On a generalization of the Jarque-Bera test using the bootstrap method

Saeid Amiri, Dietrich von Rosen, and Silvelyn Zwanzig

Research Report
Centre of Biostochastics

Swedish University of Agricultural Sciences

Report 2009:03
ISSN 1651-8543

On a generalization of the Jarque-Bera test using the bootstrap method

Saeid Amiri[1], Dietrich von Rosen

Centre of Biostochastics
Swedish University of Agricultural Sciences
P.O. Box 7032, SE-750 07 Uppsala, Sweden

Silvelyn Zwanzig

Department of Mathematics
Uppsala University, P.O.Box 480, 751 06 Uppsala, Sweden

Abstract

This paper discusses the possibility of using the Jarque-Bera test for distributions other than the normal distribution. In addition, it presents a new idea for performing goodness-of-fit test using bootstrap methods. It is shown that the bootstrapped Jarque-Bera test can be used alongside the conventional Jarque-Bera test to increase statistical power. The proposed tests are simple to use and their properties seem appropriate.

Keywords: Bootstrap, Goodness-of-fit test, Monte Carlo simulation.

[1]E-mail address to the correspondence author: saeid.amiri@et.slu.se

1 Introduction

The study of distributions is one of the main issues in statistics which concerned with assessing the validity of models. Distributional assumptions play an important role in estimation, inference, prediction, etc. Because of the importance of this subject, it is often included in statistical textbooks. Goodness-of-fit tests are covered by D'Agostino and Stephens (1986), Rayner and Best (1989) and Thode (2002), among many authors.

The main aim of this paper is to study the Jarque-Bera test. This test is the normality test, which is known in econometrics literature and time series analysis, see Jarque and Bera (1980) and Kilian and Demiroglu (2000), and the main reason for its widespread use is its straightforward interpretation and implementation. However, its lack of generality motivates closer examination. Furthermore, Our simulations show that it suffers from small sample size. This paper also examines a new way of performing goodness-of-fit test, implemented using the bootstrap analysis.

There are a lot of different statistics to test the goodness-of-fit but some are more well-known, e.g. the Shapiro-Wilk and Kolmogrov-Smironov test, which are discussed in detail by Shapiro et.al (1968). The Shapiro-Wilk test is also used in the study of normality. Its statistic is:

$$SW = \frac{\left(\sum_{i=1}^{n} a_i X_{(i)}\right)^2}{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2},$$

where $X_{(i)}$ is the order statistics and a_i are constants obtained from the means, variances and covariances of the order statistics of a sample of size n from the normal distribution. The Shapiro-Wilk test actually compares an estimate of the standard deviation using a linear combination of the order statistics. It is one of the most frequently used test and is recommended for everyday practice, e.g. Thode (2002).

The Kolmogrov-Smirnov test is used in studies of the exponential distribution, but it can also be used for other distributions. Its statistic is:

$$D_n = \sup_{1 \leq i \leq n} |F(x_i) - F_n(x_i)| \tag{1}$$

It is a form of minimum distance estimation for comparing a sample with a reference probability distribution by using its CDF. Under the null hypothesis, $\sqrt{n}D_n$ converges to the Kolmogorov distribution. It is discussed in detail in Kendall and Stuart (1973).

The main purpose of the present paper is to produce a generalization of the Jarque-Bera test. This is presented in Section 2, which also includes a bootstrapped version of the test. Section 3 outlines a new test, Section 4 explains the bootstrap method and Section 5 includes a study of the power of the tests proposed and discusses their goodness-of-fit testing of the normal distribution and the exponential distribution to reveal their applicability.

2 Jarque-Bera test

To generalize the Jarque-Bera (JB) test , we start by defining it. Its statistics is

$$JB = \frac{n}{6}\left(\widehat{\gamma}^2 + (\widehat{K}-3)^2/4\right), \tag{2}$$

where $\widehat{\gamma}$ and $\widehat{K}$ are the sample skewness and kurtosis. Under the normality of observations, JB has an asymptotic chi-square distribution with two degrees of freedom. It is obvious that it is based on the idea of analyzing the asymptotic properties of $\widehat{\gamma}$ and $\widehat{K}$. If $X_1, \cdots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$ then

$$\begin{aligned} E(\widehat{\gamma}) &\simeq 0, \qquad V(\widehat{\gamma}) \simeq \frac{6}{n}, \\ E(\widehat{K}) &\simeq 3, \qquad V(\widehat{K}) \simeq \frac{24}{n}. \end{aligned}$$

Skewness is discussed by Kendall et al. (1998) and kurtosis by Cramér (1945). D'Agostino and Stephens (1986) discussed their usage to study normality. They also suggest a combination of kurtosis and skewness, which is similar to the JB statistics. Moreover, they classify this sort of test as a moment test because it is an approach to testing departure from normality using third and fourth moment of the random variable. Jarque and Bera (1980, 1987) discuss this test using the Lagrange-Multiplier test. Generalization of this test is rather difficult, because the expectation and variance of skewness and kurtosis are not known for other distributions.

The essential question about the Jarque-Bera test is how it can be used for testing other distributions. It is obvious that the JB statistic can be written as the quadratic form,

$$JB = (\theta - \widehat{\theta})' W (\theta - \widehat{\theta}) = \begin{bmatrix} \widehat{\gamma} & \widehat{K}-3 \end{bmatrix} \begin{bmatrix} \frac{1}{6/n} & 0 \\ 0 & \frac{1}{24/n} \end{bmatrix} \begin{bmatrix} \widehat{\gamma} \\ \widehat{K}-3 \end{bmatrix}, \tag{3}$$

where θ is the parameters concerned and W is the diagonal matrix where elements are the reciprocal of variance of skewness and kurtosis.

This motivates us to study a more general form of this formula, which includes the mean and variance because they include important information on the distribution that can be handled in test statistics because the aim is to generalize the JB test to other distributions. The generalized JB statistic is

$$\mathcal{D} = \left[\begin{array}{cccc} \overline{X}-\mu & S^2-\sigma^2 & \widehat{\gamma}-\gamma & \widehat{K}-K \end{array}\right] W \left[\begin{array}{c} \overline{X}-\mu \\ S^2-\sigma^2 \\ \widehat{\gamma}-\gamma \\ \widehat{K}-K \end{array}\right]. \tag{4}$$

One choice is $W = I$ because the reciprocal of the variance of skewness and kurtosis for other distributions are not known. Moreover, these parameters are not independent and therefore it is rather difficult to find the distribution of $\mathcal{D}$.

The aim of the bootstrap method is to estimate the standard error. Therefore it can be used to estimate W, i.e. the diagonal matrix where elements are the reciprocal of the variance of the parameters concerned. This is referred to as $\mathcal{DW}$ in the reminder of this paper.

The next section discusses the methodology of the new test, which is referred to here as a $\mathcal{G}$-test. In explaining the procedure the normal distribution is used, but it can be generalized to other distributions.

3 New test

As mentioned earlier, the aim is to study the distribution using the parameters concerned; mean, variance, skewness and kurtosis. This type of statistic test is called a moments test by D'Agostino and Stephen (1986). Letting $X_1, \cdots, X_n \overset{iid}{\sim} F$, we can study the following hypothesis:

$$H_0 : F_0(x) \overset{d}{=} F(x). \tag{5}$$

This idea is to construct new parameters by considering the fact that the first four moments, $\mu_i = E(X-\mu)^i,\ i = 1, \cdots, 4$, are functions of the first, second, third and fourth power of the variable, respectively. Consider arbitrary values of U_1, U_2, U_3, U_4 and U_5, where $U_1 < U_2 < U_3 < U_4 < U_5$ or vice versa. There is one more parameter than the parameters concerned. Later, we discuss why five values should be used. It is logical that if U_i can be used to find μ_1, then we should focus on U_i^j which can be used to find μ_j. The following equations show how U_i^j can be used to find μ_j.

$$\begin{aligned}
T_1U_1 + T_2U_2 + T_3U_3 + T_4U_4 + T_5U_5 &= \mu_1, & (6)\\
T_1U_1^2 + T_2U_2^2 + T_3U_3^2 + T_4U_4^2 + T_5U_5^2 &= \mu_2, & (7)\\
T_1U_1^3 + T_2U_2^3 + T_3U_3^3 + T_4U_4^3 + T_5U_5^3 &= \mu_3, & (8)\\
T_1U_1^4 + T_2U_2^4 + T_3U_3^4 + T_4U_4^4 + T_5U_5^4 &= \mu_4, & (9)\\
T_1 + T_2 + T_3 + T_4 + T_5 &= C, & (10)
\end{aligned}$$

Assume that C is given, which is the summation of the coefficients. It actually helps to control the coefficients that plays the role as the penalty. U and θ are defined as:

$$U = \begin{pmatrix} U_1 & U_2 & U_3 & U_4 & U_5 \\ U_1^2 & U_2^2 & U_3^2 & U_4^2 & U_5^2 \\ U_1^3 & U_2^3 & U_3^3 & U_4^3 & U_5^3 \\ U_1^4 & U_2^4 & U_3^4 & U_4^4 & U_5^4 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix}, \tag{11}$$

$$\theta = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \\ C \end{pmatrix}. \tag{12}$$

There is a solution for T_i because $T = U^{-1}\theta$ which can be used for study of the distribution. Matrix U is the Vandermonde matrix and it is nonsingular, so therefore there exists a unique values for T_i.

If the distribution is standard normal, then the estimation of moments must be close to $\mu_1 = 0$, $\mu_2 = 1$, $\mu_3 = 0$ and $\mu_4 = 3$ respectively. Therefore by substituting these values in equation (6)-(10), the solutions are the values that can be expected if the underlying distribution of observations is standard normal. Hence they are referred to as the theoretical values, $T_i, i = 1, ..., 5$.

Based on the sample, the equations are as bellow,

$$\begin{aligned}
O_1U_1 + O_2U_2 + O_3U_3 + O_4U_4 + O_5U_5 &= \widehat{\mu}_1 & (13)\\
O_1U_1^2 + O_2U_2^2 + O_3U_3^2 + O_4U_4^2 + O_5U_5^2 &= \widehat{\mu}_2 & (14)\\
O_1U_1^3 + O_2U_2^3 + O_3U_3^3 + O_4U_4^3 + O_5U_5^3 &= \widehat{\mu}_3 & (15)\\
O_1U_1^4 + O_2U_2^4 + O_3U_3^4 + O_4U_4^4 + O_5U_5^4 &= \widehat{\mu}_4 & (16)\\
O_1 + O_2 + O_3 + O_4 + O_5 &= C, & (17)
\end{aligned}$$

where $\widehat{\mu}_i$ is an estimator of μ_i. If the chosen distribution is correct then it is expected that O_i will be close to T_i. O is

$$O = U^{-1}\widehat{\theta} \tag{18}$$

Any suggested criterion should include comparison of T_i and O_i which are referred to as the theoretical and observed value. Here the discrepancy measure is the square distance of the theoretical value and observed value:

$$\mathcal{G} = \|T - O\| = (\widehat{\theta} - \theta)'(U^{-1})'U^{-1}(\widehat{\theta} - \theta). \tag{19}$$

It is obvious that $(U^{-1})'U^{-1}$ plays a role as the weight in the JB test. The main question that may arise concerns the distribution of the suggested criterion. The bootstrap method is used to obtain the distribution. The following section describes use of the bootstrap method to handle the distribution of $\mathcal{D}$, $\mathcal{DW}$ and $\mathcal{G}$ which are given in (4) and (22).

Another question is which value should be used for U_i. There are many ways to choose the value. The first option is to use the original observations such as $U = (q_{10}, q_{25}, q_{50}, q_{70}, q_{85})$ where q is the quantile of observation. The second is to use fixed values such as $U = (1, 2, 3, 4, 5)$. In Sections 5 and 6, these choices are discussed by simulation.

4 Semiparametric bootstrap

Over the last three decades since the seminal paper by Efron (1979), a great deal of effort has been devoted to the theory and methods of bootstrap. It is known as an inferential tool to study the uncertainty in statistics which makes it interesting to use. There are two basic categories of approaches, the parametric and nonparametric bootstrap. The nonparametric bootstrap is based on the original sample where the empirical distribution, F_n, is involved in the resampling process. In contrast, the parametric bootstrap is based on an assumed distribution which can be used in resampling. Thus the parametric bootstrap approach is not based on resampling from original observations. For details about the bootstrap method, see Efron and Tibshirani (1993) and Davison and Hinkley (1997).

It is obvious that the nonparametric bootstrap cannot be used directly in the study of goodness-of-fit test because it does not include any information on the given distribution. From this point of view, the semiparametric bootstrap is more useful. It is the adjusted nonparametric bootstrap that uses the underlying distribution in the nonparametric approach. To see how to choose a suitable p_i for observation under the null hypothesis, see Hall and Wilson (1991), Davison and Hinkley (1996) and Efron and Tibshirani (1993). In the nonparametric bootstrap, any observation has $\frac{1}{n}$ chance of participate in resampling, because the real empirical distribution which gives the appropriate p_i to observations is not known.

Steps in the semiparametric bootstrap are as follows:

1. Suppose $\mathcal{X} = (X_1, \cdots, X_n)$ is an i.i.d. random sample of the distribution F.

2. We are interested in $\mathcal{D}(F)$ and $\mathcal{G}(F)$ which are given in (4) and (19), respectively.

3. Let $\{G_\lambda, \lambda \in \Lambda\}$ be the distribution under H_0. p_i can be found by:

$$p_i = \begin{cases} G((x_{[1]} + x_{[2]})/2) & i = 1 \\ G((x_{[i+1]} + x_{[i]})/2) - G((x_{[i]} + x_{[i-1]})/2) & 2 \leq i \leq n-1 \\ 1 - G((x_{[i]} + x_{[i-1]})/2) & i = n \end{cases} \tag{20}$$

   where G is the CDF of the G_λ and $x_{[i]}$ is the order statistics.

4. For $i = 1, \cdots, B$, resample $(X^*_{i1}, \cdots, X^*_{in})$ from the empirical distribution $G_n(x) = \sum_{i=1}^n p_i I(X_i \leq x)$.

5. Calculate P-value $= \frac{\#\{\mathcal{G}(X^*) > \mathcal{G}(X)\}+1}{B+1}$.

The idea of the semiparametric bootstrap is not given to the bootstrap method in the way that we have considered. For example, Hall et.al (2000) use the bootstrap as continuous rather than as a discrete stochastic process and use the term semiparametric. Furthermore, Good (2004, p. 48) uses the semiparametric test to study the variance as scale parameter in the statistical distribution using the bootstrap method. Actually the term semiparametric is considered for the semiparametric model, not the semiparametric resampling.

In the reminder of this paper, $.SB$ and $.PB$ are added to the names of the statistics concerned ($\mathcal{G}$, $\mathcal{D}$ and $\mathcal{DW}$) to denote the semiparametric and parametric bootstrap version, respectively.

5 Power of tests

This section studies the validity of the tests proposed for the standard normal and exponential distribution, it is done by study the probability of type I error and the statistical power. It presents some Monte Carlo experiments in order to study finite sample properties of tests, implemented by 1000 simulations. In Tables 1-4 the proposed tests are simultaneously based on the same simulated data, which increases the accuracy of the comparisons. The bootstrap resampling is done by $B = 500$. In making a comparative evaluation of testing procedures we seek certain desirable features such as high power and applicability.

5.1 Normal distribution

There are several goodness-of-fit tests for the standard normal distribution, each with its own relative merits. As it is the essential statistical distribution, many powerful tests have been devised, such as the JB and SW test. The study is carried out by simulating i.i.d samples from $N(0, 1)$. In this case, we assume that the mean and variance are known. Obviously, this is unrealistic in practice, although it is a good benchmark for treating realistic cases where the parameters are unknown and have to be estimated. Tables 1-4 include the study of proposed tests in comparison with the JB test and the SW test. It should be mentioned that the first two rows are the JB and SW test, the second two rows are the $\mathcal{G}$ test with $U1 = (q_{10}, q_{25}, q_{50}, q_{70}, q_{85})$ and the third two rows $U2 = (1, 2, 3, 4, 5)$. The last four rows include the semiparametric and parametric bootstrap of $\mathcal{D}$ and $\mathcal{DW}$.

Table 1 includes simulations of type I error at the levels $\alpha = 0.05$ and $\alpha = 0.10$, which is an essential feature of any given statistical test. Tables 2, 3 and 4 show the results of statistical power at the level $1 - \beta = 0.90$ and $1 - \beta = 0.80$. The entries in Table 1 are the ratio of times that normality is rejected when the underlying distribution is the normal whereas in Tables 2-4, the entries are the ratio of times that the normality is rejected when the underlying distribution is not normal. The distributions are $t(df = 5)$, χ^2 $(df = 5)$ and $Exp(\lambda = 1)$, respectively. In terms of type I error, see Table 1, the JB test has the lowest error. Although the $\mathcal{G}$ and $\mathcal{D}$ test are not as strong as the JB test at the level of $\alpha = 0.05$ and $\alpha = 0.10$, they are appropriate. In many cases, the given tests have better results than the SW test. For example the $\mathcal{D}.SP$ test is very close to the JB test and better than the SW test. In this case, the semiparametric bootstrap is better than the parametric bootstrap test.

Table 2 includes the study of statistical power, the underlying distribution is t_5. At first glance can see the SW test has more power than the JB test for small sample size, but for moderate sample, they have same statistical power. Many of the proposed tests have better power than the JB and SW test, for

Table 1: Study of type I error, $\alpha = 0.05$ and 0.10. The underlying distribution is normal.

	$\alpha = 0.05$				$\alpha = 0.1$			
test \ sample size	10	30	50	100	10	30	50	100
SW test	0.049	0.042	0.038	0.052	0.097	0.089	0.093	0.098
JB test	0.008	0.016	0.031	0.037	0.013	0.034	0.044	0.064
$\mathcal{G}.SB$ test[1]	0.009	0.042	0.035	0.061	0.035	0.091	0.071	0.115
$\mathcal{G}.PB$ test[1]	0.042	0.057	0.040	0.049	0.115	0.116	0.075	0.119
$\mathcal{G}.SB$ test[2]	0.014	0.042	0.063	0.075	0.035	0.121	0.134	0.135
$\mathcal{G}.PB$ test[2]	0.044	0.039	0.042	0.046	0.079	0.097	0.088	0.100
$\mathcal{D}.SB$ test	0.012	0.044	0.044	0.072	0.027	0.085	0.073	0.104
$\mathcal{D}.PB$ test	0.045	0.036	0.043	0.050	0.097	0.096	0.071	0.101
$\mathcal{DW}.SB$ test	0.009	0.049	0.050	0.074	0.036	0.121	0.097	0.133
$\mathcal{DW}.PB$ test	0.047	0.037	0.041	0.045	0.097	0.102	0.086	0.097

[1]: $U1 = (q_{10}, q_{25}, q_{50}, q_{70}, q_{85})$; [2]: $U2 = (1, 2, 3, 4, 5)$. Same notation in the tables hereafter.

Table 2: Study of power, $1 - \beta = 0.90$ and 0.80. The underlying distribution is t.

	$1-\beta = 0.90$				$1-\beta = 0.80$			
test\ sample size	10	30	50	100	10	30	50	100
SW test	0.172	0.315	0.425	0.653	0.257	0.437	0.527	0.748
JB test	0.069	0.305	0.438	0.674	0.109	0.364	0.513	0.745
$\mathcal{G}.SB$ test[1]	0.069	0.291	0.431	0.692	0.155	0.367	0.506	0.747
$\mathcal{G}.PB$ test[1]	0.136	0.318	0.445	0.672	0.231	0.390	0.520	0.741
$\mathcal{G}.SB$ test[2]	0.112	0.335	0.404	0.500	0.258	0.469	0.524	0.591
$\mathcal{G}.PB$ test[2]	0.206	0.341	0.393	0.485	0.321	0.456	0.501	0.578
$\mathcal{D}.SB$ test	0.082	0.354	0.495	0.738	0.153	0.426	0.560	0.791
$\mathcal{D}.PB$ test	0.171	0.385	0.494	0.730	0.256	0.450	0.570	0.778
$\mathcal{DW}.SB$ test	0.110	0.373	0.511	0.742	0.198	0.467	0.595	0.797
$\mathcal{DW}.PB$ test	0.193	0.399	0.509	0.723	0.259	0.479	0.594	0.789

example $\mathcal{G}.PB$ and $\mathcal{DW}.PB$. It is obvious that the parametric bootstrap has more ability to diagnose non-normality from normality.

Consider another example of their statistical power. In Table 3, the underlying distribution is χ^2_5, a skewed distribution. Here, the SW test is better than the JB test. Most of the parametric bootstrap has better performance than the JB test.

In Table 4, the underlying distribution is Exp(1). It is obvious that the JB test is quite weak for small sample sizes, whereas the $\mathcal{D}.PB$ test and

Table 3: Study of non-normality, $1-\beta = 0.90$ and 0.80. The underlying distribution is χ^2_5.

	$1-\beta=0.90$				$1-\beta=0.80$			
test\ sample size	10	30	50	100	10	30	50	100
SW test	0.301	0.753	0.934	0.997	0.436	0.865	0.975	1
JB test	0.085	0.504	0.786	0.990	0.128	0.630	0.894	0.997
$\mathcal{G}.SB$ test[1]	0.173	0.758	0.916	0.996	0.365	0.876	0.973	1
$\mathcal{G}.PB$ test[1]	0.280	0.735	0.912	0.994	0.443	0.851	0.962	1
$\mathcal{G}.SB$ test[2]	0.166	0.772	0.925	0.995	0.372	0.879	0.977	1
$\mathcal{G}.PB$ test[2]	0.286	0.736	0.916	0.998	0.449	0.879	0.969	1
$\mathcal{D}.SB$ test	0.109	0.438	0.581	0.823	0.193	0.532	0.677	0.904
$\mathcal{D}.PB$ test	0.216	0.490	0.656	0.926	0.336	0.660	0.852	0.985
$\mathcal{DW}.SB$ test	0.159	0.666	0.880	0.991	0.278	0.807	0.942	0.997
$\mathcal{DW}.PB$ test	0.243	0.670	0.888	0.995	0.380	0.833	0.9649	0.998

Table 4: Study of non-normality, $1-\beta = 0.90$ and 0.80. The underlying distribution is Exp(1).

	$1-\beta=0.90$				$1-\beta=0.80$			
test\ sample size	10	30	50	100	10	30	50	100
SW test	0.551	0.978	1	1	0.689	0.992	1	1
JB test	0.176	0.813	0.981	1	0.246	0.906	0.996	1
$\mathcal{G}.SB$ test[1]	0.261	0.684	0.683	0.621	0.509	0.771	0.736	0.667
$\mathcal{G}.PB$ test[1]	0.420	0.813	0.840	0.878	0.655	0.885	0.894	0.919
$\mathcal{G}.SB$ test[2]	0.301	0.933	0.990	1	0.553	0.985	0.997	1
$\mathcal{G}.PB$ test[2]	0.470	0.958	0.995	1	0.660	0.988	1	1
$\mathcal{D}.SB$ test	0.193	0.670	0.854	0.983	0.284	0.752	0.905	0.995
$\mathcal{D}.PB$ test	0.322	0.768	0.941	1	0.467	0.907	0.989	1
$\mathcal{DW}.SB$ test	0.265	0.896	0.984	1	0.417	0.963	0.998	1
$\mathcal{DW}.PB$ test	0.397	0.938	0.994	1	0.580	0.985	1	1

$\mathcal{DW}.PB$ test have better results. In this case the $\mathcal{G}.SB$ test with $U1 = (q_{10}, q_{25}, q_{50}, q_{70}, q_{85})$ does not have good performance.

The appendix (Figures 1-4) contains the violin plots of the simulated tests which help to illustrate the performance of these tests. The violin plot is a combination of a box plot and a kernel density plot, see Hintze and Nelson (1998). It helps to compare the distribution of simulations. Figure 1 is the violin plot of P-value when the underlying distribution is normal, It is obvious that for a sample size of 10, the JB test has higher P-values than the $\mathcal{D}.SB$ test

Table 5: Study of type I error, $\alpha = 0.05$ and 0.10. The underlying distribution is Exp(1)

	$\alpha = 0.05$				$\alpha = 0.10$			
test\ sample size	10	30	50	100	10	30	50	100
K-S test	0.003	0.010	0.005	0.008	0.019	0.019	0.021	0.014
$\mathcal{G}.SB$ test^1	0.003	0.017	0.017	0.027	0.013	0.027	0.035	0.052
$\mathcal{G}.PB$ test^1	0.005	0.020	0.023	0.042	0.028	0.034	0.037	0.056
$\mathcal{G}.SB$ test^2	0.031	0.063	0.067	0.077	0.083	0.129	0.151	0.139
$\mathcal{G}.PB$ test^2	0.026	0.027	0.028	0.041	0.061	0.079	0.074	0.081
$\mathcal{D}.SB$ test	0.002	0.009	0.019	0.034	0.006	0.014	0.037	0.071
$\mathcal{D}.PB$ test	0.043	0.048	0.039	0.055	0.083	0.097	0.090	0.103
$\mathcal{DW}.SB$ test	0.034	0.044	0.031	0.051	0.057	0.079	0.074	0.081
$\mathcal{DW}.PB$ test	0.035	0.042	0.039	0.052	0.090	0.088	0.085	0.112

and the $\mathcal{G}.SB$ test with $U1$. With increasing the sample size, the accuracy of the JB test decreases. Figure 2 is the violin plot of P-value when the underlying distribution is t_5. The change in performance of the JB test from a sample size of 10 to 30 is interesting, it clarifies that the JB is unreliable for the small sample size. It is obvious that the $\mathcal{G}.SB$ and $\mathcal{G}.PB$ tests with $U2$ have more statistical power than the others. The same results holds for the χ^2_5 and Exp(1), the plots for which are given in the Figures 3 and 4.

5.2 Exponential distribution

Here the same discussions as for the normal distribution is applied to the exponential distribution which has $\mu = \lambda$, $\sigma^2 = \lambda^2$, $\gamma = 2$ and $K = 9$. Tables 5 and 6 show the results for the type I error and the statistical power test, respectively. The entries in Tables 5 and 6 are the Kolmogorov-Smirnov test in comparison with the proposed tests.

Table 5 confirms that the method discussed has an appropriate level of accuracy for the study of the exponential distribution. The $\mathcal{G}.SB$ test with $u1$ and the $\mathcal{D}.SB$ test perform better than the others.

Table 6 includes the statistical power of the tests when the underlying distribution is χ^2_5. This agrees with the efficiency of the proposed tests, not for $DW.SB$, and confirms that the $\mathcal{G}$ and $\mathcal{D}$ tests have quite high statistical power to diagnosis the exponential distribution from the chi-square distribution. Tables 6 also shows that the $\mathcal{DW}.SB$ test is not appropriate. This is expected, because p_1, (20), gives high weight to the first observation, hence the generated samples from semiparametric bootstrap seems come from the

Table 6: Study of power, at the level $1-\beta=0.90$ and 0.80, the underlying distribution is χ^2_5

	$1-\beta=0.90$				$1-\beta=0.80$			
test\ sample size	10	30	50	100	10	30	50	100
K-S test	0.135	0.663	0.930	1.000	0.322	0.847	0.982	1.000
$\mathcal{G}.SB$ test[1]	0.340	0.849	0.957	0.998	0.541	0.939	0.985	0.998
$\mathcal{G}.PB$ test[1]	0.063	0.550	0.866	0.994	0.522	0.885	0.966	0.998
$\mathcal{G}.SB$ test[2]	0.250	0.783	0.943	0.997	0.438	0.920	0.980	0.998
$\mathcal{G}.PB$ test[2]	0.018	0.425	0.794	0.993	0.399	0.850	0.958	0.998
$\mathcal{D}.SB$ test	0.055	0.603	0.938	1.000	0.390	0.925	0.980	1.000
$\mathcal{D}.PB$ test	0.018	0.722	0.931	0.996	0.521	0.940	0.984	1.000
$\mathcal{DW}.SB$ test	0.004	0.004	0.008	0.018	0.014	0.006	0.019	0.024
$\mathcal{DW}.PB$ test	0.313	0.532	0.655	0.827	0.505	0.702	0.788	0.901

exponential distribution, although it is not problem for parametric distribution. It should be mentioned that $D.SP$ does not include the weight hence it does not affect by the appropriate value of P_1.

Figures 5 and 6 in the appendix show the violin plots of the P-value when the underlying distribution is Exp(1) and χ^2_5, respectively.

6 Conclusion

There are always some advantages and disadvantages for any given statistics. These statistics can be easily calculated and implemented into statistical softwares and have appropriate power in comparison with the discussed tests.

Although comparison of the SW test and JB test was not our object, it is obvious that the JB test has less type I error, but the SW test has more statistical power than JB.

In the case of $\mathcal{G}$, it has appropriate accuracy, although it is rather difficult to give an optimal u in general, it obvious that $u1$ and $u2$ perform well for the exponential and normal distribution, respectively. $\mathcal{G}$ has fairly interesting result and performs well. The $\mathcal{G}.PB$ performs better than $\mathcal{G}.SB$ for the study of normality, for the study of exponential distribution, it is inverse.

In the case of D, which is extracted directly based on JB, the result of study of normal is quite good. But the $\mathcal{DW}.\mathcal{SB}$ test cannot perform well for study of the exponential distribution, although in the normal case it works well.

The bootstrapped version of JB, $DW.SP$, has high statistical power, rendering it suitable for use along with the JB test in the study of normal distribution. However in the non-normal distribution, $\mathcal{G}$ test is appropriate.

The new bootstrapped test, $\mathcal{G}$, proposed includes simultaneously moments and can thus be considered as a multiparametric test. The procedure for this test can also be considered an extension of work on the bootstrap method, which shows the applicability of semiparametric bootstrap, specially in the normal distribution.

It is arguably possible to find procedures that are more sensitive to some distribution properties, for example can add more equations by using other parameters such as the fifth power of centered moment, if it exists.

Following the outlines described, the viewpoint of generality are more interesting, as it is discussed for the exponential distribution along with the normal distribution. As previously mentioned, the JB test does not directly use the mean and variance which include information on the observations and also the assumptions of distribution. These parameters are used in the new tests presented here and are a features of these tests that seems to be particularly relevant. The most obvious property of a statistical procedure is that it should be trustworthy, as the simulated type I error and the statistical power show. These results highlight the benefits of the bootstrap approach, which are more apparent, even richer and probably is less well understood.

References

[1] Cramér, H. (1945). *Mathematical Methods of Statistics.* Almqvist & Wiksells, Uppsala.

[2] D'Agostino, R.B. and Stephen, M.A. (1986). *Goodness-Of-Fit Techniques.* New York: Marcel Dekker.

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

[4] Efron, B. and Tibshirani, R. (1993). *Introduction to the Bootstrap.* Chapman & Hall, New York.

[5] Hall, P., Härdle, W., Kleinow, T., and Schmidt, P. (2000). Semiparametric bootstrap approach to hypothesis tests and confidence intervals for the hurst coefficient. *Statistical Inference for Stochastic Processes.* **3**(3), 263-276.

[6] Hall, P. and Wilson, R.S. (1991). Two guidelines for bootstrap hypothesis testing. *Biometrics*, **47**(2), 757-762.

[7] Hintze, J.L. and Nelson, R.D. (1998). Violin plots: a box plot-density trace synergism. *The American Statistician*, **52**(2), 181-184.

[8] Kendall, W.S., Barndorff-Nielson, O., and van Lieshout, M.C. (1998). *Current Trends in Stochastic Geometry: Likelihood and Computation.* Boca Raton, FL. CRC Press.

[9] Kendall, M,G. and Stuart. A. (1973). *The Advanced Theory of Statistics.* Vol 1. Charles Griffin & Company Limited London.
[10] Kilian, L. and Demiroglu, U. (2000) Residual-based tests for normality in autoregressions: asymptotic theory and simulation evidence. *Journal of Business & Economic Statistics*, **18**(1), 40-50.
[11] Jarque, W.M. and Bera, A.K. (1987). Efficient tests for normality, homoscedasticity and serial independence of regression residuals, *Economics Letters* **6**, 255-259.
[12] Jarque, W M. and Bera, A.K. (1987). A test for normaility of observations and regression residuals. *International Statistical Review*, **55**(2), 163-172.
[13] Rayner, J.C.W. and Best, D.J. (1989). *Smooth Tests of Goodness of Fit.* Oxford University Press. New York.
[14] Shapiro, S.S., Wilk, M.B., and Chen, M.H.J. (1968). A comparative study of various tests for normality. *Journal of the American Statistical Association*, **63**(324), 1343-1372.
[15] Thode, Jr., H.C. (2002). *Testing for Normality*, Marcel Dekker, New York.

Appendix

Here the violin plots are given for the study of simulations.

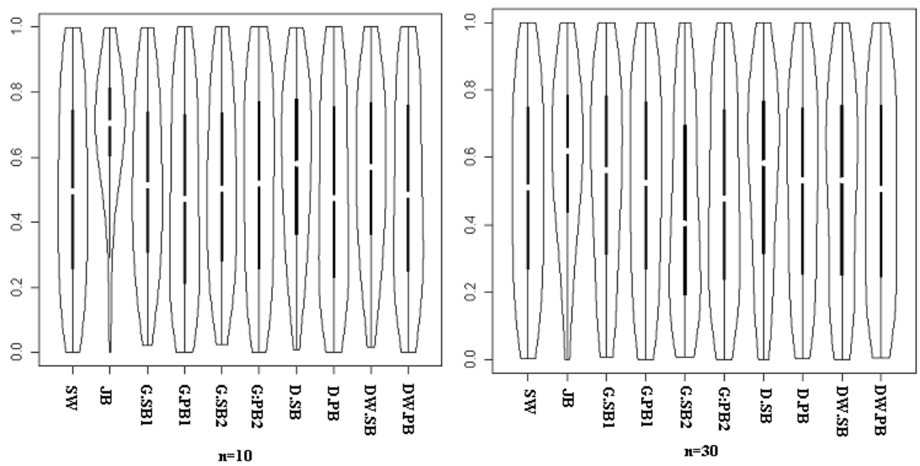

Figure 1: Violin plot of the simulated P-value when distribution is normal.

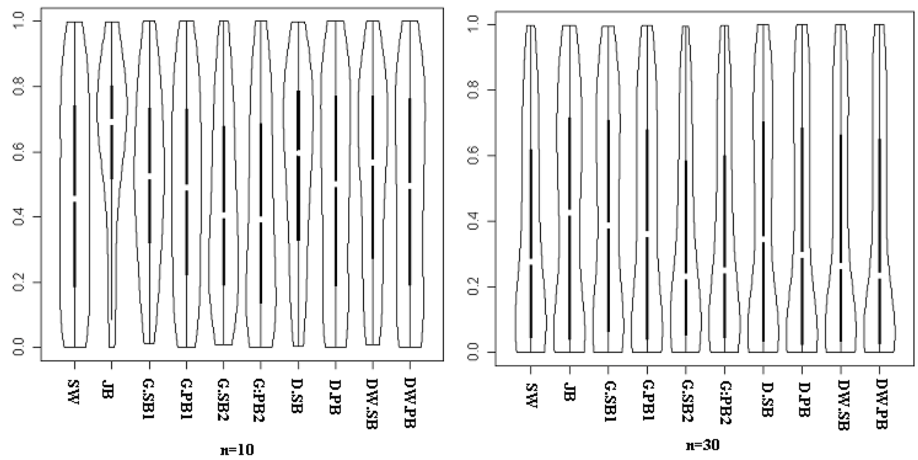

Figure 2: Violin plot of the simulated P-value for the study of normality when the distribution is t_5.

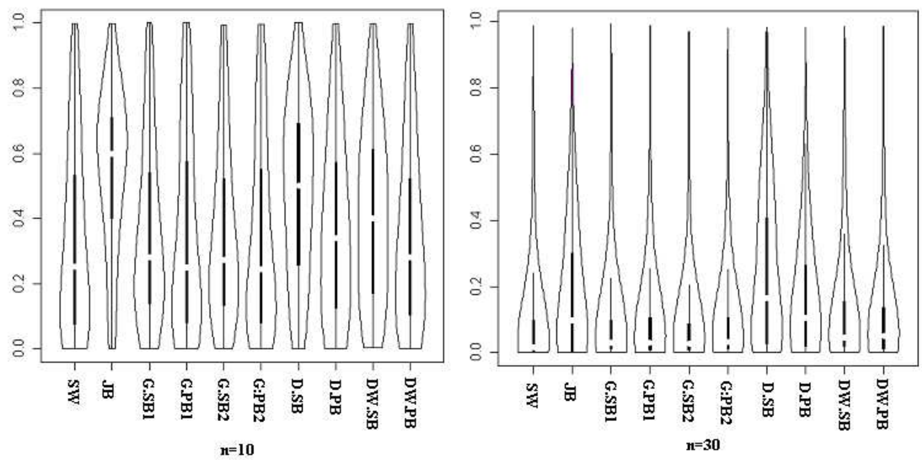

Figure 3: Violin plot of the simulated P-value for the study of normality when the distribution is χ^2_5.

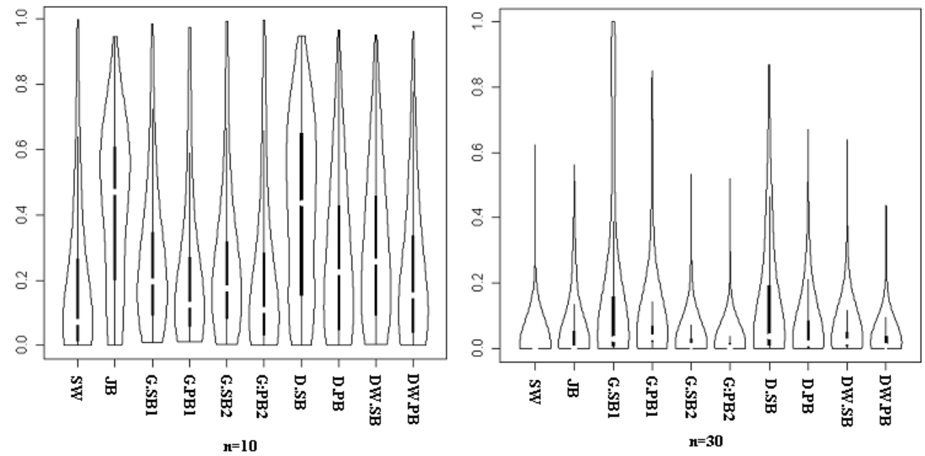

Figure 4: Violin plot of the simulated P-value for the study of normality when the distribution is Exp(1).

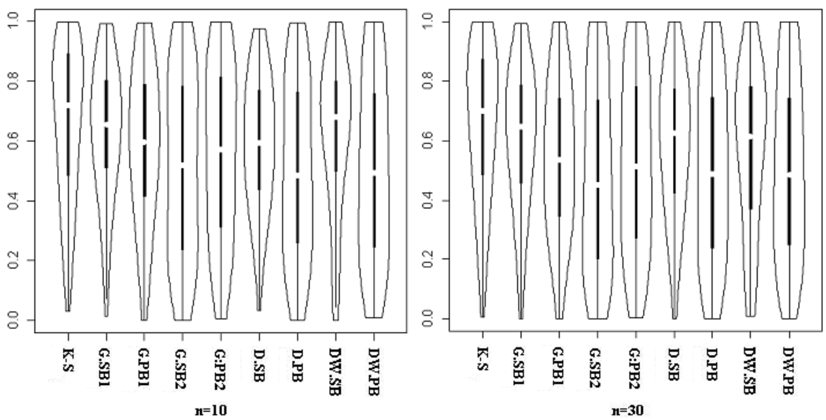

Figure 5: Violin plot of the simulated P-value for the study Exp(1) when the underlying distribution is Exp(1).

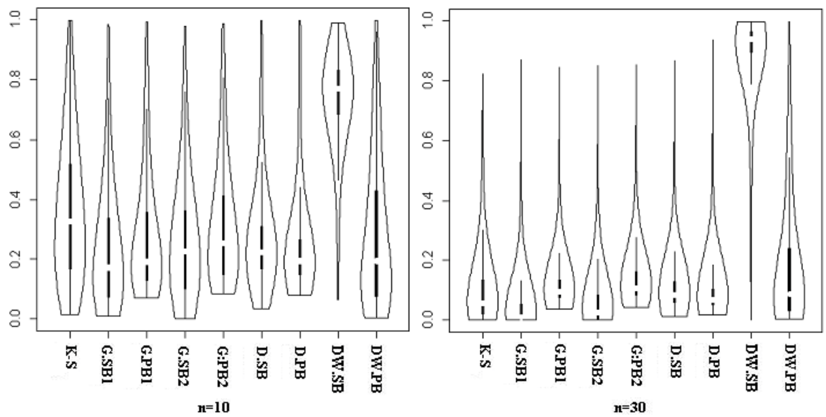

Figure 6: Violin plot of the simulated P-value for the study Exp(1) when the underlying distribution is χ^2_5.