Econometrics and Statistics 2 (2017) 50–60

Contents lists available at ScienceDirect

Econometrics and Statistics

journal homepage: www.elsevier.com/locate/ecosta

A distance test of normality for a wide class of stationary processes

Zacharias Psaradakis[a,*], Marián Vávra[b]

[a] Department of Economics, Mathematics and Statistics, Birkbeck, University of London, Malet Street, London WC1E 7HX, United Kingdom
[b] The Governor and Research Departments, National Bank of Slovakia, Slovakia

ARTICLE INFO

Article history:
Received 10 August 2016
Revised 23 November 2016
Accepted 23 November 2016
Available online 8 December 2016

Keywords:
Distance test
Fractionally integrated process
Sieve bootstrap
Normality

ABSTRACT

A distance test for normality of the one-dimensional marginal distribution of stationary fractionally integrated processes is considered. The test is implemented by using an autoregressive sieve bootstrap approximation to the null sampling distribution of the test statistic. The bootstrap-based test does not require knowledge of either the dependence parameter of the data or of the appropriate norming factor for the test statistic. The small-sample properties of the test are examined by means of Monte Carlo experiments. An application to real-world data is also presented.

1. Introduction

Testing whether a sample of observations comes from a Gaussian distribution is a problem that has attracted a great deal of attention over the years. This is not perhaps surprising in view of the fact that normality is a common maintained assumption in a wide variety of statistical procedures, including estimation, inference and forecasting procedures. In model building, a test for normality is often a useful diagnostic for assessing whether a particular type of stochastic model may provide an appropriate characterization of the data (for instance, non-linear models are unlikely to be an adequate approximation to a time series having a Gaussian one-dimensional marginal distribution). Normality tests may also be useful in evaluating the validity of different hypotheses and models to the extent that the latter rely on or imply Gaussianity, as is the case, for example, with some option pricing, asset pricing, and dynamic stochastic general equilibrium models found in the economics and finance literature. Other examples where normality or otherwise of the marginal distribution is of interest, include value-at-risk calculations (e.g., Cotter, 2007) and copula-based modeling for multivariate time series with the marginal distribution and the copula function being specified separately. Kilian and Demiroglu (2000) and Bontemps and Meddahi (2005) give further examples from economics, finance and econometrics where testing for normality is of interest.

Although most of the voluminous literature on the subject of testing for univariate normality has focused on the case of independent and identically distributed (i.i.d.) observations (see Thode, 2002, for an extensive review), a small number of tests which are valid for dependent data have also been considered. The latter include tests based on the bispectrum (e.g., Hinich, 1982; Nusrat and Harvill, 2008; Berg et al., 2010), the characteristic function (Epps, 1987), moment conditions implied by Stein's characterization of the Gaussian distribution (Bontemps and Meddahi, 2005), and classical measures of

* Corresponding author.
E-mail address: z.psaradakis@bbk.ac.uk (Z. Psaradakis).

http://dx.doi.org/10.1016/j.ecosta.2016.11.005
2452-3062/

skewness and kurtosis involving standardized third and fourth central moments (Lobato and Velasco, 2004; Bai and Ng, 2005). A feature shared by these tests is that they all rely on asymptotic results obtained under dependence conditions which typically require the autocovariances of the data to decay towards zero, as the lag parameter goes to infinity, sufficiently fast to be (at least) absolutely summable. It has long been recognized, however, that such short-range dependence conditions may not accord well with the slowly decaying autocovariances of many observed time series.

The purpose of this paper is to discuss a test for normality which may be used in the presence of not only short-range dependence but also long-range dependence and antipersistence. The defining characteristic of stochastic processes with such dependence structures is that their autocovariances decay to zero as a power of the lag parameter and, in the case of long-range dependence, slowly enough to be non-summable. Models that allow for long-range dependence have been found to be useful for modeling data occurring in fields as diverse as economics, geophysics, hydrology, meteorology, and telecommunications; a summary of some of the empirical evidence on long-range dependence can be found in the collection of papers in Doukhan et al. (2003).

The normality test we consider here is based on the Anderson–Darling distance statistic involving the weighted quadratic distance of the empirical distribution function of the data from a Gaussian distribution function (Anderson and Darling, 1952). Unlike tests based on measures of skewness and kurtosis, which can only detect deviations from normality that are reflected in the values of such measures, normality tests based on the empirical distribution function are known to be consistent against any fixed non-Gaussian alternative. The Anderson–Darling test also fares well in small-sample power comparisons for i.i.d. data relatively to the popular correlation test of Shapiro and Wilk (1965) and the moment-based tests of Bowman and Shenton (1975) and Jarque and Bera (1987) (see, e.g., Stephens, 1974, and Thode, 2002, Ch. 7). Furthermore, the Anderson–Darling test is superior, in terms of asymptotic relative efficiency, to distance tests such as those based on the (unweighted) Cramér–von Mises and Kolmogorov–Smirnov statistics (Koziol, 1986; Arcones, 2006).

Our analysis extends earlier work by considering the case of correlated data from stationary (in the strict sense) fractionally integrated processes, which may be short-range dependent, long-range dependent or antipersistent depending on the value of their dependence parameter. Unfortunately, however, inference based on conventional large-sample approximations is anything but straightforward in such a setting because the weak limit of the null distribution of the test statistic, as well as the appropriate norming factor, depend on the unknown dependence parameter of the data and on the particular estimators of location and scale parameters that are used in the construction of the test statistic.

As a practical way of overcoming these difficulties, we propose to use the bootstrap to estimate the null sampling distribution of the Anderson–Darling distance statistic and thus obtain estimates of P-values and/or critical values for a normality test. Our approach relies on the autoregressive sieve bootstrap, which is based on the idea of approximating the data-generating mechanism by an autoregressive sieve, that is a sequence of autoregressive models that increase in order as the sample size increases without bound (Kreiss, 1992; Bühlmann, 1997). The bootstrap-based normality test is easy to implement and requires knowledge (or estimation) of neither the value of the dependence parameter of the data nor of the appropriate norming factor for the test statistic. Furthermore, the bootstrap scheme is the same under short-range dependence, long-range dependence and antipersistence.

We note that Beran and Ghosh (1991) and Boutahar (2010) obtained results relating to the asymptotic behavior of moment-based and distance-based statistics under long-range dependence, but did not discuss how operational tests for normality might be constructed. To the best of our knowledge, the problem of developing an operational normality test which is valid for data that are neither independent nor short-range dependent has not been tackled in the existing literature.

The plan of the paper is as follows. Section 2 formulates the problem and introduces the test statistic and the class of stochastic processes of interest. Section 3 discusses the autoregressive sieve bootstrap approach to implementing the distance test of normality. Section 4 examines the finite-sample properties of the proposed test by means of Monte Carlo experiments. Section 5 presents an application to a set of U.S. economic and financial time series. Section 6 summarizes and concludes.

2. Assumptions and test statistic

Suppose $\mathbf{X}_n := \{X_1, X_2, \ldots, X_n\}$ are consecutive observations from a stationary stochastic process $\mathbf{X} := \{X_t\}_{t\in\mathbb{Z}}$ satisfying

$$X_t - \mu = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \qquad t \in \mathbb{Z}, \tag{1}$$

for some $\mu \in \mathbb{R}$, where $\{\psi_j\}_{j\in\mathbb{Z}^+}$ is a square-summable sequence of real numbers (with $\psi_0 = 1$) and $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is a sequence of i.i.d., real-valued, zero-mean random variables with variance $\sigma^2 \in (0, \infty)$. The objective is to test the null hypothesis that the one-dimensional marginal distribution of $\mathbf{X}$ is Gaussian,

$$\mathcal{H}_0: \; F(\mu + \gamma_0^{1/2} x) - \Phi(x) = 0 \quad \text{for all } x \in \mathbb{R}, \tag{2}$$

where $\gamma_k := \mathrm{Cov}(X_k, X_0) = \sigma^2 \sum_{j=0}^{\infty} \psi_{j+|k|}\psi_j$ for $k \in \mathbb{Z}$, F is the distribution function of X_0, and Φ denotes the standard normal distribution function. Notice that (2) holds if ε_0 is normally distributed. Conversely, (2) implies normality of the distribution of ε_0, which in turn implies Gaussianity of the causal linear process $\mathbf{X}$ (see, e.g., Rosenblatt, 2000, Sec. 1.1).

To allow for different types of dependence, it will be maintained throughout that the transfer function $\psi(z) := \sum_{j=0}^{\infty}\psi_j z^j$, $z \in \mathbb{C}$, associated with (1) satisfies

$$\psi(z) = (1-z)^{-d}\delta(z), \qquad |z| < 1, \tag{3}$$

for some real $|d| < \frac{1}{2}$, where $\delta(z) := \sum_{j=0}^{\infty}\delta_j z^j$, $z \in \mathbb{C}$, and $\{\delta_j\}_{j\in\mathbb{Z}^+}$ is an absolutely summable sequence of real numbers such that $\delta(0) = 1$ and $\delta(1) \neq 0$. Under (1) and (3), $\mathbf{X}$ is a fractionally integrated process with dependence (memory/fractional differencing) parameter d. Using the power series expansion

$$(1-z)^{-d} = \sum_{j=0}^{\infty}\frac{\Gamma(d+j)}{\Gamma(d)\Gamma(1+j)}z^j, \qquad |z| < 1,\ d \neq 0,$$

where Γ denotes the gamma function, it is not difficult to see that, for $d \neq 0$,

$$\psi_j = \sum_{s=0}^{j}\frac{\delta_{j-s}\Gamma(d+s)}{\Gamma(d)\Gamma(1+s)}, \qquad j \geq 1,$$

and so $\psi_j \sim \{\delta(1)/\Gamma(d)\}j^{d-1}$ as $j \to \infty$, under the additional assumption that $j^{1-d}\delta_j \to 0$ as $j \to \infty$ if $0 < d < \frac{1}{2}$ (Hassler and Kokoszka, 2010); the tilde signifies that the limiting value of the quotient of the left-hand side by the right-hand side is 1. Hence, since $\gamma_k \sim c_\gamma |k|^{2d-1}$ as $|k| \to \infty$ for $d \neq 0$, with $c_\gamma := \{\sigma\delta(1)\}^2\Gamma(1-2d)/\{\Gamma(d)\Gamma(1-d)\}$, $\sum_{k=-\infty}^{\infty}\gamma_k = \infty$ for $0 < d < \frac{1}{2}$ and $\mathbf{X}$ is long-range dependent; if $-\frac{1}{2} < d < 0$, then $\sum_{k=-\infty}^{\infty}\gamma_k = 0$ and $\mathbf{X}$ is antipersistent. Short-range dependence corresponds to $d = 0$, with $\sum_{k=-\infty}^{\infty}\gamma_k = \{\sigma\delta(1)\}^2$. The class of stochastic processes defined by (1) and (3) is rich enough to include a wide range of processes having slowly decaying autocovariances. A well-known example are autoregressive fractionally integrated moving average (ARFIMA) processes with $\delta(z) = \vartheta(z)/\varphi(z)$, $\vartheta(z)$ and $\varphi(z)$ being relatively prime polynomials of finite degree with $\varphi(z) \neq 0$ for $|z| \leq 1$ (see, e.g., Palma, 2007, Sect. 3.2).

The test for the hypothesis in (2) considered here is based on the Anderson–Darling distance statistic

$$\mathcal{A} := \int_{-\infty}^{\infty}\frac{\{\hat{F}(\bar{X} + \hat{\gamma}_0^{1/2}x) - \Phi(x)\}^2}{\Phi(x)\{1-\Phi(x)\}}\,\mathrm{d}\Phi(x), \tag{4}$$

where $\hat{F}(x) := n^{-1}\sum_{t=1}^{n}\mathbb{I}(X_t \leq x)$, $x \in \mathbb{R}$, is the empirical distribution function of $\mathbf{X}_n$, $\bar{X} := n^{-1}\sum_{t=1}^{n}X_t$, $\hat{\gamma}_k := n^{-1}\sum_{t=1}^{n-|k|}(X_{t+|k|} - \bar{X})(X_t - \bar{X})$ for $|k| < n$, and $\mathbb{I}$ denotes the indicator function. Note that $\mathcal{A}$ may be expressed as

$$\mathcal{A} = -1 - n^{-2}\sum_{t=1}^{n}(2t-1)[\log\Phi(Y_t) + \log\{1 - \Phi(Y_{n-t+1})\}], \tag{5}$$

where $Y_t := \hat{\gamma}_0^{-1/2}(X_{n:t} - \bar{X})$ and $X_{n:t}$ is the t-th order statistic from $\mathbf{X}_n$ (cf. Anderson and Darling, 1954).

As is well known, the asymptotic null distribution of distance statistics such as $\mathcal{A}$ is closely related to the weak limit, as n tends to infinity, of (a suitably normalized version of) the random function $\hat{K}(x) := \hat{F}(x) - \Phi(\hat{\gamma}_0^{-1/2}(x - \bar{X}))$, $x \in \mathbb{R}$ (see, e.g., Shorack and Wellner, 1986, Ch. 5, for the case of i.i.d. data). However, unless one is dealing with the classical problem of testing a simple null hypothesis (μ and γ_0 known) for i.i.d. data, inference is complicated by the presence of both estimated parameters in $\hat{K}$ and dependence in $\mathbf{X}_n$. To complicate matters further in our setup, the norming factor needed in order to obtain weak convergence of $\hat{K}$, and hence of the distribution of $\mathcal{A}$, as well as the weak limits themselves, depend on the value of the dependence parameter d and on the estimators of μ and γ_0 used in the construction of $\hat{K}$ (see Beran and Ghosh, 1991; Ho, 2002; Kulik, 2009). For example, when ε_0 is normally distributed, the random function $K(x) := \hat{F}(x) - \Phi(\gamma_0^{-1/2}(x - \mu))$, $x \in \mathbb{R}$, converges weakly to a non-degenerate Gaussian process at the usual $\sqrt{n}$ rate for $d = 0$ (Doukhan and Surgailis, 1998) and to a semi-deterministic Gaussian process at the slower rate $n^{(1-2d)/2}$ for $0 < d < \frac{1}{2}$ (Dehling and Taqqu, 1989; Giraitis and Surgailis, 1999). Replacing μ and γ_0 by estimates improves the rate of convergence under long-range dependence, with $\hat{K}$ converging at rate $\sqrt{n}$ for $0 \leq d < \frac{1}{3}$ (Beran and Ghosh, 1991). This dependence of the appropriate norming factor and of the weak limit of $\hat{K}$ on the unknown value of the parameter d, combined with the complicated covariance structure of the relevant limiting processes, make inference based on conventional large-sample asymptotics for $\mathcal{A}$ extremely cumbersome. It is worth remarking that similar difficulties also arise in the case of moment-based skewness and kurtosis statistics (cf. Boutahar, 2010; Ho, 2002).

As a practical way of circumventing the problems mentioned above, we propose to use an autoregressive sieve bootstrap procedure to obtain P-values and/or critical values for the normality test based on $\mathcal{A}$. The principal advantage of the sieve bootstrap is that it can be used to approximate the sampling properties of $\mathcal{A}$ without knowledge or estimation of the dependence parameter. Moreover, because bootstrap approximations are constructed from replicates of $\mathcal{A}$, there is no need to derive analytically, nor to make assumptions about, the appropriate norming factor for $\mathcal{A}$ or its asymptotic null distribution.

3. Autoregressive sieve bootstrap approximation

The autoregressive sieve bootstrap is motivated by the observation that, under (1), (3) and the additional assumption that $\delta(z) \neq 0$ for $|z| \leq 1$, $\mathbf{X}$ admits the representation

$$\sum_{j=0}^{\infty} \phi_j (X_{t-j} - \mu) = \varepsilon_t, \qquad t \in \mathbb{Z}, \tag{6}$$

for a square-summable sequence of real numbers $\{\phi_j\}_{j\in\mathbb{Z}^+}$ (with $\phi_0 = 1$) such that $\phi(z) := \sum_{j=0}^{\infty} \phi_j z^j = (1-z)^d/\delta(z)$ for $|z| < 1$ (with $\phi_j \sim \{\delta(1)\Gamma(-d)\}^{-1} j^{-d-1}$ as $j \to \infty$ when $d \neq 0$). The idea is to approximate (6) by a finite-order autoregressive model and use this as the basis of a semi-parametric bootstrap scheme. If the order of the autoregressive approximation is allowed to increase simultaneously with n at an appropriate rate, the distribution of the process in (6) will be matched asymptotically (cf. Kreiss, 1992; Bühlmann, 1997; Kapetanios and Psaradakis, 2006; Poskitt, 2008).

It is important to point out that, as discussed in Poskitt (2007), the autoregressive representation (6) provides a meaningful approximation even if $\delta(z)$ has zeros in the unit disc $|z| < 1$. In this case, the transfer function $\phi(z)$ associated with (6) may be viewed as arising, not from the inversion of $\psi(z)$, but as the limit of $\sum_{j=0}^{p} \phi_{pj} z^j$ ($\phi_{p0} := 1$) as p tends to infinity, where, for any integer $p > 0$, $(-\phi_{p1}, \ldots, -\phi_{pp})$ are the coefficients of the best (in the mean-square sense) linear predictor of X_0 based on $\{X_{-1}, \ldots, X_{-p}\}$. Since $\gamma_0 > 0$ and $\gamma_k \to 0$ as $|k| \to \infty$ under (1) and (3), $(\phi_{p1}, \ldots, \phi_{pp})$ are uniquely determined as the solution of the set of equations $\sum_{j=0}^{p} \phi_{pj} \gamma_{k-j} = 0$ $(k = 1, \ldots, p)$ (Brockwell and Davis, 1991, Corollary 5.1.1), and are such that $\sum_{j=0}^{p} \phi_{pj} z^j \neq 0$ for $|z| \leq 1$ and $\mathrm{Var}[\sum_{j=0}^{p} \phi_{pj}(X_{-j} - \mu)] \to \sigma^2$ as $p \to \infty$.

The bootstrap procedure used to approximate the sampling properties of the statistic $\mathcal{A}$ under the normality hypothesis is as follows. For an integer $p > 0$ (chosen as a function of n so that $p^{-1} + n^{-1}p \to 0$ as $n \to \infty$), let $(\hat{\phi}_{p1}, \ldots, \hat{\phi}_{pp})$ and $\hat{\sigma}_p$ be estimators (based on $\mathbf{X}_n$) of the coefficients and the noise standard deviation, respectively, of an autoregressive model of order p for $X_t - \bar{X}$. Bootstrap replicates $\mathbf{X}^* := \{X_t^*\}_{t\in\mathbb{Z}}$ of $\mathbf{X}$ are then defined via the recursion

$$\sum_{j=0}^{p} \hat{\phi}_{pj} (X_{t-j}^* - \bar{X}) = \hat{\sigma}_p \varepsilon_t^*, \qquad t \in \mathbb{Z}, \tag{7}$$

where $\hat{\phi}_{p0} := 1$ and $\{\varepsilon_t^*\}_{t\in\mathbb{Z}}$ are i.i.d. standard normal random variables (independent of $\mathbf{X}_n$). Finally, the bootstrap analogue $\mathcal{A}^*$ of $\mathcal{A}$ is defined as in (4) but with $\mathbf{X}_n^* := \{X_1^*, X_2^*, \ldots, X_n^*\}$ replacing $\mathbf{X}_n$. The conditional distribution of $\mathcal{A}^*$, given $\mathbf{X}_n$, constitutes the sieve bootstrap approximation to the null sampling distribution of $\mathcal{A}$.

It is worth noting that, by requiring ε_t^* in (7) to be Gaussian, $\mathbf{X}^*$ is constructed in a way which reflects the normality hypothesis under test even though $\mathbf{X}$ may not satisfy (2). This is important for ensuring that the bootstrap test has reasonable power against departures from normality (cf. Hall and Wilson, 1991; Lehmann and Romano, 2005, Sect. 15.6). The estimator $(\hat{\phi}_{p1}, \ldots, \hat{\phi}_{pp}, \hat{\sigma}_p)$ used in (7) to define $\mathbf{X}^*$ may be the Yule–Walker estimator or any other asymptotically equivalent estimator (e.g., the least-squares estimator).

Recalling that $\mathcal{A}$ may be expressed as in (5), consistency of the sieve bootstrap estimator of the null sampling distribution of $\mathcal{A}$ follows from Lemma 1, Theorem 2 and Remark 2 of Poskitt (2008) under a suitable assumption about the rate of increase of p. More specifically, let $\rho(H, H^*) := \{\int_0^1 |H^{-1}(u) - H^{*-1}(u)|^2 \mathrm{d}u\}^{1/2}$ stand for the Mallows–Wasserstein distance between the distribution function H of $\mathcal{A}$ and the conditional distribution function $H^*$of $\mathcal{A}^*$ given $\mathbf{X}_n$ (where $g^{-1}(u) := \inf\{x : g(x) \geq u\}$ for any non-decreasing function g). Then, if $\mathbf{X}$ satisfies (1) and (3), the distribution of ε_0 is Gaussian, and $p \to \infty$ and $(\log n)^{-\nu} p = O(1)$ as $n \to \infty$ for some $\nu \geq 1$, we have $\rho(H, H^*) \to 0$ with probability 1 as $n \to \infty$. We note that, although Poskitt (2008) considers a bootstrap scheme in which ε_t^* in (7) is drawn from the empirical distribution of the residuals $\hat{\varepsilon}_t := \sum_{j=0}^{p} \hat{\phi}_{pj}(X_{t-j} - \bar{X})$ $(t = p+1, \ldots, n)$, standardized to have mean 0 and variance 1, it is not difficult to see that the arguments in the proof of his Theorem 2 go through with little or no change when ε_t and ε_t^* are Gaussian.

The bootstrap estimator of the P-value for a test that rejects for large values of $\mathcal{A}$ is $P_{\mathcal{A}}^* := 1 - H^*(\mathcal{A})$, and so normality is rejected at a given level of significance $\alpha \in (0, 1)$ if $P_{\mathcal{A}}^* \leq \alpha$. Since convergence with respect to ρ implies weak convergence (Bickel and Freedman, 1981, Lemma 8.3), the bootstrap P-value $P_{\mathcal{A}}^*$ is asymptotically equivalent to the P-value based on the null sampling distribution of $\mathcal{A}$ under the conditions stated above. While H^* is unknown, an approximation (of any desired accuracy) can be obtained by Monte Carlo simulation. Specifically, if $(\mathcal{A}_1^*, \ldots, \mathcal{A}_m^*)$ are copies of $\mathcal{A}^*$, obtained from m independent bootstrap pseudo-samples $\mathbf{X}_n^*$ from (7), then the empirical distribution function of $(\mathcal{A}_1^*, \ldots, \mathcal{A}_m^*)$ provides an approximation to H^*. Hence, $P_{\mathcal{A}}^*$ may be approximated by $\hat{P}_{\mathcal{A}}^* := m^{-1} \sum_{i=1}^{m} \mathbb{I}(\mathcal{A}_i^* > \mathcal{A})$, so that $\hat{P}_{\mathcal{A}}^* \to P_{\mathcal{A}}^*$ with probability 1 as $m \to \infty$. A bootstrap critical value of nominal level α for $\mathcal{A}$ is given by $H^{*-1}(1-\alpha)$, which may be approximated by $\inf\{x \in \mathbb{R} : m^{-1} \sum_{i=1}^{m} \mathbb{I}(\mathcal{A}_i^* \leq x) \geq 1 - \alpha\}$.

In the implementation of the bootstrap procedure in practice, replicates $\mathbf{X}_n^*$ may be obtained according to (7) by setting $X_{-p+1}^* = \cdots = X_0^* = \bar{X}$, generating $n + b$ replicates X_t^*, $t \geq 1$, for some large integer $b > 0$, and then discarding the initial b replicates to eliminate start-up effects (this procedure, with $b = 500$, is used in the sequel). The order of the autoregressive sieve may be selected as the minimizer of Akaike's information criterion $\mathrm{AIC}(p) := \log \hat{\sigma}_p^2 + 2n^{-1}p$ over $1 \leq p \leq p_{\max}$ for some suitable maximal order $p_{\max}$. Under mild regularity conditions, a data-dependent choice of p based on AIC is asymptotically efficient, in the sense defined by Shibata (1980), for all $|d| < \frac{1}{2}$ (Poskitt, 2007, Theorem 9); furthermore, it satisfies, with probability 1, the growth conditions required for the asymptotic validity of the sieve bootstrap as long as $p_{\max}$ grows to infinity with n so that $(\log n)^{-h} p_{\max}$ is eventually bounded for some $h \geq 1$ (Psaradakis, 2016, Proposition 2). Alternative criteria for order selection that may be used include, among many others, the Bayesian information criterion $\mathrm{BIC}(p) := \log \hat{\sigma}_p^2 + n^{-1} p \log n$ and Mallows' criterion $\mathrm{MC}(p) := \hat{\sigma}_\infty^{-2} \hat{\sigma}_p^2 - 1 + 2n^{-1} p$, where $\hat{\sigma}_\infty^2 := 2\pi \exp(C + n_0^{-1} \sum_{i=1}^{n_0} \log I_i)$. Here, $I_i := (2\pi n)^{-1} |\sum_{t=1}^{n} X_t \exp(-\omega_i t \sqrt{-1})|^2$ is the periodogram ordinate of $\mathbf{X}_n$ at the Fourier frequency $\omega_i := 2\pi i/n$, $C = -\Gamma'(1)$ is Euler's constant, and $n_0 := \lfloor (n-1)/2 \rfloor$, $\lfloor \cdot \rfloor$ denoting the greatest-integer function. The quantity $\hat{\sigma}_\infty^2$ is a non-parametric estimator of σ^2 motivated by the observation that $\sigma^2 = 2\pi \exp\{(2\pi)^{-1} \int_{-\pi}^{\pi} \log \Lambda(\omega) d\omega\}$, where Λ is the spectral density of $\mathbf{X}$. Like AIC, MC is an asymptotically efficient selection criterion.

We conclude this section by remarking that the linear structure imposed by (1) and (3) may arguably be considered as somewhat restrictive. However, since non-linear processes with a Gaussian one-dimensional marginal distribution appear to be a rarity (see, e.g., Tong, 1990, Sect. 4.2), the focus on linear dependence is not perhaps unjustifiable when the objective is to test for normality. In any case, the results of Bickel and Bühlmann (1997) suggest that linearity may not be too onerous a requirement in the sense that the closure (with respect to the total variation metric) of the class of linear processes is quite large; roughly speaking, for any stationary non-linear process, there exists another process in the closure of linear processes having identical sample paths with probability exceeding 0.36. This suggests that the autoregressive sieve bootstrap is likely to yield reasonably good approximations within a class of processes larger than that associated with (1) or (6).

4. Simulation study

In this section, we present and discuss the results of a simulation study examining the small-sample properties of the distance-based test of normality under various data-generating mechanisms.

4.1. Experimental design and simulation

In the first set of experiments, we examine the performance of the test based on $\mathcal{A}$ under different types of dependence by considering artificial data generated according to the ARFIMA process

$$\text{M1: } (1–0.7L)X_t = (1–0.3L)(1-L)^{-d}\varepsilon_t, \quad d \in \{-0.4, -0.25, 0, 0.25, 0.4\},$$

where L denotes the lag operator and $\{\varepsilon_t\}$ are i.i.d. random variables. (Note that the stationary solution of M1 satisfies (1) with $\mu = 0$, $\psi_1 = 0.4 + \zeta_1$ and $\psi_j = 0.7\psi_{j-1} - 0.3\zeta_{j-1} + \zeta_j$ for $j \geq 2$, where $\zeta_j := \Gamma(d+j)/\{\Gamma(d)\Gamma(1+j)\}$ for $j \geq 1$ and $d \neq 0$.) The distribution of ε_t is normal (labeled N in what follows), lognormal (labeled LN), or a member of the family of generalized lambda distributions having quantile function $F_\varepsilon^{-1}(u) = \lambda_1 + \lambda_2^{-1}\{u^{\lambda_3} - (1-u)^{\lambda_4}\}$, $0 < u < 1$; the parameter values used in the experiments are taken from Bai and Ng (2005) and can be found in Table 1. The distributions S1–S3 are symmetric, whereas A1–A3 and LN are asymmetric. Throughout this section, $\{\varepsilon_t\}$ are standardized to have mean 0 and variance 1.

In the second set of experiments, we assess the robustness of the test based on $\mathcal{A}$ with respect to departures from the linearity assumption underlying the autoregressive sieve bootstrap by using artificial data from the models

$$\text{M2: } X_t = 0.5X_{t-1} - 0.3X_{t-1}\varepsilon_{t-1} + \varepsilon_t,$$
$$\text{M3: } X_t = (0.9X_{t-1} + \varepsilon_t)\mathbb{I}(|X_{t-1}| \leq 1) - (0.3X_{t-1} - 2\varepsilon_t)\mathbb{I}(|X_{t-1}| > 1),$$
$$\text{M4: } X_t = 0.5X_{t-1} + \varepsilon_t\varepsilon_{t-1}.$$

M2 is a bilinear model, M3 is a threshold autoregressive model, and M4 is an autoregressive model with one-dependent, uncorrelated noise. In all three cases, $\{X_t\}$ is short-range dependent and does not admit the representation (1) or (6). Furthermore, the distribution of X_t is non-Gaussian even if ε_t is normally distributed. In addition, we consider artificial data generated according to

$$\text{M5: } X_t = \Phi^{-1}(F_\xi(\xi_t)), \quad \xi_t = \beta|\xi_{t-1}| + \varepsilon_t, \quad \beta \in \{-0.5, 0.5\},$$

Table 1
Parameters of a generalized lambda distribution and selected descriptive statistics.

	λ_1	λ_2	λ_3	λ_4	Skewness	Kurtosis
N	–	–	–	–	0.0	3.0
S1	0.000000	−1.000000	−0.080000	−0.080000	0.0	6.0
S2	0.000000	−0.397912	−0.160000	−0.160000	0.0	11.6
S3	0.000000	−1.000000	−0.240000	−0.240000	0.0	126.0
A1	0.000000	−1.000000	−0.007500	−0.030000	1.5	7.5
A2	0.000000	−1.000000	−0.100900	−0.180200	2.0	21.1
A3	0.000000	−1.000000	−0.001000	−0.130000	3.2	23.8
LN	–	–	–	–	6.2	113.9

Table 2
Rejection frequencies of $\mathcal{A}$ test under M1.

	$n=100$					$n=200$					$n=500$				
Distr.\d	−0.40	−0.25	0	0.25	0.40	−0.40	−0.25	0	0.25	0.40	−0.40	−0.25	0	0.25	0.40
N	0.06	0.05	0.05	0.05	0.06	0.05	0.05	0.05	0.05	0.06	0.05	0.06	0.05	0.05	0.07
S1	0.48	0.44	0.25	0.12	0.12	0.75	0.70	0.41	0.14	0.09	0.98	0.97	0.75	0.16	0.11
S2	0.72	0.73	0.44	0.21	0.15	0.94	0.92	0.66	0.23	0.13	1.00	1.00	0.97	0.33	0.15
S3	0.88	0.85	0.61	0.31	0.21	0.99	0.99	0.88	0.37	0.21	1.00	1.00	1.00	0.58	0.24
A1	0.97	0.95	0.77	0.25	0.15	1.00	1.00	0.98	0.42	0.16	1.00	1.00	1.00	0.71	0.23
A2	0.90	0.86	0.64	0.29	0.20	0.99	0.99	0.91	0.43	0.21	1.00	1.00	1.00	0.69	0.32
A3	1.00	1.00	1.00	0.59	0.29	1.00	1.00	1.00	0.83	0.33	1.00	1.00	1.00	0.99	0.54
LN	1.00	1.00	1.00	0.78	0.41	1.00	1.00	1.00	0.94	0.54	1.00	1.00	1.00	1.00	0.82

where F_ξ is the distribution function of ξ_t. When ε_t is normally distributed, the threshold autoregressive process $\{\xi_t\}$ is stationary with

$$F_\xi(x) = \left\{2(1-\beta^2)/\pi\right\}^{1/2} \int_{-\infty}^{x} \exp\left\{-\tfrac{1}{2}(1-\beta^2)y^2\right\} \Phi(\beta y) \mathrm{d}y, \qquad x \in \mathbb{R},$$

for all $|\beta| < 1$ (see Anděl and Ranocha, 2005), and X_t is a standard normal random variable for each t.

In the final set of experiments, we compare the distance test based on $\mathcal{A}$ to the moment-based test discussed in Bai and Ng (2005). The latter is based on the statistic

$$\mathcal{B} := n\{\hat{\tau}_3^{-2}\hat{\kappa}_3^2 + \hat{\tau}_4^{-2}(\hat{\kappa}_4 - 3)^2\},$$

where $\hat{\kappa}_r := n^{-1}\sum_{t=1}^{n}\{\hat{\gamma}_0^{-1/2}(X_t - \bar{X})\}^r$ $(r = 3, 4)$, and $\hat{\tau}_3^2$ and $\hat{\tau}_4^2$ are estimators of the asymptotic variance of $\sqrt{n}\hat{\kappa}_3$ and $\sqrt{n}(\hat{\kappa}_4 - 3)$, respectively, that are consistent under normality. As in Bai and Ng (2005), $\hat{\tau}_3^2$ and $\hat{\tau}_4^2$ are constructed using a non-parametric kernel estimator with Bartlett weights and a data-dependent bandwidth selected according to the procedure of Andrews (1991). When $\{X_t\}$ is a Gaussian process with absolutely summable autocovariances, the asymptotic distribution of $\mathcal{B}$ is chi-square with 2 degrees of freedom. (We note that the normality test of Lobato and Velasco (2004) is also based on $\mathcal{B}$, but their choice for $\hat{\tau}_3^2$ and $\hat{\tau}_4^2$ is different from that of Bai and Ng (2005). Following Bai and Ng (2005), the data-generating mechanism is the autoregressive model

M6: $X_t = \varphi X_{t-1} + \varepsilon_t, \quad \varphi \in \{0, 0.5, 0.8\}$.

For each design point, 1000 independent realizations of $\{X_t\}$ of length $500 + n$, with $n \in \{100, 200, 500\}$, are generated. The first 500 data points of each realization are then discarded in order to eliminate start-up effects and the remaining n data points are used to compute the value of the test statistic of interest. P-values for the distance test are computed from $m = 1000$ bootstrap replicates of $\mathcal{A}$. Unless stated otherwise, the sieve order is selected by AIC with $p_{\max} = \lfloor(\log n)^2\rfloor$; the approximating autoregressive model is fitted by least-squares, which is preferred over the Yule–Walker method because it produces estimates that exhibit smaller finite-sample bias.

4.2. Simulation results

The Monte Carlo rejection frequencies, under M1, of the distance test at 5% significance level ($\alpha = 0.05$) are reported in Table 2. The null rejection probabilities of the test are generally insignificantly different from the nominal level across all values of the dependence parameter. The test also performs well under non-Gaussianity, its rejection frequencies improving with larger sample sizes and smaller values of the dependence parameter. Asymmetry in the distribution of ε_t leads, perhaps unsurprisingly, to higher rejection rates. For long-range dependent data, the test generally suffers a loss in power compared to the short-range dependent or antipersistent cases, a loss which becomes more pronounced the larger the value of d is. Psaradakis (2016) reports a similar finding for bootstrap-based tests of distributional symmetry about an unspecified centre.

As a robustness check, we repeat the experiments setting $p_{\max} = 2\lfloor(\log n)^2\rfloor$. The rejection frequencies of a 5%-level distance test shown in Table 3 (for $n = 200$) reveal that, irrespective of the value of the dependence parameter, there are no substantial changes in the empirical level and power of the test as a result of the increase in the maximal sieve order.

To assess the sensitivity of results with respect to the method used to determine the order of the autoregressive sieve, we consider selecting the latter by minimizing BIC and MC in addition to AIC. The rejection frequencies under M1 of a 5%-level test based on $\mathcal{A}$, with $n = 200$ and $p_{\max} = \lfloor(\log n)^2\rfloor$, are reported in Table 4. It is clear that there is little to choose between BIC, MC and AIC, the rejection frequencies not being notably different across the three criteria for any given combination of noise distribution and value of d. It is worth noting that results from experiments based on artificial time series of length $n = 100$ and $n = 500$ from M1 confirm the robustness of the properties of the distance test with respect to the choice of order selection criterion and maximal sieve order.

The distance test based on $\mathcal{A}$ also works very well for data-generating processes which are not representable as (1) or (6). This can be seen in Table 5, which shows the rejection frequencies of a 5%-level test under M2, M3 and M4. The

Table 3
Rejection frequencies of $\mathcal{A}$ test under M1, $n = 200$.

	$p_{\max} = \lfloor(\log n)^2\rfloor$					$p_{\max} = 2\lfloor(\log n)^2\rfloor$				
Distr.\d	−0.40	−0.25	0	0.25	0.40	−0.40	−0.25	0	0.25	0.40
N	0.05	0.05	0.05	0.05	0.06	0.05	0.04	0.04	0.06	0.06
S1	0.75	0.70	0.41	0.14	0.09	0.75	0.67	0.43	0.14	0.12
S2	0.94	0.92	0.66	0.23	0.13	0.92	0.95	0.66	0.21	0.15
S3	0.99	0.99	0.88	0.37	0.21	0.99	0.99	0.87	0.37	0.23
A1	1.00	1.00	0.98	0.42	0.16	1.00	1.00	0.97	0.45	0.18
A2	0.99	0.99	0.91	0.43	0.21	0.99	0.98	0.89	0.43	0.24
A3	1.00	1.00	1.00	0.83	0.33	1.00	1.00	1.00	0.83	0.35
LN	1.00	1.00	1.00	0.94	0.54	1.00	1.00	1.00	0.94	0.55

Table 4
Rejectionfrequencies of $\mathcal{A}$ test under M1, $n = 200$, and $p_{\max} = \lfloor(\log n)^2\rfloor$.

	AIC					BIC					MC				
Distr.\d	−0.40	−0.25	0	0.25	0.40	−0.40	−0.25	0	0.25	0.40	−0.40	−0.25	0	0.25	0.40
N	0.05	0.05	0.05	0.05	0.06	0.06	0.04	0.06	0.07	0.07	0.06	0.03	0.04	0.06	0.05
S1	0.75	0.70	0.41	0.14	0.09	0.75	0.71	0.38	0.135	0.10	0.71	0.70	0.36	0.13	0.12
S2	0.94	0.92	0.66	0.23	0.13	0.93	0.92	0.68	0.26	0.14	0.95	0.92	0.67	0.24	0.13
S3	0.99	0.99	0.88	0.37	0.21	0.99	0.99	0.88	0.39	0.23	0.99	0.98	0.86	0.35	0.21
A1	1.00	1.00	0.98	0.42	0.16	1.00	1.00	0.97	0.478	0.17	1.00	1.00	0.97	0.41	0.16
A2	0.99	0.99	0.91	0.43	0.21	0.99	0.99	0.91	0.45	0.20	1.00	0.99	0.89	0.40	0.19
A3	1.00	1.00	1.00	0.83	0.33	1.00	1.00	1.00	0.82	0.38	1.00	1.00	1.00	0.80	0.35
LN	1.00	1.00	1.00	0.94	0.54	1.00	1.00	1.00	0.95	0.59	1.00	1.00	1.00	0.95	0.51

Table 5
Rejection frequencies of $\mathcal{A}$ test under M2–M4.

	$n = 100$			$n = 200$			$n = 500$		
Distr.	M2	M3	M4	M2	M3	M4	M2	M3	M4
N	0.37	0.28	0.94	0.64	0.43	1.00	0.93	0.82	1.00
S1	0.68	0.65	0.99	0.93	0.91	1.00	1.00	1.00	1.00
S2	0.81	0.79	0.99	0.97	0.97	1.00	1.00	1.00	1.00
S3	0.87	0.89	1.00	0.99	0.99	1.00	1.00	1.00	1.00
A1	0.57	0.86	0.97	0.84	0.99	1.00	0.99	1.00	1.00
A2	0.71	0.86	0.99	0.91	0.99	1.00	1.00	1.00	1.00
A3	0.96	1.00	0.97	1.00	1.00	1.00	1.00	1.00	1.00
LN	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00

Table 6
Rejection frequencies of $\mathcal{A}$ test under M5.

β	$n = 100$	$n = 200$	$n = 500$
0.5	0.053	0.057	0.052
−0.5	0.040	0.050	0.052

rejection rate of the distance test exceeds 57% for any design point with non-Gaussian noise, even for the smallest sample size considered. In the case of artificial time series from M5, the one-dimensional marginal distribution of which is Gaussian, the test has rejection rates that do not differ substantially from the nominal level, as can be seen in Table 6.

Let us finally turn to Table 7, which contains the rejection frequencies, at the 5% significance level, of the tests based on $\mathcal{A}$ and $\mathcal{B}$ under M6. Unlike the distance test, the moment-based test is prone to level distortion. The differences in the empirical levels of the two tests notwithstanding, the distance test has a clear advantage under non-normality, outperforming the moment-based test for every design point in our simulations. The differences are particularly striking for symmetric noise distributions, cases in which the $\mathcal{B}$ test has little or no power to detect non-Gaussianity when $\varphi \neq 0$.

Since the comparison between the distance-based and moment-based tests is perhaps somewhat unfair given that the latter relies on a conventional large-sample approximation to the null distribution of $\mathcal{B}$, we also consider a bootstrap-based version of the $\mathcal{B}$ test. Specifically, the same autoregressive sieve bootstrap procedure that is used in the case of the distance test is employed to compute P-values and critical values for a normality test based on $\mathcal{B}$. The rejection frequencies, at the 5% significance level, of the bootstrap-based $\mathcal{B}$ test for $n = 200$ are shown in Table 8 under the heading $\mathcal{B}_B$. Although the test controls the probability of Type I error marginally better than its asymptotic counterpart $\mathcal{B}$, it is clearly dominated by the distance $\mathcal{A}$ test in terms of power.

Table 7
Rejection frequencies of $\mathcal{A}$ and $\mathcal{B}$ tests under M6.

	$n=100$						$n=200$						$n=500$					
Distr.\ φ	0.0		0.5		0.8		0.0		0.5		0.8		0.0		0.5		0.8	
	$\mathcal{A}$	$\mathcal{B}$	$\mathcal{A}$	$\mathcal{B}$	$\mathcal{A}$	$\mathcal{B}$	$\mathcal{A}$	$\mathcal{B}$	$\mathcal{A}$	$\mathcal{B}$	$\mathcal{A}$	$\mathcal{B}$	$\mathcal{A}$	$\mathcal{B}$	$\mathcal{A}$	$\mathcal{B}$	$\mathcal{A}$	$\mathcal{B}$
N	0.05	0.05	0.05	0.03	0.08	0.01	0.05	0.09	0.06	0.05	0.06	0.02	0.05	0.08	0.05	0.09	0.05	0.04
S1	0.50	0.06	0.27	0.01	0.13	0.00	0.78	0.13	0.43	0.04	0.15	0.00	0.99	0.53	0.76	0.22	0.16	0.02
S2	0.72	0.07	0.43	0.04	0.17	0.01	0.95	0.22	0.70	0.09	0.24	0.02	1.00	0.50	0.97	0.34	0.39	0.06
S3	0.89	0.09	0.64	0.04	0.30	0.01	1.00	0.20	0.89	0.11	0.37	0.03	1.00	0.37	1.00	0.33	0.62	0.09
A1	0.98	0.81	0.79	0.21	0.26	0.00	1.00	1.00	0.98	0.83	0.44	0.03	1.00	1.00	1.00	1.00	0.81	0.46
A2	0.92	0.22	0.66	0.10	0.28	0.01	1.00	0.52	0.90	0.35	0.44	0.06	1.00	0.87	1.00	0.79	0.76	0.36
A3	1.00	0.95	1.00	0.45	0.59	0.00	1.00	0.99	1.00	0.97	0.87	0.03	1.00	1.00	1.00	1.00	1.00	0.73
LN	1.00	0.84	1.00	0.73	0.81	0.02	1.00	0.94	1.00	0.93	0.97	0.16	1.00	0.99	1.00	0.97	1.00	0.88

Table 8
Rejection frequencies of $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{B}_B$ tests under M6 and $n=200$.

Distr.\ φ	0.0			0.5			0.8		
	$\mathcal{A}$	$\mathcal{B}$	$\mathcal{B}_B$	$\mathcal{A}$	$\mathcal{B}$	$\mathcal{B}_B$	$\mathcal{A}$	$\mathcal{B}$	$\mathcal{B}_B$
N	0.05	0.09	0.07	0.06	0.05	0.05	0.06	0.02	0.05
S1	0.78	0.13	0.07	0.43	0.04	0.03	0.15	0.00	0.02
S2	0.95	0.22	0.10	0.70	0.09	0.05	0.24	0.02	0.04
S3	1.00	0.20	0.10	0.89	0.11	0.08	0.37	0.03	0.04
A1	1.00	1.00	0.98	0.98	0.83	0.90	0.44	0.03	0.26
A2	1.00	0.52	0.36	0.90	0.35	0.37	0.44	0.06	0.16
A3	1.00	0.99	0.97	1.00	0.97	0.98	0.87	0.03	0.56
LN	1.00	0.94	0.90	1.00	0.93	0.89	0.97	0.16	0.58

5. Real-data application

The bootstrap-based distance test for normality is applied to a large set of U.S. economic and financial time series. The data set consists of 79 time series associated with the financial markets (12 series), the labour market (22 series), prices (17 series), money and credit (7 series), output, income and capacity (14 series), and surveys (7 series). All time series are monthly, spanning the period 1971–2013, seasonally adjusted, and (with the exception of survey series) transformed to stationarity by differencing either the raw series (indicated by Δ in what follows) or their natural logarithms (indicated by $\Delta\log$). The data were downloaded from Haver Analytics.

P-values for the normality test based on $\mathcal{A}$ are presented in Table 9. These are computed from 10, 000 bootstrap replications, with the data-dependent sieve order $\hat{p}$ determined as the minimizer of AIC over $1 \le p \le \lfloor(\log n)^2\rfloor$. For comparison, we also report asymptotic P-values for the test based on $\mathcal{B}$ (with $\hat{\tau}_3^2$ and $\hat{\tau}_4^2$ computed as in Section 4). Finally, we report a semi-parametric estimate $\hat{d}$ of the dependence parameter of each time series, obtained using the local Whittle estimator, that is, the minimizer over $|d| \le 0.499$ of the objective function

$$\log\left(\ell^{-1}\sum_{i=1}^{\ell}\omega_i^{2d}I_i\right) - 2d\ell^{-1}\sum_{i=1}^{\ell}\log\omega_i,$$

where ℓ is a positive integer chosen as a function of n so that $\ell^{-1}+n^{-1}\ell \to 0$ as $n \to \infty$ (see Robinson, 1995). We set $\ell = \lfloor\{16(-2.19\hat{c})^2\}^{-1/5}n^{4/5}\rfloor$, where $\hat{c}$ is the least-squares estimate of the third coefficient in the pseudo-regression of $\log I_i$ on $(1, -2\log\omega_i, \omega_i^2/2)$, $i=1,\ldots,\lfloor 0.3n^{8/9}\rfloor$ (cf. Henry and Robinson, 1996; Andrews and Sun, 2004).

Evidence in favour of non-Gaussianity in the U.S. economic time series is overwhelming: the null hypothesis is rejected, at the 5% significance level, for 95% of the series on the basis of the $\mathcal{A}$ test. Interestingly, non-normality is found to be a characteristic feature for all six categories of time series. By comparison, the moment-based $\mathcal{B}$ test leads to rejection of normality in only 30% of the cases. It must be borne in mind, however, that the validity of the test based on $\mathcal{B}$ relies heavily on the assumption of short-range dependence in the data. Such an assumption does not accord well with the estimates of the dependence parameter shown in Table 9, on the basis of which short-range dependence ($d=0$) is rejected in favour of long-range dependence ($d>0$) for almost 80% of the time series under consideration. It is also worth recalling from our earlier simulation study that, even for short-range dependent data, the moment-based test appears to be considerably less successful than the distance-based test at detecting deviations from Gaussianity. The non-normality of the marginal distribution of the time series under consideration may, of course, be due to a variety of sources, including non-normal noise in a linear representation like (1), non-linearity, and conditional heteroskedasticity.

Table 9
P-values of the $\mathcal{A}$ and $\mathcal{B}$ tests.

Series	Transformation	$\mathcal{A}$	$\mathcal{B}$	$\hat{p}$	$\hat{d}$	$se(\hat{d})$
(A) Financial market						
10-year treasury constant maturity rate	Δ	0.00	0.10	22	-0.03	0.06
1-year treasury constant maturity rate	Δ	0.00	0.17	19	-0.10	0.06
3-month treasury bill: secondary market rate	Δ	0.00	0.16	20	-0.01	0.05
5-year treasury constant maturity rate	Δ	0.00	0.16	22	-0.07	0.07
Effective federal funds rate	Δ	0.00	0.12	38	-0.13	0.08
Moody's seasoned aaa corporate bond yield	Δ	0.00	0.04	15	0.01	0.05
Moody's seasoned baa corporate bond yield	Δ	0.00	0.12	5	0.16	0.07
Foreign exchange rate (Yen per US dolar)	Δlog	0.00	0.01	14	0.16	0.04
Foreign exchange rate (Pound per US dolar)	Δlog	0.00	0.12	3	-0.04	0.08
Foreign Exchange rate (Franc per US dolar)	Δlog	0.00	0.01	11	0.05	0.05
SP 500 composite index (1941–43=10)	Δlog	0.00	0.05	11	0.03	0.07
SP industrialIndex (1941–43=10)	Δlog	0.00	0.04	22	0.02	0.07
(B) Employment, hours, Earnings						
All employees: construction	Δlog	0.00	0.00	38	0.49	0.07
All employees: durable goods	Δlog	0.00	0.05	36	0.43	0.08
All employees: financial activities	Δlog	0.55	0.53	38	0.49	0.06
All employees: goods-producing industries	Δlog	0.00	0.12	37	0.49	0.07
All employees: government	Δlog	0.00	0.10	18	0.32	0.07
All employees: manufacturing	Δlog	0.00	0.05	36	0.44	0.08
All employees: mining and logging: mining	Δlog	0.00	0.16	38	0.24	0.07
All employees: nondurable goods	Δlog	0.00	0.33	38	0.22	0.09
All employees: retail trade	Δlog	0.01	0.03	36	0.49	0.09
All employees: service-providing Industries	Δlog	0.01	0.21	25	0.49	0.09
All employees: total nonfarm	Δlog	0.00	0.04	36	0.49	0.06
All employees: trade, transportation and utilities	Δlog	0.00	0.14	36	0.49	0.07
All employees: wholesale trade	Δlog	0.02	0.22	37	0.49	0.06
Average hourly earnings of production: construction	Δlog	0.00	0.15	35	0.49	0.08
Averagehourly earnings of production: goods-Producing	Δlog	0.00	0.00	33	0.49	0.08
Average hourly earnings of production: manufacturing	Δlog	0.00	0.00	34	0.49	0.08
Average weekly hours of production: goods-producing	Δ	0.00	0.37	35	-0.06	0.07
Average weekly hours of production: Manufacturing	Δ	0.00	0.35	36	-0.04	0.06
Average weekly overtime hours: manufacturing	Δ	0.00	0.27	35	-0.07	0.04
Civilian employment	Δlog	0.00	0.04	37	0.35	0.06
Civilian labor force	Δ	0.00	0.24	37	0.28	0.08
Civilian unemployment rate	Δ	0.00	0.21	38	0.49	0.07
(C) Prices						
Consumer price index: all items	Δlog	0.00	0.00	18	0.49	0.07
Consumer price index: all items less food	Δlog	0.00	0.02	15	0.41	0.07
Consumer price index: apparel	Δlog	0.00	0.03	18	0.39	0.09
Consumerprice index: commodities	Δlog	0.00	0.19	17	0.18	0.06
Consumer price index: durables	Δlog	0.00	0.00	37	0.28	0.05
Consumer price index: medical Care	Δlog	0.00	0.09	36	0.49	0.08
Consumer price index: services	Δlog	0.00	0.07	31	0.49	0.09
Consumer price index: transportation	Δlog	0.00	0.32	12	0.06	0.08
Personal consumption expenditures	Δlog	0.00	0.00	17	0.49	0.08
Personal consumption expenditures: durable goods	Δlog	0.03	0.16	37	0.49	0.08
Personal consumption expenditures: nondurable goods	Δlog	0.00	0.10	16	0.18	0.04
Personal consumption expenditures: services	Δlog	0.00	0.00	36	0.49	0.09
Producer price index: commodities: metals	Δlog	0.00	0.06	20	0.00	0.09
Producer price index: crude materials	Δlog	0.00	0.04	38	-0.17	0.09
Producer price index: finished consumer goods	Δlog	0.00	0.16	17	0.21	0.05
Producer price index: finished goods	Δlog	0.00	0.11	17	0.25	0.06
Producer price index: intermediate materials	Δlog	0.00	0.11	14	0.15	0.09
(D) Money and credits						
M1 money stock	Δlog	0.00	0.15	36	0.45	0.08
M2 money stock	Δlog	0.00	0.06	34	0.34	0.05
M3 money stock	Δlog	0.00	0.05	34	0.35	0.05
Commercial and industrial loans, all commercial banks	Δlog	0.01	0.08	36	0.49	0.04
Real estate loans, all commercial banks	Δlog	0.07	0.12	7	0.49	0.08
Real M2 money stock	Δlog	0.00	0.26	18	0.26	0.06
Total nonrevolving credit owned, outstanding	Δlog	0.00	0.34	38	0.49	0.07
(E) Output, income and capacity						
Industrial production: business equipment	Δlog	0.00	0.21	12	0.40	0.05
Industrial production: consumer goods	Δlog	0.00	0.17	24	0.13	0.06
Industrial production: durable materials	Δlog	0.00	0.08	36	0.28	0.07
Industrial production: final products	Δlog	0.00	0.07	36	0.26	0.08
Industrial production: fuels	Δlog	0.00	0.42	36	-0.15	0.09
Industrial production index	Δlog	0.00	0.18	38	0.31	0.08

(*continued on next page*)

Table 9 (*continued*)

Series	Transformation	$\mathcal{A}$	$\mathcal{B}$	$\hat{p}$	$\hat{d}$	$se(\hat{d})$
Industrial production: manufacturing	Δlog	0.00	0.18	38	0.23	0.08
Industrial production: materials	Δlog	0.00	0.09	37	0.30	0.07
Industrial production: nondurable goods	Δlog	0.76	0.55	37	0.17	0.08
Industrial production: nondurable materials	Δlog	0.00	0.17	37	-0.08	0.08
Personal income	Δlog	0.00	0.09	15	0.49	0.09
Real personal income	Δlog	0.00	0.10	17	0.37	0.09
Real personal income excluding current transfers	Δlog	0.00	0.14	14	0.36	0.10
Capacity utilization: manufacturing	Δ	0.00	0.22	38	0.14	0.08
(F) Surveys						
ISM manufacturing: employment index	–	0.04	0.06	38	0.49	0.08
ISM manufacturing: inventories index	–	0.02	0.01	37	0.49	0.07
ISM manufacturing: new orders index	–	0.00	0.01	38	0.35	0.08
ISM manufacturing: PMI composite index	–	0.01	0.02	38	0.49	0.08
ISM manufacturing: prices index	–	0.11	0.03	17	0.41	0.09
ISM manufacturing: production index	–	0.00	0.03	38	0.28	0.09
ISM manufacturing: supplier deliveries	–	0.00	0.11	38	0.34	0.09

6. Conclusion

This paper considered an Anderson–Darling distance test for normality of the one-dimensional marginal distribution of stationary, fractionally integrated processes. As a practical way of implementing the test, we proposed using an autoregressive sieve bootstrap procedure to estimate finite-sample *P*-values and/or critical values. The bootstrap-based test is valid for short-range dependent, long-range dependent and antipersistent processes, and does not require knowledge or estimation of the dependence parameter of the data or of the appropriate norming factor for the test statistic. Monte Carlo simulations showed that the distance test has good size and power properties in small samples, although it tends to lack power when the dependence parameter is large and positive. Under short-range dependence, the distance test was found to be more successful than the popular skewness/kurtosis test in detecting departures from normality. An application to a set of U.S. economic and financial time series revealed than non-Gaussianity is a prevalent feature of the data.

Acknowledgments

The authors are grateful to Gianni Amisano, Ron Smith, an Associate Editor, and two anonymous referees for their helpful comments and suggestions.

References

Anderson, T.W., Darling, D.A., 1952. Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. Ann. Math. Stat. 23, 193–212.
Anderson, T.W., Darling, D.A., 1954. A test of goodness of fit. J. Am. Stat. Assoc. 49, 765–769.
Andrews, D.W.K., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 59, 817–858.
Andrews, D.W.K., Sun, Y., 2004. Adaptive local polynomial Whittle estimation of long-range dependence. Econometrica 72, 569–614.
Anděl, J., Ranocha, P., 2005. Stationary distribution of absolute autoregression. Kybernetika 41, 735–742.
Arcones, M.A., 2006. On the Bahadur slope of the Lilliefors and the Cramér–von Mises tests of normality. In: Giné, E., Koltchinskii, V., Li, W., Zinn, J. (Eds.), Proceedings of the Fourth International Conference on High Dimensional Probability. Institute of Mathematical Statistics, Beachwood, Ohio, pp. 196–206.
Bai, J., Ng, S., 2005. Tests for skewness, kurtosis, and normality for time series data. J. Bus. Econ. Stat. 23, 49–60.
Beran, J., Ghosh, S., 1991. Slowly decaying correlations, testing normality, nuisance parameters. J. Am. Stat. Assoc. 86, 785–791.
Berg, A., Paparoditis, E., Politis, D.N., 2010. A bootstrap test for time series linearity. J. Stat. Plan. Inference 140, 3841–3857.
Bickel, P.J., Bühlmann, P., 1997. Closure of linear processes. J. Theor. Prob. 10, 445–479.
Bickel, P.J., Freedman, D.A., 1981. Some asymptotic theory for the bootstrap. Ann. Stat. 9, 1196–1217.
Bontemps, C., Meddahi, N., 2005. Testing normality: a GMM approach. J. Econom. 124, 149–186.
Boutahar, M., 2010. Behaviour of skewness, kurtosis and normality tests in long memory data. Stat. Methods Appl. 19, 193–215.
Bowman, K.O., Shenton, L.R., 1975. Omnibus test contours for departures from normality based on $\sqrt{b_1}$ and b_2. Biometrika 62, 243–250.
Brockwell, P.J., Davis, R.A., 1991. Time Series: Theory and Methods, 2nd edition. Springer, New York.
Bühlmann, P., 1997. Sieve bootstrap for time series. Bernoulli 3, 123–148.
Cotter, J., 2007. Varying the VaR for unconditional and conditional environments. J. Int. Money Financ. 26, 1338–1354.
Dehling, H., Taqqu, M.S., 1989. The empirical process of some long-range dependent sequences with an application to *U*-statistics. Ann. Stat. 17, 1767–1783.
Doukhan, P., Oppenheim, G., Taqqu, M.S. (Eds.), 2003, Theory and Applications of Long-Range Dependence. Birkhäuser, Boston.
Doukhan, P., Surgailis, D., 1998. Functional central limit theorem for the empirical process of short memory linear processes. Comptes Rendus Acad. Sci. – Series I 326, 87–92.
Epps, T.W., 1987. Testing that a stationary time series is Gaussian. Ann. Stat. 15, 1683–1698.
Giraitis, L., Surgailis, D., 1999. Central limit theorem for the empirical process of a linear sequence with long memory. J. Stat. Plan. Inference 80, 81–93.
Hall, P., Wilson, S.R., 1991. Two guidelines for bootstrap hypothesis testing. Biometrics 47, 757–762.
Hassler, U., Kokoszka, P., 2010. Impulse responses of fractionally integrated processes with long memory. Econom. Theory 26, 1855–1861.
Henry, M., Robinson, P.M., 1996. Bandwidth choice in Gaussian semiparametric estimation of long range dependence. In: Robinson, P.M., Rosenblatt, M. (Eds.), Athens Conference on Applied Probability and Time Series Analysis, Vol. II: Time Series Analysis in Memory of E. J. Hannan. Springer–Verlag, New York, pp. 220–232.
Hinich, M.J., 1982. Testing for Gaussianity and linearity of a stationary time series. J. Time Ser. Anal. 3, 169–176.
Ho, H.-C., 2002. On functionals of linear processes with estimated parameters. Stat. Sin. 12, 1171–1190.

Jarque, C.M., Bera, A.K., 1987. A test for normality of observations and regression residuals. Int. Stat. Rev. 55, 163–172.
Kapetanios, G., Psaradakis, Z., 2006. Sieve Bootstrap for Strongly Dependent Stationary Processes. Department of Economics, Queen Mary, University of London. Working Paper.
Kilian, L., Demiroglu, U., 2000. Residual-based tests for normality in autoregressions: asymptotic theory and simulation evidence. J. Bus. Econ. Stat. 18, 40–50.
Koziol, J.A., 1986. Relative efficiencies of goodness of fit procedures for assessing univariate normality. Ann. Inst. Stat. Math. 38, 485–493.
Kreiss, J.-P., 1992. Bootstrap procedures for AR(∞) processes. In: Jöckel, K.-H., Rothe, G., Sendler, W. (Eds.), Bootstrapping and Related Techniques. Springer-Verlag, Heidelberg, pp. 107–113.
Kulik, R., 2009. Empirical process of long-range dependent sequences when parameters are estimated. J. Stat. Plan. Inference 139, 287–294.
Lehmann, E.L., Romano, J.P., 2005. Testing Statistical Hypotheses, 3rd ed. Springer, New York.
Lobato, I.N., Velasco, C., 2004. A simple test of normality for time series. Econom. Theory 20, 671–689.
Nusrat, J., Harvill, J.L., 2008. Bispectral-based goodness-of-fit tests of Gaussianity and linearity of stationary time series. Commun. Stat. – Theory Methods 37, 3216–3227.
Palma, W., 2007. Long-Memory Time Series: Theory and Methods. Wiley, New York.
Poskitt, D.S., 2007. Autoregressive approximation in nonstandard situations: the fractionally integrated and non-invertible cases. Ann. Inst. Stat. Math. 59, 697–725.
Poskitt, D.S., 2008. Properties of the sieve bootstrap for fractionally integrated and non-invertible processes. J. Time Ser. Anal. 29, 224–250.
Psaradakis, Z., 2016. Using the bootstrap to test for symmetry under unknown dependence. J. Bus. Econ. Stat. 34, 406–415.
Robinson, P.M., 1995. Gaussian semiparametric estimation of long range dependence. Ann. stat. 23, 1630–1661.
Rosenblatt, M., 2000. Gaussian and Non-Gaussian Linear Time Series and Random Fields. Springer, New York.
Shapiro, S.S., Wilk, M.B., 1965. An analysis of variance test for normality (complete samples). Biometrika 52, 591–611.
Shibata, R., 1980. Asymptotically efficient selection of the order of the model for estimating parameters of a linear process. Ann. Stat. 8, 147–164.
Shorack, G.R., Wellner, J.A., 1986. Empirical Processes with Applications to Statistics. Wiley, New York.
Stephens, M.A., 1974. EDF statistics for goodness of fit and some comparisons. J. Am. Stat. Assoc. 69, 730–737.
Thode, H.C., 2002. Testing for Normality. Marcel Dekker, New York.
Tong, H., 1990. Non-linear Time Series: A Dynamical System Approach. Oxford University Press, Oxford.