Econometrics and Statistics 000 (2016) 1–11

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 xxx

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/

Please cite this article as: Z. Psaradakis, M. Vávra, A distance test of normality for a wide class of stationary processes, Econometrics and Statistics (2016), http://dx.doi.org/10.1016/j.ecosta.2016.11.005

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).

Please cite this article as: Z. Psaradakis, M. Vávra, A distance test of normality for a wide class of stationary processes, Econometrics and Statistics (2016), http://dx.doi.org/10.1016/j.ecosta.2016.11.005