

# Selection and Validation of Comparative Study of Normality Test

Babalola M. Afeez<sup>1,\*</sup>, Obubu Maxwell<sup>2</sup>, Oluwaseun A. Otekunrin<sup>3</sup>, Obiora-Ilono Happiness<sup>2</sup>

<sup>1</sup>Department of Statistics, University of Ilorin, Ilorin, Nigeria

<sup>2</sup>Department of Statistics, Nnamdi Azikiwe University, Awka, Nigeria

<sup>3</sup>Department of Statistics, University of Ibadan, Nigeria

**Abstract** The goal of this study is to investigate the best goodness-of-fit test among five selected normality tests under various continuous non-normal distributions using power as criteria. The tests were compared when the normal parameters are unknown and sample sizes are 10, 30, 50, 100, 300, 500 and 1000 were iterated 1000 times each with 0.01, 0.05, and 0.10 level of significance, using the Monte Carlo technique. We study the procedures based on five well-known normality tests: the Anderson–Darling, Cramer–von Mises, Shapiro–Wilk, Jarque–Bera and Chi-Square. Evidence from the simulation study reveals that the performance of the five normality test statistics varies with the level of significance, sample size and alternative distributions.

**Keywords** Normal Distributions, Shapiro-Wilk test (SW), Cramer von Misses test (CVM), Jarque-Bera test (JB), Chi-Square test (CSQ) and Anderson-Darling test (AD)

## 1. Introduction

Tests of normality are statistical inference procedures design to verify that an underlying distribution of a random variable is normally distributed. The problem of testing normality is fundamental in both theoretical and empirical research. Indeed, the validity of parametric statistical inference procedures in finite samples (in the sense that their size is controlled) depends crucially on the underlying distributional assumptions. Consequently, there has been extensive focus on whether hypothesized distributions are compatible with the data. Tests of normality are particularly prevalent because the assumption of normality is quite often made in statistical analysis, e.g. in econometric studies. In this respect, the reviews by D'Agostino and Stephens (1986, Ch. 9) and Dufour et al. (1998) report nearly 40 different normality tests.

However, the tests of normality are based on different characteristics of the normal distribution and the power of these tests differs depending on the nature of non-normality (Seier, 2002). And a test is said to be powerful when it has a high probability of rejecting the null hypothesis of normality when the sample under study is taken from a non-normal distribution.

The effort of developing techniques to detect departures from normality was initiated by Pearson (1895) who worked on the skewness and kurtosis coefficients (Althouse et al., 1998). Tests of normality differ in the characteristics of the normal distribution they focus on, such as its skewness and kurtosis values, its distribution or characteristic function, and the linear relationship existing between the distribution of the variable and the standard normal variable,  $Z$ . The tests also differ in the level at which they compare the empirical distribution with the normal distribution, in the complexity of the test statistic and the nature of its distribution (Seier, 2002).

However, Seier (2002) classified the tests of normality into four major sub-categories which are skewness and kurtosis test, empirical distribution test, regression and correlation test and other special test. Arshad et al. (2003) also categorized the tests of normality into four major categories which are tests of chi-square types, moment ratio techniques, tests based on correlation and tests based on the empirical distribution function.

As it has been stated above that there are various normality tests in the literature. Some are either too simple to be powerful or too difficult to apply. Most of them are only applicable for certain situations due to various kinds of biases. But, this study is mainly focused on comparing the power of five normality tests, which are: Shapiro-Wilk test (SW), Cramer von Misses test (CVM), Jarque-Bera test (JB), Chi-Square test (CSQ) and Anderson-Darling test (AD). These formal normality tests were selected for two reasons. First, they are among the best and commonly used normality tests, Arshad et al. (2003) and Seier (2002). Second, they are

\* Corresponding author:

babalolaafeez222@gmail.com (Babalola M. Afeez)

Published online at <http://journal.sapub.org/ajms>

Copyright © 2018 The Author(s). Published by Scientific & Academic Publishing

This work is licensed under the Creative Commons Attribution International

License (CC BY). <http://creativecommons.org/licenses/by/4.0/>

available in most statistical packages, and are widely used in practice (Yap & Sim. 2011).

## 2. Simulation Methodology

### 2.1. The Normal Distribution

One of the most, if not the most, used distribution in statistical analysis is the normal distribution. The first time the normal distribution appeared was in one of the works of Abraham De-Moivre (1667–1754) in 1733, when he investigated large-sample properties of the binomial distribution. He discovered that the probability for sums of binomially distributed random variables to lie between two distinct values follows approximately a certain distribution the distribution we today call the normal distribution.

Another consideration which favors the normal distribution is the fact that sampling distributions based on a parent normal distribution are fairly manageable analytically. In making inferences about populations from samples, it is necessary to have the distributions for various functions of the sample observations.

Gaussian distribution is one of the most important continuous distributions. It is bell-shaped and symmetric about its mean.

Let  $X$  be normally distributed i.e.  $X \sim N(\mu, \sigma^2)$

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp \frac{(x-\mu)^2}{2\sigma^2} \text{ if } -\infty < x < \infty, \sigma^2 > 0.$$

The parameters  $\mu$  and  $\sigma^2$  are the mean and variance respectively. The mean is the same as the median and the mode.

### 2.2. Tests of Normality

It may be useful to summarize very briefly previous work in so far as it is strictly relevant to this study. Throughout the years several statistics have been proposed to test the assumption of normality. The hypotheses in question can be written as follows:

$$H_0 : X_i \sim N(\mu, \sigma^2) \text{ for } i=1, \dots, n$$

$Vs$

$$H_1 : X_i \sim F_1 \text{ for } i = 1, \dots, n$$

Where  $F_1$  denotes a distribution different from the normal distribution. We proceed by reviewing a few of the more common testing procedures.

### 2.3. Empirical Distribution Function (EDF) Tests

The idea of the EDF tests in testing normality of data is to compare the empirical distribution function which is estimated based on the data with the cumulative distribution function (CDF) of normal distribution to see if there is a good agreement between them. Dufour et al. (1998) described EDF tests as those based on a measure of discrepancy between the empirical and hypothesized distributions. The EDF tests can be further subdivided into those belong to supremum and square class of the

discrepancies. Arshad et al. (2003) and Seier (2002) claimed that the most crucial and widely known EDF tests are Kolmogorov-Smirnov, Anderson-Darling and Cramer Von Mises tests.

### 2.4. Regression and Correlation Tests

Dufour et al. (1998) defined correlation tests as those based on the ratio of two weighted least-squares estimates of scale obtained from order statistics. The two estimates are the normally distributed weighted least squares estimates and the sample variance from other population. Some of the regression and correlation tests are Shapiro-Wilk test, Shapiro-Francia test and Ryan-Joiner test.

### 2.5. Moment Tests

In addition to the types of normality test categorized by Seier (2002) above, there are also other types of normality test. One of these types is called the moment tests. Moment tests are those derived from the recognition that the departure of normality may be detected based on the sample moments which are the skewness and kurtosis. The procedures for individual skewness and kurtosis tests can be found in D'Agostino and Stephens (1986). The two most widely known are the tests proposed by D'Agostino-Pearson (1973) and Jarque-Bera (1987).

### 2.6. Chi Square Test

The oldest and most well-known goodness-of-fit test is the CSQ test for goodness of fit, first presented by Pearson (1900). However, the CSQ test is not highly recommended for continuous distribution since it uses only the counts of observations in each cell rather than the observations themselves in computing the test statistic.

## 3. Simulation Result

This study employs Monte Carlo experiment to empirically investigate the power of SW, AD, CSQ, CVM, and JB test statistics to verify if a random sample  $x$  of size  $n$  independent observations come from a normal population. The random samples were generated independently in R environment. In which SW, AD, CSQ, CVM and JB were accessed via *Shapiro.test*, *ad.test*, *pearson.test*, *cvm.test* and *jarque.bera.test* functions respectively. Since there is no standard setting that has been established in research for empirical simulation studies for tests of normality. In the conducted studies of this work we tried to choose our parameters as conventional as possible.

Seven different sample sizes  $n = (10, 30, 50, 100, 300, 500 \& 1000)$  relating to small, medium, and large sample sizes were iterated over 1000 times from each of the selected alternative distributions and the levels of significance considered were 1%, 5% and 10%.

The simulated samples were generated in order to test the hypotheses as follow:

$$H_0: X_i \sim N(\mu, \sigma^2) \text{ for } i=1, \dots, n$$

vs

$$H_1: \text{Not } H_0$$

The alternative non-normal distributions were classified in accordance to Shapiro et al. (1968), which proposed the classification of continuous non-normal distributions by their skewness ( $\sqrt{\beta_1}$ ) and kurtosis ( $\beta_2$ ) into five categories, which are: Near normal distribution, Symmetric long-tailed distribution, Symmetric short-tailed distribution, Asymmetric long-tailed distribution, and Asymmetric short-tailed distribution.

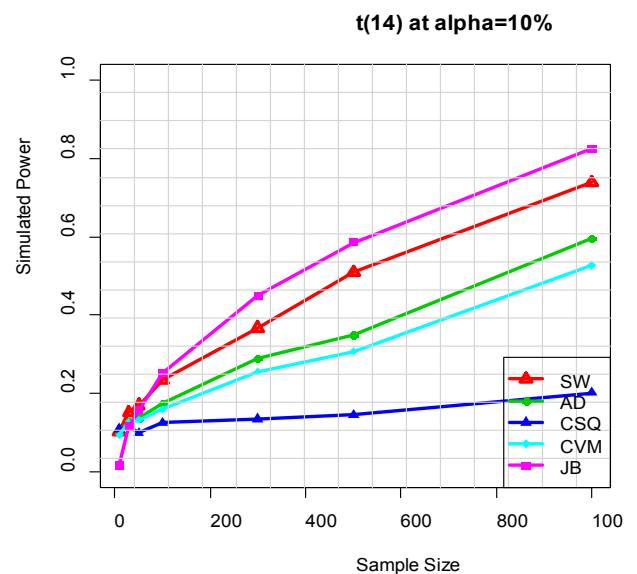
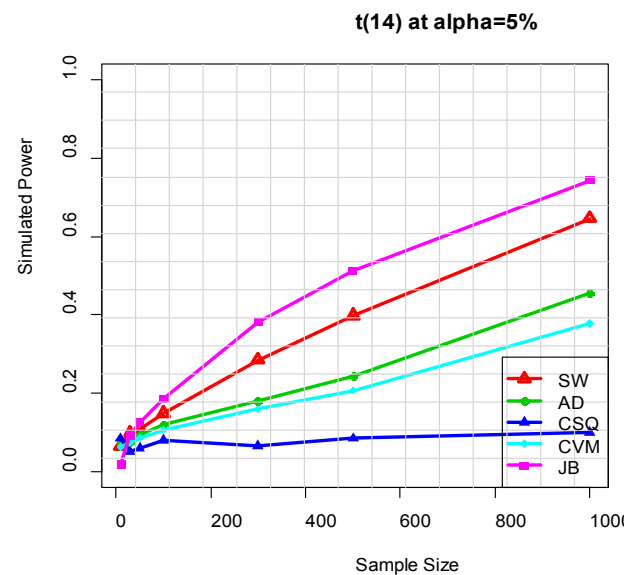
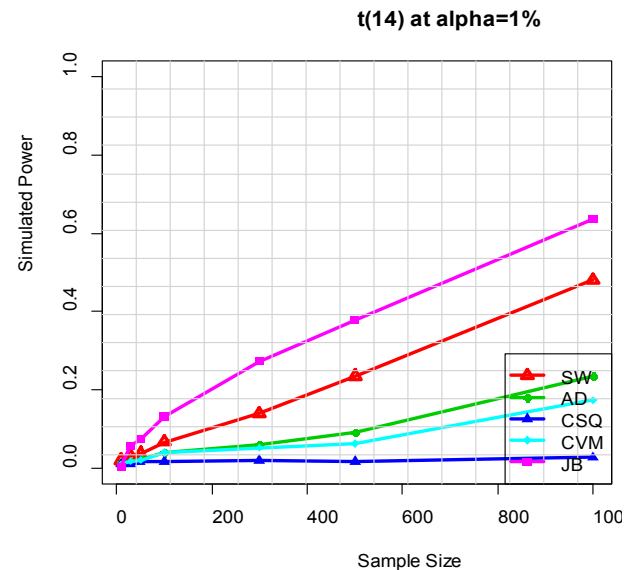
Then the different non-normal distributions considered in this study are essentially a subset of those investigated by Shapiro et al. (1968), Chaichatschwal & Budsaba (2007) and Yap & Sim (2010), in order to complement their results and examine whether the present study agrees or disagrees with previous ones.

**Table 1.** Classification of distributions under study

Skewness: $\sqrt{\beta_1}$ Kurtosis: $\beta_2$	Case	Distributions used in the study
$\sqrt{\beta_1} \leq .3$ , $2.5 \leq \beta_2 \leq 4.5$	Near Normal	t (14), t(10)
$\sqrt{\beta_1} \leq .3$ , $\beta_2 > 4.5$	Symmetric long-tailed	t (7), t(5)
$\sqrt{\beta_1} \geq .3$ , $\beta_2 < 2.5$	Symmetric short-tailed	U (0,1), Beta (1.5,1.5)
$\sqrt{\beta_1} > .3$ , $\beta_2 > 3.0$	Asymmetric long-tailed	Weibull (2.211,1), Beta (7,2)
$\sqrt{\beta_1} > .3$ , $\beta_2 < 3.0$	Asymmetric short-tailed	Beta (2,1), Beta (2,4)

**Table 2.1.** Simulated Power of the near normal distributions

Alternative distribution	Sample size (n)	Normality tests				
		SW	AD	CSQ	CVM	JB
t(14) at $\alpha=1\%$	10	0.016	0.017	0.013	0.015	0.007
	30	0.033	0.020	0.010	0.017	0.055
	50	0.041	0.027	0.014	0.028	0.074
	100	0.078	0.034	0.015	0.027	0.143
	300	0.135	0.064	0.022	0.050	0.241
	500	0.278	0.124	0.020	0.103	0.416
	1000	0.487	0.245	0.031	0.191	0.635
t(14) at $\alpha=5\%$	10	0.068	0.069	0.040	0.067	0.023
	30	0.095	0.077	0.050	0.075	0.086
	50	0.087	0.071	0.057	0.061	0.110
	100	0.154	0.109	0.059	0.105	0.184
	300	0.271	0.163	0.068	0.142	0.370
	500	0.416	0.250	0.077	0.211	0.501
	1000	0.685	0.495	0.118	0.409	0.770
t(14) at $\alpha=10\%$	10	0.121	0.131	0.105	0.117	0.036
	30	0.159	0.158	0.133	0.147	0.111
	50	0.185	0.164	0.099	0.149	0.173
	100	0.226	0.178	0.124	0.156	0.247
	300	0.405	0.282	0.147	0.237	0.467
	500	0.509	0.372	0.130	0.317	0.578
	1000	0.758	0.595	0.176	0.505	0.830

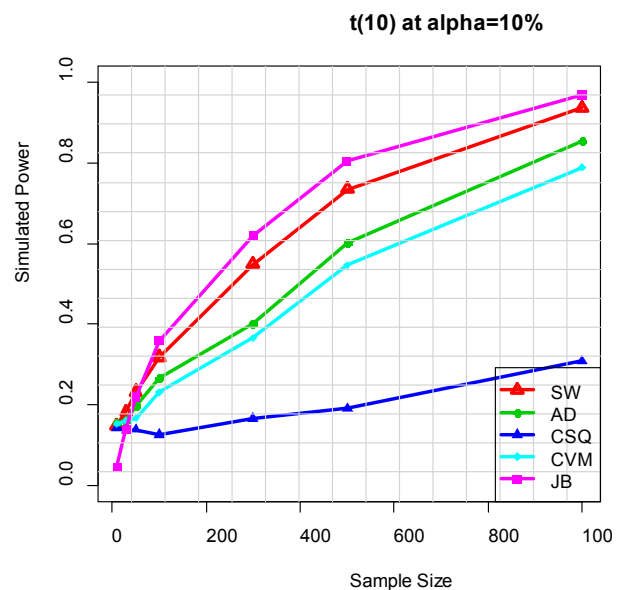
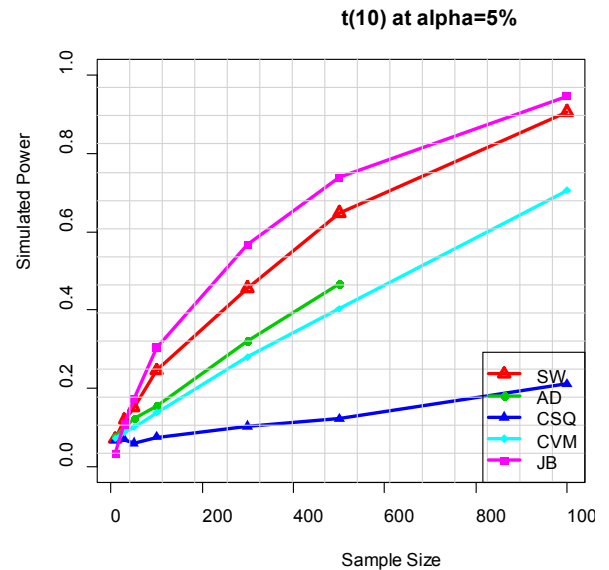
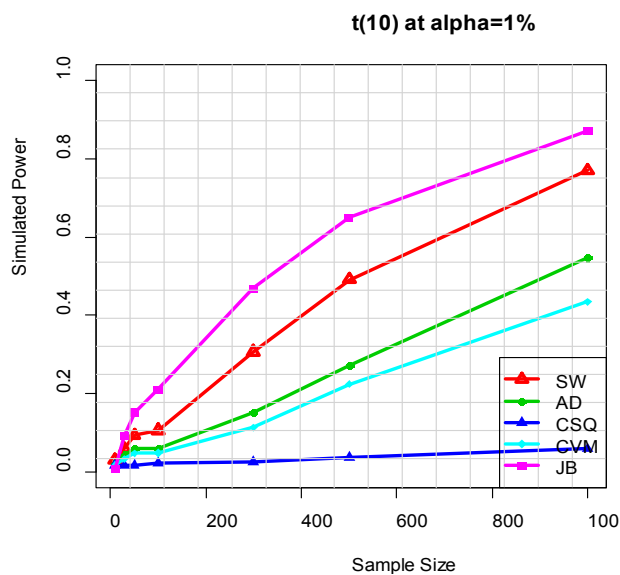


**Figure 1.1.** Simulated Power of the near normal distributions at various alpha levels

Table 2.1 show the simulated power for selected Near normal distributions, while Figure 1.1 show the plot of power for all tests against selected near normal distributions for 1%, 5% and 10% significance levels. Which reveal that: for Near Normal distributions, the power of JB have better power compared to others, at all significance levels considered in the study. And the result also shows that the power of CSQ is very poor and it required much larger sample size to achieve comparable power with other tests. It can also be observed that as alpha increases, the power of the tests also increased.

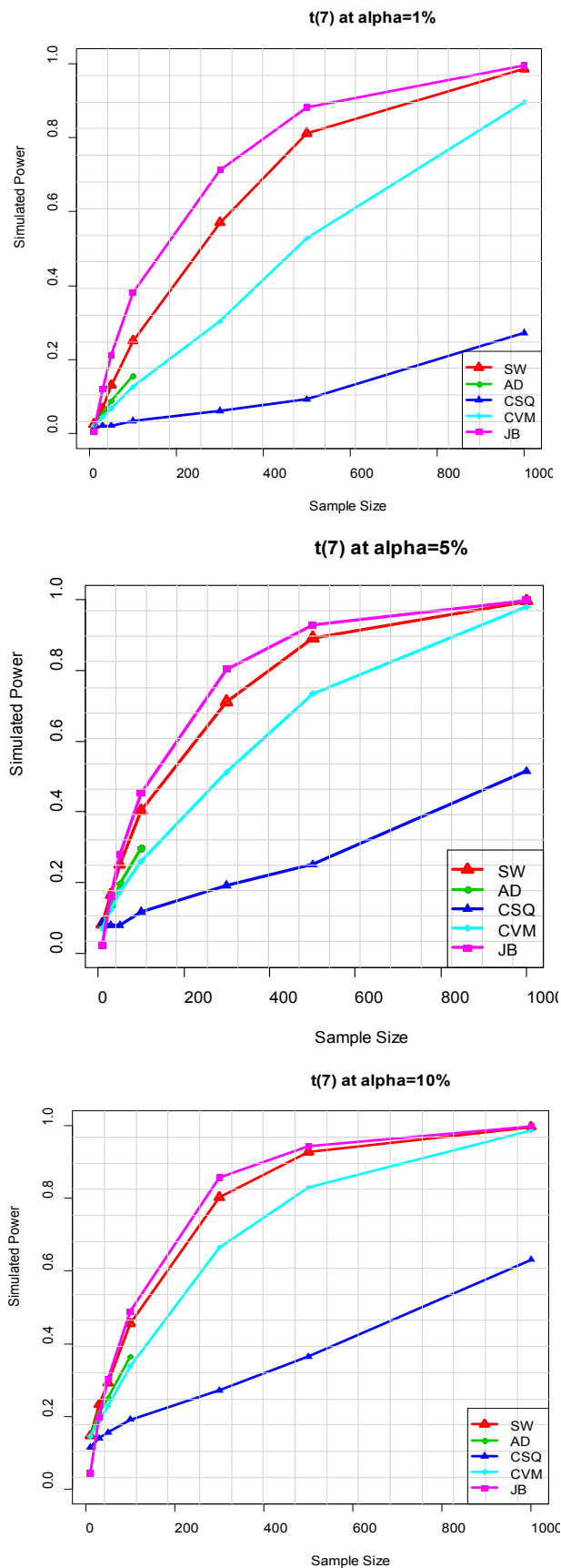
**Table 2.2.** Power for all tests against selected near normal distributions

Alternative distribution	Sample size (n)	Normality tests				
		SW	AD	CSQ	CVM	JB
t(10) at $\alpha=1\%$	10	0.029	0.019	0.017	0.016	0.005
	30	0.058	0.043	0.017	0.033	0.092
	50	0.093	0.059	0.017	0.047	0.151
	100	0.106	0.060	0.022	0.048	0.210
	300	0.306	0.151	0.025	0.113	0.469
	500	0.491	0.273	0.036	0.224	0.650
	1000	0.770	0.547	0.058	0.434	0.871
t(10) at $\alpha=5\%$	10	0.070	0.077	0.066	0.072	0.032
	30	0.118	0.100	0.069	0.087	0.108
	50	0.151	0.121	0.059	0.100	0.173
	100	0.245	0.155	0.075	0.137	0.303
	300	0.456	0.320	0.103	0.279	0.566
	500	0.648	0.466	0.122	0.403	0.739
	1000	0.906	NA	0.211	0.706	0.947
t(10) at $\alpha=10\%$	10	0.147	0.155	0.142	0.152	0.044
	30	0.182	0.161	0.137	0.161	0.139
	50	0.232	0.197	0.138	0.166	0.220
	100	0.318	0.267	0.126	0.231	0.359
	300	0.548	0.402	0.166	0.365	0.619
	500	0.735	0.601	0.190	0.546	0.805
	1000	0.937	0.854	0.308	0.789	0.969



**Figure 1.2.** Power for all tests against selected near normal distributions

Table 2.2 and Figure 1.2 also show the plot of power for all tests against selected near normal distributions for 1%, 5% and 10% significance levels. Which reveal that: for Near Normal distributions, the power of JB have better power compared to others, follow by SW at all significance levels considered in the study. And the result also show that the power of CSQ is very poor and it required much larger sample size to achieve comparable power with other tests. It can also be observed that as alpha increases, the power of the tests also increased.



**Figure 2.1.** Power for all tests against selected Symmetric long-tailed distribution

**Table 3.1.** Simulated Power of the Symmetric long- tailed distributions

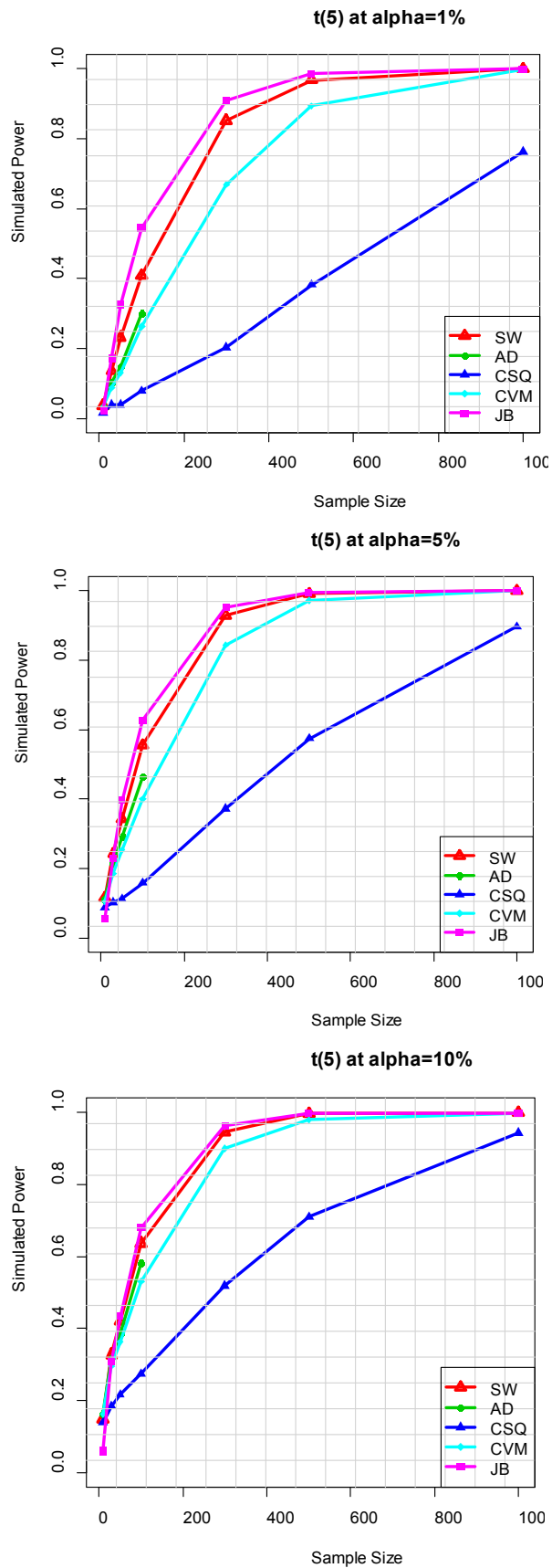
Alternative distribution	Sample size (n)	Normality tests				
		SW	AD	CSQ	CVM	JB
t(7) at $\alpha=1\%$	10	0.025	0.020	0.013	0.021	0.006
	30	0.067	0.052	0.021	0.044	0.118
	50	0.130	0.088	0.021	0.068	0.212
	100	0.250	0.157	0.034	0.126	0.381
	300	0.570	NA	0.061	0.303	0.712
	500	0.812	NA	0.092	0.527	0.882
	1000	0.985	NA	0.272	0.896	0.995
t(7) at $\alpha=5\%$	10	0.081	0.078	0.085	0.070	0.022
	30	0.165	0.135	0.079	0.122	0.161
	50	0.251	0.197	0.079	0.174	0.279
	100	0.403	0.296	0.117	0.259	0.453
	300	0.711	NA	0.192	0.512	0.803
	500	0.892	NA	0.251	0.734	0.928
	1000	0.996	NA	0.515	0.979	0.999
t(7) at $\alpha=10\%$	10	0.146	0.148	0.115	0.145	0.043
	30	0.233	0.208	0.140	0.188	0.199
	50	0.292	0.251	0.156	0.229	0.303
	100	0.455	0.366	0.191	0.338	0.489
	300	0.803	NA	0.271	0.665	0.857
	500	0.927	NA	0.364	0.829	0.942
	1000	0.996	NA	0.631	0.985	0.998

Table 3.1 shows the simulated power for selected Symmetric long-tailed distributions, while Figure 2.1 show the plot of power for all tests against selected Symmetric long-tailed distribution for 1%, 5% and 10% significance levels. Which show that: for Symmetric long-tailed distributions, the power of JB is the best, follow closely by SW at all level of significance. And also show that the power of CSQ is very low compares to other tests.

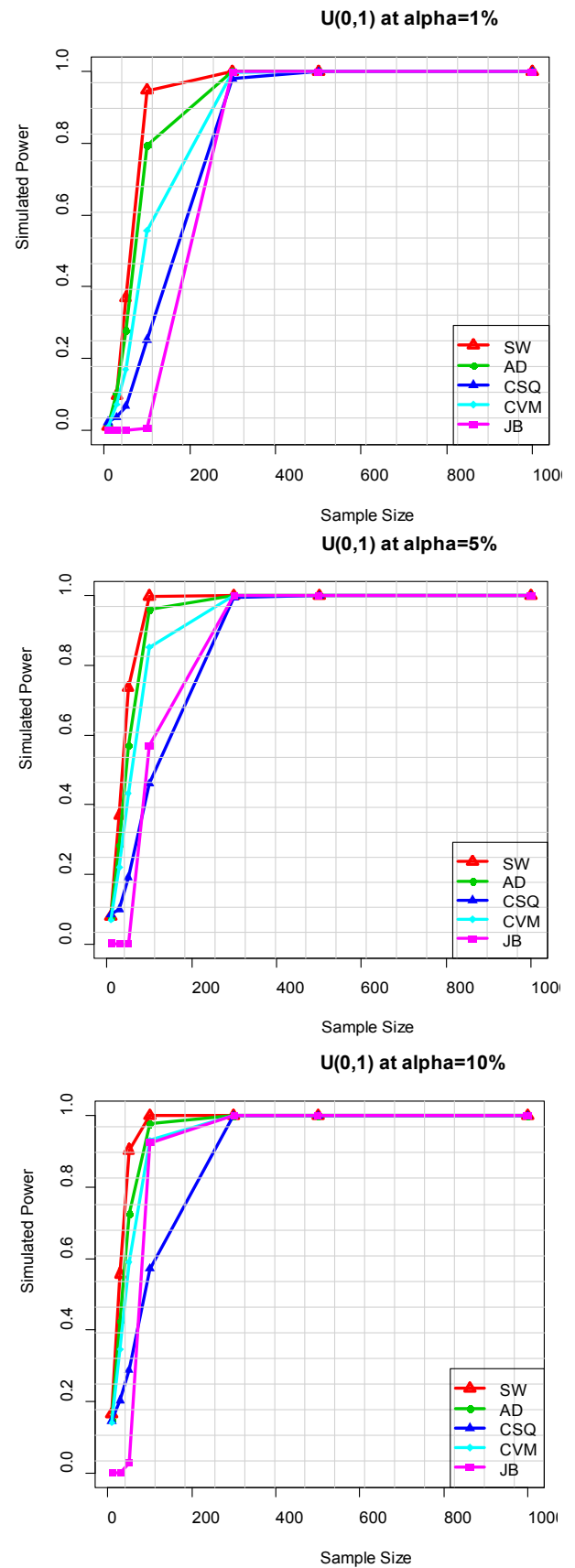
**Table 3.2.** Power for all tests against selected Symmetric long-tailed distribution

Alternative distribution	Sample size (n)	Normality tests				
		SW	AD	CSQ	CVM	JB
t(5) at $\alpha=1\%$	10	0.037	0.034	0.017	0.032	0.021
	30	0.135	0.103	0.039	0.087	0.171
	50	0.232	0.149	0.040	0.130	0.326
	100	0.409	0.300	0.079	0.263	0.544
	300	0.852	NA	0.203	0.669	0.909
	500	0.967	NA	0.382	0.893	0.985
	1000	1.000	NA	0.762	0.998	1.000
t(5) at $\alpha=5\%$	10	0.115	0.110	0.087	0.104	0.054
	30	0.242	0.213	0.102	0.185	0.229
	50	0.343	0.291	0.113	0.254	0.396
	100	0.555	0.463	0.158	0.400	0.626
	300	0.929	NA	0.372	0.844	0.952
	500	0.991	NA	0.574	0.972	0.994
	1000	1.000	NA	0.896	1.000	1.000
t(5) at $\alpha=10\%$	10	0.148	0.162	0.138	0.162	0.059
	30	0.326	0.316	0.184	0.295	0.307
	50	0.421	0.387	0.216	0.363	0.433
	100	0.636	0.581	0.274	0.531	0.680
	300	0.945	NA	0.519	0.900	0.962
	500	0.997	NA	0.710	0.980	0.998
	1000	0.999	NA	0.943	0.999	0.999

NA means Not Applicable



**Figure 2.2.** Power for all tests against selected Symmetric long-tailed distribution



**Figure 3.1.** Simulated Power of the Symmetric short tailed distributions

**Table 4.1.** Simulated Power of the Symmetric short tailed distributions

Alternative distribution	Sample size (n)	Normality tests				
		SW	AD	CSQ	CVM	JB
U(0,1) at $\alpha=1\%$	10	0.011	0.012	0.022	0.011	0.000
	30	0.098	0.088	0.038	0.060	0.000
	50	0.355	0.256	0.077	0.164	0.000
	100	0.952	0.792	0.287	0.584	0.000
	300	1.000	1.000	0.988	1.000	1.000
	500	1.000	1.000	1.000	1.000	1.000
	1000	1.000	1.000	1.000	1.000	1.000
U(0,1) at $\alpha=5\%$	10	0.078	0.083	0.084	0.078	0.002
	30	0.379	0.299	0.103	0.224	0.000
	50	0.723	0.564	0.181	0.432	0.000
	100	0.996	0.956	0.448	0.861	0.548
	300	1.0000	1.0000	0.994	1.0000	1.0000
	500	1.0000	1.0000	1.0000	1.0000	1.0000
	1000	1.0000	1.0000	1.0000	1.0000	1.0000
U(0,1) at $\alpha=10\%$	10	0.179	0.171	0.144	0.149	0.004
	30	0.572	0.433	0.211	0.357	0.002
	50	0.877	0.730	0.283	0.600	0.025
	100	0.999	0.979	0.581	0.918	0.909
	300	1.0000	1.0000	0.9960	1.0000	1.0000
	500	1.0000	1.0000	1.0000	1.0000	1.0000
	1000	1.0000	1.0000	1.0000	1.0000	1.0000

**Table 4.2.** Power for all tests against selected symmetric short-tailed distributions

Alternative distribution	Sample size (n)	Normality tests				
		SW	AD	CSQ	CVM	JB
Beta(1.5,1.5) at $\alpha=1\%$	10	0.015	0.016	0.018	0.014	0.000
	30	0.013	0.027	0.019	0.021	0.000
	50	0.086	0.083	0.041	0.059	0.000
	100	0.426	0.281	0.085	0.185	0.000
	300	1.000	0.981	0.499	0.869	0.954
	500	1.000	1.000	0.919	0.997	1.000
	1000	1.000	1.000	1.000	1.000	1.000
Beta(1.5,1.5) at $\alpha=5\%$	10	0.049	0.049	0.053	0.046	0.003
	30	0.135	0.115	0.068	0.093	0.001
	50	0.352	0.261	0.108	0.206	0.000
	100	0.807	0.590	0.196	0.441	0.094
	300	1.000	0.999	0.721	0.978	1.000
	500	1.000	1.000	0.967	1.000	1.000
	1000	1.000	1.000	1.000	1.000	1.000
Beta(1.5,1.5) at $\alpha=10\%$	10	0.128	0.122	0.139	0.117	0.008
	30	0.277	0.226	0.148	0.205	0.001
	50	0.505	0.366	0.184	0.300	0.003
	100	0.913	0.756	0.265	0.624	0.458
	300	1.000	0.999	0.814	0.991	1.000
	500	1.000	1.000	0.984	1.000	1.000
	1000	1.000	1.000	1.000	1.000	1.000

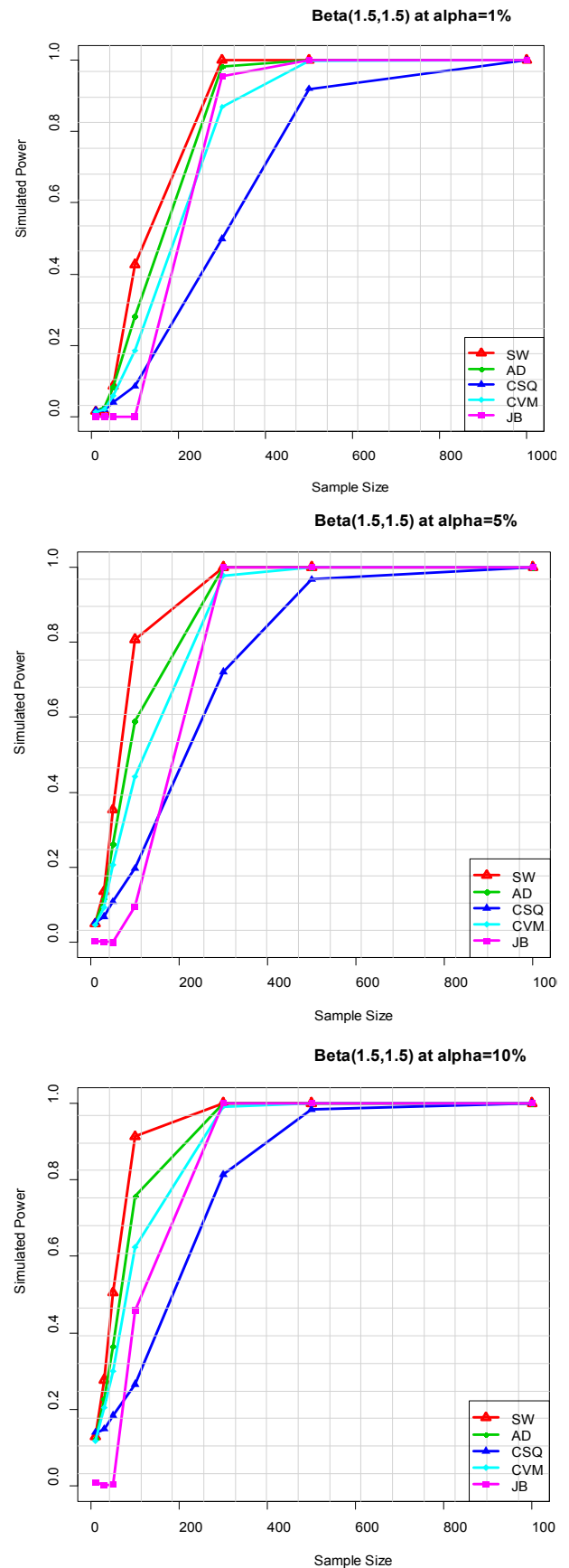
**Figure 3.2.** Power for all tests against selected symmetric short-tailed distributions

Table 3.2 shows the simulated power for selected Symmetric long-tailed distributions, while Figure 2.2 show the plot of power for all tests against selected Symmetric long-tailed distribution for 1%, 5% and 10% significance levels, which show that: for Symmetric long-tailed distributions, the power of JB is the best, follow closely by SW at all level of significance. And also show that the power of CSQ is very low compares to other tests.

Table 4.1 and Figure 3.1 reveal that, for symmetric short-tailed distributions, the power of SW is the best followed closely by AD and CVM tests and the power of JB and CSQ are relatively low, though, JB outperform CSQ.

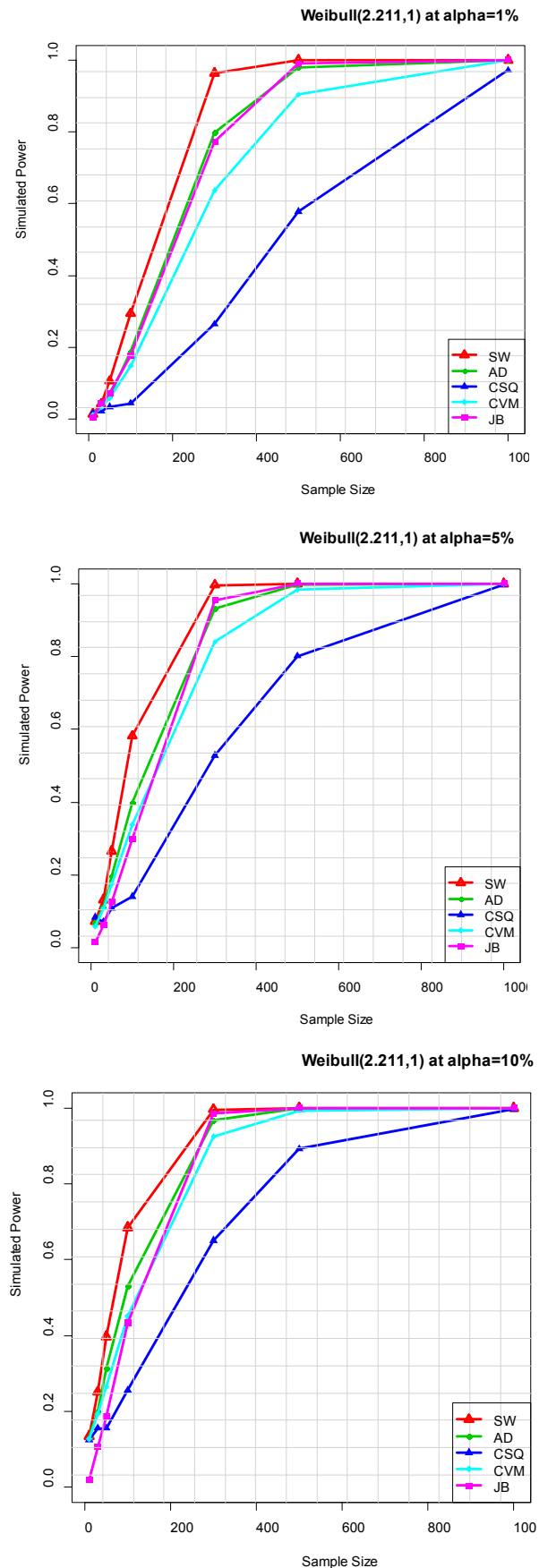
Table 4.2 and Figure 3.2 reveal that, for symmetric short-tailed distributions, the power of SW is the best followed closely by AD and CVM tests and the power of JB and CSQ are relatively low, though, JB outperform CSQ.

**Table 5.1.** Simulated Power of the Asymmetric long-tailed distributions

Alternative distribution	Sample size (n)	Normality tests				
		SW	AD	CSQ	CVM	JB
Weibull(2.211,1) at $\alpha=1\%$	10	0.014	0.011	0.016	0.011	0.004
	30	0.042	0.040	0.022	0.032	0.044
	50	0.106	0.064	0.034	0.058	0.072
	100	0.294	0.188	0.044	0.148	0.176
	300	0.964	0.798	0.264	0.637	0.772
	500	1.000	0.980	0.578	0.904	0.990
Weibull(2.211,1) at $\alpha=5\%$	10	0.073	0.067	0.084	0.058	0.014
	30	0.132	0.112	0.071	0.107	0.062
	50	0.265	0.197	0.108	0.175	0.125
	100	0.582	0.400	0.141	0.338	0.298
	300	0.996	0.931	0.528	0.840	0.955
	500	1.000	0.998	0.801	0.983	1.000
Weibull(2.211,1) at $\alpha=10\%$	10	0.134	0.131	0.125	0.127	0.019
	30	0.250	0.200	0.155	0.193	0.105
	50	0.397	0.313	0.155	0.264	0.186
	100	0.685	0.531	0.255	0.452	0.433
	300	0.996	0.967	0.651	0.925	0.985
	500	1.000	0.999	0.892	0.994	1.000
	1000	1.000	1.000	0.998	1.000	1.000

Table 5.1 summarize the simulated power for selected asymmetric long-tailed distributions for  $\alpha=1\%$ , 5% and 10%, while Figure 4.1 show the plot of power for all tests against selected asymmetric long-tailed distributions for all level of significance considered. They show that, for asymmetric long-tailed distributions, SW outperforms AD, CVM, JB and CSQ tests.

Table 5.2 summarize the simulated power for selected asymmetric long-tailed distributions for  $\alpha=1\%$ , 5% and 10%, while Figure 4.2 show the plot of power for all tests against selected asymmetric long-tailed distributions for all level of significance considered. Again for asymmetric long-tailed distributions, SW outperforms AD, CVM, JB and CSQ tests.



**Figure 4.1.** Simulated Power of the Asymmetric long-tailed distributions



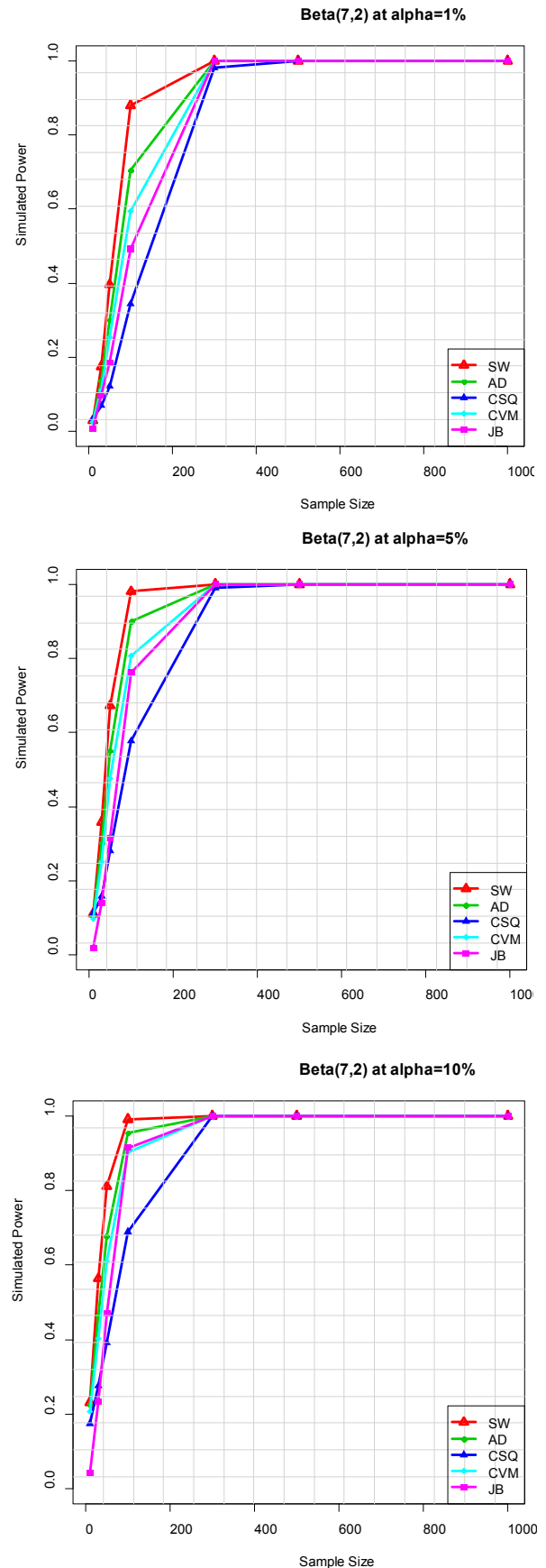
**Table 5.2.** Simulated power for selected asymmetric long-tailed distributions

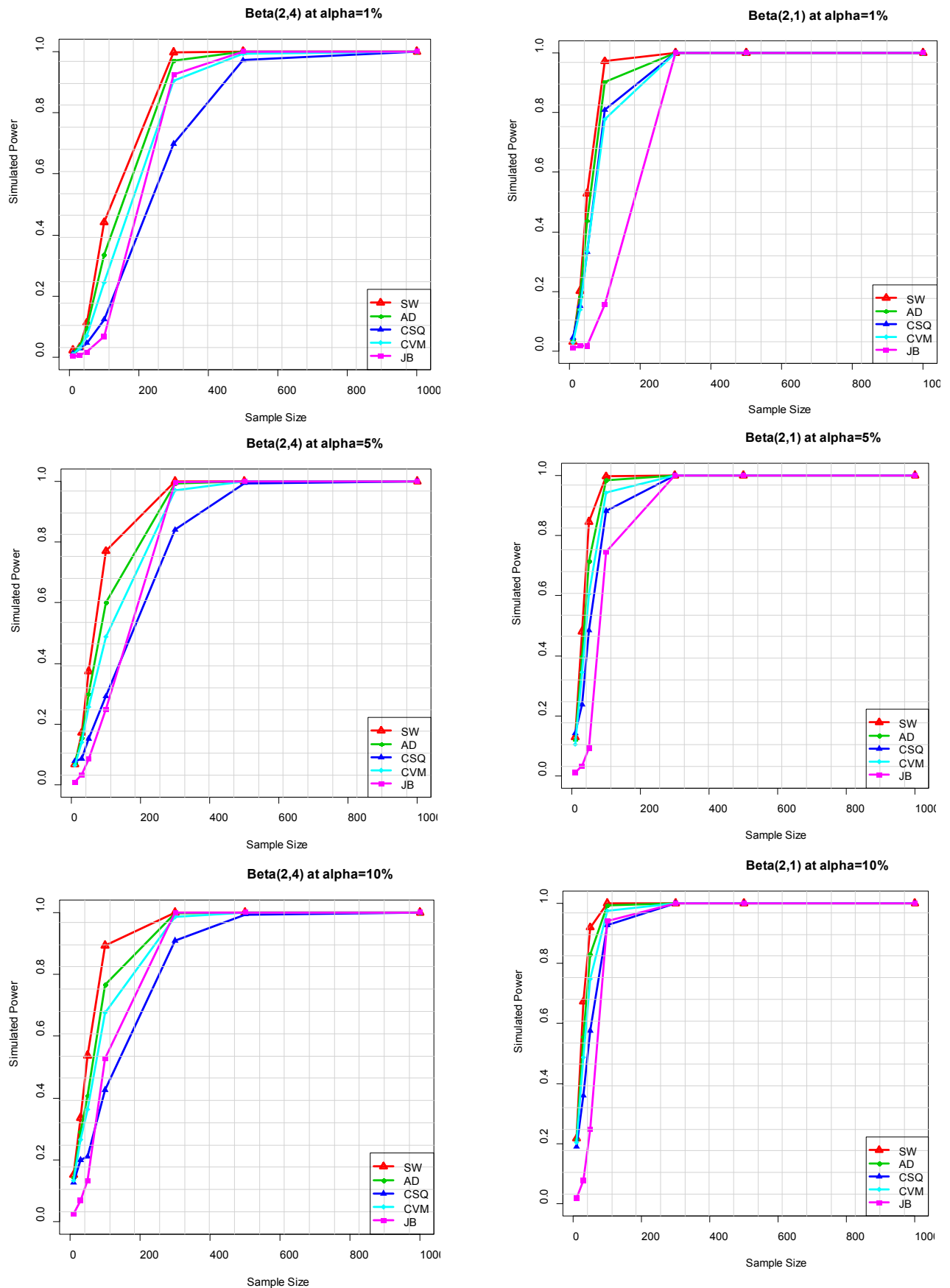
Alternative distribution	Sample size (n)	Normality tests				
		SW	AD	CSQ	CVM	JB
Beta(7,2) at $\alpha=1\%$	10	0.028	0.028	0.033	0.025	0.007
	30	0.175	0.135	0.069	0.111	0.097
	50	0.396	0.299	0.121	0.253	0.185
	100	0.880	0.704	0.344	0.594	0.492
	300	1.000	1.000	0.981	0.999	1.000
	500	1.000	1.000	1.000	1.000	1.000
	1000	1.000	1.000	1.000	1.000	1.000
Beta(7,2) at $\alpha=5\%$	10	0.111	0.104	0.113	0.097	0.020
	30	0.358	0.301	0.157	0.251	0.140
	50	0.672	0.551	0.281	0.476	0.317
	100	0.981	0.901	0.577	0.807	0.762
	300	1.000	1.000	0.991	1.000	1.000
	500	1.000	1.000	1.000	1.000	1.000
	1000	1.000	1.000	1.000	1.000	1.000
Beta(7,2) at $\alpha=10\%$	10	0.231	0.225	0.175	0.207	0.041
	30	0.563	0.464	0.276	0.403	0.233
	50	0.811	0.678	0.391	0.609	0.470
	100	0.991	0.955	0.689	0.903	0.914
	300	1.000	1.000	1.000	1.000	1.000
	500	1.000	1.000	1.000	1.000	1.000
	1000	1.000	1.000	1.000	1.000	1.000

**Table 6.1.** Simulated Power of the Asymmetric short-tailed distributions

Alternative distribution	Sample size (n)	Normality tests				
		SW	AD	CSQ	CVM	JB
Beta(2,4) at $\alpha=1\%$	10	0.138	0.133	0.134	0.135	0.018
	30	0.311	0.260	0.181	0.237	0.070
	50	0.575	0.480	0.251	0.408	0.130
	100	0.875	0.740	0.400	0.633	0.499
	300	1.000	0.999	0.917	0.993	1.000
	500	1.000	1.000	0.999	1.000	1.000
	1000	1.000	1.000	1.000	1.000	1.000
Beta(2,4) at $\alpha=5\%$	10	0.121	0.057	0.078	0.054	0.011
	30	0.318	0.173	0.089	0.154	0.038
	50	0.524	0.268	0.133	0.241	0.057
	100	0.886	0.583	0.281	0.497	0.250
	300	1.000	1.000	0.873	0.981	0.998
	500	1.000	1.000	0.990	1.000	1.000
	1000	1.000	1.000	1.000	1.000	1.000
Beta(2,4) at $\alpha=10\%$	10	0.138	0.133	0.134	0.135	0.018
	30	0.311	0.260	0.181	0.237	0.070
	50	0.575	0.480	0.251	0.408	0.130
	100	0.875	0.740	0.400	0.633	0.499
	300	1.000	0.999	0.917	0.993	1.000
	500	1.000	1.000	0.999	1.000	1.000
	1000	1.000	1.000	1.000	1.000	1.000

Table 6.1 reveals the simulated power for selected asymmetric short-tailed distributions for  $\alpha = 1\%$ ,  $5\%$  and  $10\%$  while Figure 5.1 show the plot of power for all tests against selected asymmetric long-tailed distributions for  $1\%$ ,  $5\%$  and  $10\%$  significance levels. Also for asymmetric long-tailed distributions, SW outperforms other tests.

**Figure 4.2.** Simulated power for selected asymmetric long-tailed distributions



**Figure 5.1.** Simulated Power of the Asymmetric short-tailed distributions

**Figure 5.2.** Simulated power for selected asymmetric short-tailed distributions

**Table 6.2.** Simulated power for selected asymmetric short-tailed distributions

Alternative distribution	Sample size (n)	Normality tests				
		SW	AD	CSQ	CVM	JB
Beta(2,1) at $\alpha=1\%$	10	0.032	0.035	0.045	0.029	0.010
	30	0.202	0.182	0.152	0.139	0.019
	50	0.528	0.438	0.332	0.334	0.018
	100	0.972	0.902	0.809	0.777	0.156
	300	1.000	1.000	1.000	1.000	1.000
	500	1.000	1.000	1.000	1.000	1.000
	1000	1.000	1.000	1.000	1.000	1.000
Beta(2,1) at $\alpha=5\%$	10	0.129	0.121	0.142	0.105	0.011
	30	0.481	0.396	0.238	0.345	0.033
	50	0.846	0.713	0.484	0.606	0.093
	100	0.997	0.983	0.882	0.942	0.745
	300	1.000	1.000	1.000	1.000	1.000
	500	1.000	1.000	1.000	1.000	1.000
	1000	1.000	1.000	1.000	1.000	1.000
Beta(2,1) at $\alpha=10\%$	10	0.217	0.216	0.189	0.204	0.018
	30	0.672	0.556	0.360	0.487	0.077
	50	0.919	0.829	0.576	0.748	0.248
	100	1.000	0.993	0.927	0.975	0.941
	300	1.000	1.000	1.000	1.000	1.000
	500	1.000	1.000	1.000	1.000	1.000
	1000	1.000	1.000	1.000	1.000	1.000

Table 6.2 reveals the simulated power for selected asymmetric short-tailed distributions for  $\alpha = 1\%$ ,  $5\%$  and  $10\%$  while Figure 5.2 show the plot of power for all tests against selected asymmetric long-tailed distributions for  $1\%$ ,  $5\%$  and  $10\%$  significance levels. Also for asymmetric long-tailed distributions, *SW* outperforms other tests.

In order to get a clear picture of the performance of these normality test. The ranking procedure was employed. The rank of 1 was given to the test with the highest power while rank of 5 was given to the test which has the lowest power (since there were five tests of normality considered in this study). The ranks were then summed to get the grand total of ranks. As the lowest number was given to the test with the highest power, therefore the test which had the lowest total rank was nominated as the best test to detect the departure from normality. Table 7.1 and Table 7.2 show the rank of power based on sample size and the type of alternative distribution, respectively.

The results of the total rank based on sample size in Table 7.1 below show that *SW* has the best test for all sample size since it consistently has the lowest total rank for all sample sizes considered.

Also from Table 7.2, it can be clearly seen that *JB* is the best test to be adopted for Near normal and Symmetric long-tailed distributions, and it is followed closely by *SW*. While *SW* is the best test to be adopted for both symmetric short-tailed, asymmetric long-tailed and asymmetric short-tailed distribution, and it is followed rather closely by the *AD* test since it has the lowest total rank (for all level of significances) among all the five tests considered.

**Table 7.1.** Rank of power based on sample size for all the alternative distributions

Level of significance		Sample size							
		10	30	50	100	300	500	1000	Total
$\alpha=1\%$	SW	19.0	18.0	14.0	14.0	14.0	14.0	13.0	106
	AD	24.0	24.0	26.0	24.0	26.0	24.0	22.0	170
	CSQ	24.0	44.0	46.0	46.0	44.0	31.0	25.0	260
	CVM	32.0	35.0	34.0	37.0	32.0	24.0	20.0	214
	JB	49.0	29.0	30.0	29.0	14.0	11.0	10.0	172
$\alpha=5\%$	SW	21.0	10.0	14.0	14.0	21.0	23.5	25.0	128.5
	AD	23.5	24.0	24.0	24.0	31.5	33.0	35.0	195
	CSQ	20.0	44.0	46.0	46.0	51.0	42.0	37.0	286
	CVM	35.5	34.0	34.0	34.0	35.0	33.0	31.0	236.5
	JB	50.0	38.0	32.0	32.0	23.5	19.5	22.0	217
$\alpha=10\%$	SW	16.0	11.0	12.0	14.0	20.5	23.5	25.5	122.5
	AD	19.5	22.0	24.0	24.0	35.5	33.0	33.5	191.5
	CSQ	36.0	40.0	46.0	50.0	44.0	42.0	38.0	296
	CVM	28.5	32.0	34.0	35.0	34.5	32.0	30.5	226.5
	JB	50.0	45.0	34.0	27.0	17.5	19.5	22.5	215.5

**Table 7.2.** Rank of power based on the type of Alternative distribution

		Near-normal	Symmetric long-tailed	Symmetric short-tailed	Asymmetric long-tailed	Asymmetric short-tailed
$\alpha=1\%$	SW	4.00	4.00	2.00	2.00	2.00
	AD	6.00	10.00	4.00	4.00	4.00
	CSQ	10.00	8.00	8.00	10.00	8.00
	CVM	8.00	6.0	6.00	6.00	6.00
	JB	2.00	2.00	10.00	8.00	10.00
$\alpha=5\%$	SW	4.00	4.00	2.00	2.00	2.00
	AD	6.00	10.00	4.00	4.00	4.00
	CSQ	10.00	8.00	8.00	10.00	8.00
	CVM	8.00	6.0	6.00	6.00	6.00
	JB	2.00	2.00	10.00	8.00	10.00
$\alpha=10\%$	SW	4.00	4.00	2.00	2.00	2.00
	AD	6.00	10.00	4.00	4.00	4.00
	CSQ	10.00	8.00	8.00	10.00	8.00
	CVM	8.00	6.0	6.00	6.00	6.00
	JB	2.00	2.00	10.00	8.00	10.00

## 4. Conclusions

In general, it can be concluded that:

- I. The powers of all the tests considered in this study increases as the sample size and significance level increases. This is in agreement with the findings of Razali et al. (2010).
- II. The results also supports the finding of Oztuna et. al. (2006) that the SW test has good power properties over a wide range of non-normal distributions.
- III. In consonance with the finding of Nakagawa et al. (2007), the JB power is very poor for Symmetric short-tailed distribution. But if the distribution is near normal and symmetric long-tailed, then JB tests is considered to be appropriate.
- IV. The results also agree with Stephen (1993) who reported that CSQ tests for normality have poor power property for all the alternative continuous non-normal distribution.

Finally, based on the findings from this study and in concurrence with relevant literature reviewed, it was discovered that most of these normality tests were developed in the 1960s, 1970s and 1980s. After the development of the Jarque-Bera test in the late 1980s, only few new tests seemed to appear after that time. However, none of these new tests could establish itself as a real alternative for testing

normality with an essential improvement of empirical power behavior. This leads us to the conclusion that the theory of testing for normality, in particular the search for the tests of normality with better power behavior than the already existing tests seems to be mostly closed.

## 5. Recommendations

As a concluding remark, it is recommended that a researcher should combine the graphical technique with formal normality test and also some measures of skewness and kurtosis should be put in consideration before choosing the appropriate normality test in order to avoid erroneous conclusion.

## REFERENCES

- [1] Anderson, T.W., and Darling, D.A., A Test of Goodness of fit, *Journal of the American Statistical Association*, 49; 1954: 765-769.
- [2] Shapiro, S.S., and Wilk, M.B., An Analysis of Variance Test for Normality, *Biometrika*, 52; 1965: 591-611.
- [3] Shapiro, S.S., Wilk, M.B. and Chen, H.J., A Comparative Study of Various Tests for Normality, *Journal of the American Statistical Association*, 3; 1968: 343-370.
- [4] Tae-Hwan Kim and Halbert White, On More Robust Estimation of skewness and Kurtosis: Simulation and Application to the S&P500 Index.
- [5] Thomas A. Ryan, Jr. and Brian L. Joiner, Normal Probability Plots and Tests for Normality, *Statistics Department, The Pennsylvania State University* 1976.
- [6] Derya Öztuna, Elhan, and Tüccar, Investigation of Four Different Normality Tests in Terms of Type 1 Error Rate and Power under Different Distributions, *Turk J Med Sci*, 2006; 36 (3): 171-176.
- [7] Seier Edith, Douglas G. Bonett, A test of Normality with high uniform power, *Computational statistics & Data Analysis* 40 (2002) 435-445.
- [8] Nakagawa, Niki and Hashiguchi. An Omnibus test for normality. *Proc. of the ninth Japan-China symposium on Statistics*. 191-194, 2007.
- [9] M.A. Stephens. EDF Statistics for Goodness of fit and some comparison. *Journal of the American statistical Association*. Vol. 69, No. 347 (sept, 1974), 730-737.
- [10] Nornadian Mohd Razali, Yap Bee Wah. Power comparisons of some selected normality tests. *Proceedings of the Regional Conference on Statistical Sciences 2010 (RCSS'10)* June 2010, 126-138.
- [11] James Brown D. *Shiken: JALT Testing & Evaluation SIG Newsletter*, 1 (1) April 1997 (p. 20-23).
- [12] R. C. Geary. Testing for Normality. *Biometrika*, Vol. 34, No. 3/4. (Dec., 1947), pp. 209-242.
- [13] M.A Stephen, Aspect of Goodness of fit, Technical report No. 474. Sept. 30, 1993.
- [14] Oja, H., 1981. On location, scale, skewness and kurtosis of univariate distributions. *Scand. J. Statist.* 8, 154-168.
- [15] Jean-Marie Dufour, Abdeljelil Farhat, Lucien Gardio, Lynda Khalaf. Simulation-based finite sample normality tests in linear regressions. *Econometrics Journal* (1998), volume 1, pp. 154-173.
- [16] Hun Myoung Park, 2008. Hypothesis Testing and Statistical Power of a Test. Working Paper. The University Information Technology Services (UIT) Center for Statistical and Mathematical Computing, Indiana University." <http://www.indiana.edu/~statmath/stat/all/power/index.html>.