Available online at www.sciencedirect.com
ScienceDirect
Computational Statistics & Data Analysis 51 (2007) 2734 – 2746
COMPUTATIONAL STATISTICS & DATA ANALYSIS
www.elsevier.com/locate/csda

Robust directed tests of normality against heavy-tailed alternatives

Yulia R. Gel[a,*], Weiwen Miao[b], Joseph L. Gastwirth[c]

[a]Department of Statistics and Actuarial Science, Faculty of Mathematics, University of Waterloo, 200 University Ave. W., Waterloo, Ont., Canada N2L 3G1
[b]Department of Mathematics and Computer Science, Macalester College, St.Paul, MN, USA
[c]Department of Statistics, George Washington University, Washington, DC, USA

Received 19 April 2006; received in revised form 12 August 2006; accepted 13 August 2006
Available online 14 September 2006

Abstract

Many statistical procedures rely on the assumption that the observed data are normally distributed. Consequently, there exists a vast literature on tests of normality and their statistical properties. Today the most commonly used omnibus test for general use is the Shapiro–Wilk method while the Jarque–Bera test is the most popular omnibus test in economics and related applications.

In a recent securities law case concerning an allegation of profit sharing, an inference is drawn on a relationship between the hypothetical profits, which customers made on days with special stock offerings, and the ratio of their commission business to those profits became an issue. Two common statistical tests are based on the Pearson or Spearman correlation coefficients. The results provided by the two tests are dramatically different with contradictory implications for the existence of a "profit sharing arrangement". In particular, the Pearson correlation coefficient is of -0.13 with a p-value of 0.35 while the Spearman coefficient is of -0.68 with a p-value less than 0.0001. However, if the data do not come from a bivariate normal distribution, the usual sampling theory for the Pearson correlation would not be applicable. As stock market data are typically heavy-tailed, a simple test of normality directed against heavy-tailed alternatives complemented by a graphical method is desired for presentation in a courtroom.

In this paper we utilize the idea of the Shapiro–Wilk test whose test statistic W is the ratio of the classical standard deviation s_n to the optimal L-estimator of the population spread. In particular, we replace the mean and standard deviation by the median and a robust estimator of spread in the Shapiro–Wilk statistic W. This eventually leads to a new more powerful directed test and a clearer QQ plot for assessing departures from normality in the tails. Somewhat surprisingly, the robust estimator of spread that yields the more powerful test of normality utilizes the average absolute deviation from the median (MAAD) while the statistic based on the median absolute deviation from the median (MAD) usually provides a clearer graphical display. The asymptotic distribution of the proposed test statistic is derived, and the use of the new test is illustrated by simulations and by application to data sets from the securities law, finance and meteorology.

Keywords: Tests of normality; Robust estimators of scale; Shapiro–Wilk test; Power; Graphical check of normality

1. Introduction

Many statistical procedures are based on the assumption that the data are a random sample of size n from a normal distribution. Consequently, a variety of tests have been developed to check the validity of this assumption (Lilliefors,

* Corresponding author. Tel.: +1 519 8851211; fax: +1 519 7461875.
E-mail addresses: ygl@math.uwaterloo.ca (Y.R. Gel), miao@macalester.edu (W. Miao), jlgast@gwu.edu (J.L. Gastwirth).

0167-9473/$ - see front matter
doi:10.1016/j.csda.2006.08.022

1967; Shapiro and Wilk, 1965; Jarque and Bera, 1980; D'Agostino and Stephens, 1986). Today the most popular omnibus test for normality for a general use is the Shapiro–Wilk (SW) test. The Jarque–Bera (JB) test is the most widely adopted omnibus test for normality in econometrics and related fields. In particular, JB which is based on the classical measures of skewness and kurtosis, is frequently used when one is more concerned with heavy-tailed alternatives. The Lilliefors (Kolomorov–Smirnov) (L(KS)) test is the best known omnibus test based on empirical distribution function (EDF). Being omnibus procedures, the SW, JB and L(KS) tests do not provide insight about the nature of the deviation from normality, e.g. skewness, heavy tails or outliers. Therefore, specialized tests directed at particular alternatives are desired in many practical situations.

Our interest in revisiting this classical goodness-of-fit problem was motivated by our involvement as statistical consultants for the defense in a securities law case concerning alleging "profit sharing". The relationship between the hypothetical profits (HP), which customers made on days with highly profitable initial public offerings (IPO) of stocks, and the fraction of the HP formed by the commission (C) business on those days was important. If customers were sharing their profits, which is not allowed, the ratio (C/HP) should be nearly constant as it would fluctuate around the agreed upon "share". If customers were not "profit sharing" one would expect that the C/HP ratio would be less on days when there were high "profits" even though there might be some increase in commissions as customers were making money. Applying the two commonly used measures of correlation to C/HP and HP, the Pearson and Spearman coefficients, yields dramatically different conclusions. The two commonly used statistical tests here are based on the Pearson ($r = -0.13$, $p = 0.35$) or Spearman ($\rho = -0.68$, $p < 0.0001$) correlation coefficients. However, if the data do not come from a bivariate normal distribution, or more simply, if neither variable is normally distributed, the Pearson correlation is not appropriate while the distribution-free Spearman correlation would be. As stock market data typically are heavy-tailed, our main focus was on detecting heavy-tailed alternatives.

In this paper we propose a directed test of normality against heavy-tailed alternatives based on the comparison of two estimators of the spread σ for normal data, the classical standard deviation s_n and the more robust average absolute deviation from the sample median, named J_n. Under the null hypothesis of normality, the robust deviation J_n is a consistent estimate of the population standard deviation σ. Thus, the ratio s_n/J_n should be near 1 for normally distributed data. Robust estimators of spread about the sample median M rather than the sample mean $\bar{X}$ are typically less sensitive to departures from normality in the tails such as heavy tails or outliers, which implies that the ratio of s_n vs. J_n will deviate from 1. This leads to a new directed test SJ based on the statistic $R = s_n/J_n$. SJ is similar to the Bonett–Seier test (Bonett and Seier, 2002) which is based on the modified (log-transformed) Geary's measure of kurtosis, or G-kurtosis (Geary, 1936), i.e. the ratio of the average absolute deviation from the sample mean to s_n. Both SJ and BS are directed tests of normality designed for symmetric distributions. In particular, if data fail those tests, the conclusion is that the underlying distribution has non-normal tails or outliers. However, that does not preclude the possibility of substantial skewness with respect to the normal distribution. Therefore, if skewness is expected, the directed test should be accompanied by a test of symmetry.

Under the normality assumption, the Shapiro–Wilk statistic W is also the ratio of two estimators of spread, s_n and the optimal L-estimator of σ (Chernoff et al., 1967; Sarkadi, 1982). The asymptotic distribution of W is derived by De Wet and Venter (1972) and Leslie et al. (1986). Under the null hypothesis of normality, the SW test statistic W does not follow a normal distribution. Therefore, the critical values of W should be either computed empirically or obtained using the Royston approximation of weights (Royston, 1982). Thus, the critical values of the SW test are not easily accessible without an aid of a computer. Since the new test statistic R is also a ratio of two L-estimators of σ, we can use the theory of L-statistics to derive the large sample distribution of $\sqrt{n}R$ under general alternatives. In particular, under the null hypothesis, $\sqrt{n}(R-1) \sim \mathrm{N}(0, \sigma_R)$. It will be seen that the new directed test SJ is more powerful against heavy-tailed alternatives than the omnibus Shapiro–Wilk, Jarque–Bera and Lilliefors tests for all sample sizes and than the Bonett–Seier test for small sample sizes. As expected, SJ is less powerful for different departures from normality, e.g. skewness. The fact that the test statistic s_n/J_n is easy to compute and that it follows an asymptotic normal distribution, makes it attractive as a "quick" check against heavy-tailed alternatives.

The new test SJ and its asymptotic properties are introduced in Section 2. Section 3 is devoted to the size of the new test SJ and a useful approximation to its distribution in small samples. The power of SJ is illustrated in Section 4 by simulations and application to data sets from the securities law, finance and meteorology. The final section discusses the various tests of normality examined and provides guidance for the practical use of the new test.