkhäuser, Boston

Fiorentini G, Sentana E, Calzolari G (2004) On the validity of Jarque–Bera normality test in conditionally heteroskedastic dynamic regression models. Econ Lett 83:307–312

Forsberg L, Ghysels E (2007) Why do absolute returns predict volatility so well? J Financial Econom 5:31–67

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

Gel YR, Gastwirth JL (2008) A robust modification of the Jarque–Bera test of normality. Econ Lett 99:30–32

Giraitis L, Surgailis D (1985) CLT and other limit theorems for functionals of Gaussian sequences. Z Wahrsch Verw Gebiete 70:191–212

Hassler U, Wolters J (1995) Long memory in inflation rates: international evidence. J Bus Econ Stat 13:37–45

Heinz M (2001) On the kurtosis of digitally modulated signals with timing offsets. In: Third IEEE signal processing workshop on signal processing. Advance in wireless communications, Taiwan, 20–23, March 2001

Hosking JRM (1996) Asymptotic distribution of the sample mean, autocovariances, autocorrelations of long-memory time series. J Econom 73:261–284

Jarque CM, Bera AK (1980) Efficient tests for normality, homoskedasticity and serial independence of regression residuals. Econ Lett 6:255–259

Jarque CM, Bera AK (1987) A test for normality of observations and regression residuals. Int Stat Rev 55:163–172

Kolmogorov AN (1933) Sulla determinazione empirica di une legge di distribuzione. Giornale dell'Intituto Italiano degli Attuari 4:83–91

Lobato IN, Velasco C (2004) A simple test of normality for time series. Econom Theory 20:671–689

Lomnicki Z (1961) Tests for departure from normality in the case of linear stochastic processes. Metrika 4(1):37–62

Major P (1981) Multiple Wiener–Itô integrals, lecture notes in mathematics, vol 849. Springer, New York

Pearson K (1900) On the criterion that a given system deviations from the probable in the case of a correlated system of variables in such that it can be reasonable supposed to have arisen from random sampling. Philosophical Magazine, V, 50, 157

Robinson PM (2003) Times series with long memory. Advanced texts in econometrics. Oxford University Press, Oxford

Shapiro SS, Wilk MB (1965) An analysis of variance test for normality. Biometrika 52:591–611

Taqqu MS (1975) Weak convergence to fractional brownian motion and to the Rosenblatt process. Z Wahrsch verw Gebiete 31:287–302

Yazici B, Yolacan S (2007) A comparison of various tests of normality. J Stat Comput Simul 77(2):175–183