



# Article Ranking of Normality Tests: An Appraisal through Skewed Alternative Space

# Tanweer Ul Islam

Department of Economics, National University of Sciences & Technology, Islamabad 44000, Pakistan; tanweer@s3h.nust.edu.pk

Received: 14 May 2019; Accepted: 18 June 2019; Published: 3 July 2019



**Abstract:** In social and health sciences, many statistical procedures and estimation techniques rely on the underlying distributional assumption of normality of the data. Non-normality may lead to incorrect statistical inferences. This study evaluates the performance of selected normality tests within the stringency framework for skewed alternative space. The stringency concept allows us to rank the tests uniquely. The Bonett and Seier test ( $T_w$ ) turns out to represent the best statistics for slightly skewed alternatives and the Anderson–Darling (AD); Chen–Shapiro (CS); Shapiro–Wilk (W); and Bispo, Marques, and Pestana (BCMR) statistics are the best choices for moderately skewed alternative distributions. The maximum loss of Jarque–Bera (JB) and its robust form (RJB), in terms of deviations from the power envelope, is greater than 50%, even for large sample sizes, which makes them less attractive in testing the hypothesis of normality against the moderately skewed alternatives. On balance, all selected normality tests except  $T_w$  and Daniele Coin's COIN-test performed exceptionally well against the highly skewed alternative space.

Keywords: power envelope; Neyman-Pearson tests; skewness; kurtosis

# 1. Introduction

Departures from normality can be measured in a variety of ways; however, the most common measures are skewness and kurtosis in this regard. Skewness refers to the symmetry of a distribution and kurtosis refers to the flatness or 'peakedness' of a distribution. These two statistics have been widely used to differentiate between distributions. A normal distribution has skewness and kurtosis values of 0 and 3, respectively. If the values of skewness and kurtosis significantly deviate from 0 and 3, it is assumed that the data in hand is not normally distributed. Macroeconomists are always concerned with whether the economic variables exhibit similar behavior during recessions and booms. Delong and Summers [1] applied the skewness measure to GDP, the unemployment rate, and industrial production to study whether the business cycles were symmetric. The experimental data sets generated in clinical chemistry require the use of skewness and kurtosis statistics to determine their shape and normality [2]. Blanca, Arnau, López-Montiel, Bono, and Bendayan [3] analyzed the shape of 693 real data distributions by including the measures of cognitive ability and other psychological variables in terms of skewness and kurtosis. Only 5.5% of the distributions were close to the normality assumption.

Keeping this in mind, the literature has produced few normality tests which are based on skewness and kurtosis [4–7]. Other than the moment based tests, normality literature also provides tests based on correlation and regression [8–10], empirical distribution [11–13], and special tests [14,15].

This study is devoted to analyzing the respective impact of change in skewness and kurtosis on the power of normality tests. Normality tests are developed based on the different characteristics of a normal distribution and the power of normality statistics varies, depending on the nature of non-normality [4]. Therefore, comparisons of normality tests yield ambiguous results since all normality statistics critically depend on alternative distributions which cannot be specified [16]. Fifteen normality

tests are selected for a comparison of power based on the stringency concept proposed by Islam [16]. The stringency concept allows you to rank the normality tests in a unique fashion. Neyman–Pearson (NP) tests are computed against each alternative distribution to construct the power curve. Relative efficiencies of all the tests in question are computed as the deviations of each test from the power curve. The best test is defined as the test displaying the minimum deviation from the power curve among the maximum deviations of all the tests.

# 2. Stringency Framework

Islam [16] proposed a new framework to evaluate the performance of normality tests based on the stringency concept introduced by Lehmann and Stein [17].

Let  $y = (y_1, y_2, y_3, ..., y_n)$  be the observations with the density function  $f(y, \varphi)$ , where  $\varphi$  belongs to the parameter space  $\emptyset$ . A function h(y) which takes the values {0, 1} is called a hypothesis test and belongs to  $H_\alpha$ , so all such functions are set with  $\alpha$ -level of significance.

For any test of size  $\alpha$ , maximum achievable power is defined as

$$Max_{h\in H_{\alpha}}\beta(h, \varphi) = Sup[P(h(y) = 1|\varphi \in \emptyset_a)]$$

where  $\beta(h, \varphi)$  is the power of h(y) and  $\emptyset_a$  represents the alternative parameter space. Different values of  $\varphi$  yield different optimal test statistics, which provide the power envelope. The relative power performance of a test,  $h \in H_{\alpha}$ , is measured by its deviation from the power envelope as

$$D(h(y), \varphi) = Max_{h \in H_{\alpha}}\beta(h, \varphi) - \beta(h, \varphi)$$

A test is said to be most stringent if it minimizes the maximum deviation from the power envelope. The stringency of a test is defined as the maximum deviation from the power envelope when evaluated over the entire alternative space.

$$S(h(y)) = Sup_{\varphi \in \emptyset_a} D(h(y), \varphi)$$

Only the uniformly most powerful test can have zero stringency, which is rarely found; however, slightly compromising on it can give us a test which is as good as the uniformly most powerful test [16]. Evaluating the normality tests based on their stringencies allows us to rank them in a unique manner and helps researchers to find the best test.

## 3. Tests and Alternative Distributions

Normality tests are based on different characteristics like the empirical distribution, moments, correlation, and regression, and based on special characteristics of the data distribution. Fifteen normality tests were selected (Table 1), which are the most representative of their respective class. The empirical cumulative distribution function (ECDF) class includes Zhang and Wu (Za, Zc), Anderson–Darling (AD), and Kolmogorov–Smirnov (KS) tests. The Jarque–Bera (JB), robust form of Jarque–Bera (RJB), Bowman–Shenton (K), and Bonett and Seier (Tw) test statistics represent the class of moments-based normality tests. The correlation and regression category is comprised of Shapiro–Wilk (W), Shapiro–Francia (Wsf), D'Agostino (D), Shapiro–Chen (CS), Coin (COIN), and Barrio et al. (BCMR) statistics. The Gel and Gastwirth (Rsj) statistic is a special test which focuses on detecting heavy tails and outliers of distributions. Some of our selected normality tests (e.g., Jarque–Bera, Kolmogorov–Smirnov, Anderson–Darling, Shapiro–Wilk, Shapiro–Francia etc.) are also available in popular software like MATLAB, STATA, SPSS, and EViews. Departures from normality (first and second order) depend on the skewness and kurtosis parameters. A mixture of t-distributions allows you to vary these two statistics in a wide range. It also covers the distributions used in the literature in terms of skewness and kurtosis (for details, see [16]).

	-
Test	Class of Test
Za, Zc, AD, and KS	ECDF
JB, RJB, K, and Tw	Moments
W, Wsf, D, CS, BCMR, and COIN	Correlation and Regression
Rsj	Special

Table 1. Normality tests.

This study uses a mixture of t-distributions as alternative distributional space (Appendix: Table A1). The alternative distributional space was generated by the following rule:

 $\lambda t(v_1, \mu_1) + (1 - \lambda) t(v_2, \mu_2)$ 

where  $v_1$ ,  $v_2$ ,  $\mu_1$ , and  $\mu_2$  are the degrees of freedom and the means of the respective t-distributions. We have divided our alternative space of distributions into the following three groups on the basis of skewness ( $\beta_1$ ): (i) slightly skewed, (ii) moderately skewed, and (iii) highly skewed. In each group, skewness remained within the bounds and we allowed kurtosis to vary. The benchmark value for Group-I (symmetric distributions) was defined by the following [18], and other classes were defined relatively

Group I : 
$$|\beta_1| \le 0.3$$
  
Group II :  $0.3 < |\beta_1| \le 1.5$   
Group III :  $|\beta_1| > 1.5$ 

Neyman–Pearson (NP) tests were computed against each alternative distribution in each group to construct the power curve. Relative efficiencies of all the tests in question were computed as the deviations of each test from the power curve. The best test was defined as the test displaying the minimum deviation from the power curve among the maximum deviations of all the tests.

Following Islam [16], I group alternative space into three categories based on the power of the NP test: FAR, INTERMEDIATE, and NEAR. The alternative distributions where the power of the NP test is between 90–100%, 40–90%, and 5–40% are categorized as the FAR, INTERMEDIATE, and NEAR group of alternatives, respectively.

## 4. Discussion of Results

Monte Carlo procedures were employed to investigate the powers of fifteen selected normality tests for sample sizes of 25, 50, and 75, at the 5% level of significance with 100,000 replications.

#### 4.1. Slightly Skewed Alternatives

When considering all the selected normality tests, Tw is the best test against the slightly skewed alternatives (Figures A1–A10 and Table 2) for all sample sizes (n = 25, 50, 75), whereas the performance of JB and RJB tests is very poor, with an 80.5%–99.5% maximum loss of power.

# 4.1.1. Performance of the Moments-Based Tests

Among the moments-based class of normality tests, Tw is the best test for all sample sizes for slightly skewed alternatives (Table 2 and Figure A1). The K2 test occupies the fourth (for n = 25, 50) and third (for n = 75) rank, with maximum power losses of 42.6%, 44.8%, and 44.7%, respectively (Figure A3).

For all sample sizes, the JB and RJB tests are the least favorable options in terms of their maximum deviations (gaps) from the power curve (Figure A2). The worst distributions for JB and RJB statistics belong to the symmetric and short-tailed class of alternatives (Figure A2 and Appendix Table A2). These results corroborate with the findings in [12,19,20]. To decide about the worst or best performance of a test, we need an invariant benchmark: a power envelope. The worst performances of JB, in the

aforementioned studies, have been evaluated by using an arbitrary reference (e.g., W and AD); however, we computed the power curve by using the most powerful NP test, which yielded the exact deviations of the JB test from the power curve.

	n = 25			n = 50			n = 75	
Test	Rank	Gap	Test	Rank	Gap	Test	Rank	Gap
Tw	1	24.0%	Tw	1	22.9%	Tw	1	31.8%
COIN	2	34.6%	Rsj	2	26.4%	Rsj	1	32.4%
AD	2	34.7%	AD	3	38.0%	AD	1	32.6%
CS	2	34.8%	CS	3	39.8%	CS	2	38.6%
Rsj	3	36.1%	COIN	4	42.5%	W	3	43.3%
Ŵ	3	37.5%	K2	4	44.8%	K2	3	44.7%
KS	3	38.1%	W	4	45.5%	KS	3	45.1%
Zc	3	39.0%	Zc	5	48.0%	COIN	3	45.2%
BCMR	3	39.9%	BCMR	5	48.3%	BCMR	3	46.1%
K2	4	42.6%	KS	5	49.9%	Zc	4	50.5%
Za	4	43.1%	Za	6	51.9%	Za	4	51.4%
Wsf	4	46.5%	Wsf	7	61.3%	Wsf	5	56.2%
D	5	91.6%	JB	8	80.5%	D	6	85.3%
JB	6	97.2%	D	9	90.9%	JB	7	88.0%
RJB	6	98.2%	RJB	10	99.5%	RJB	8	92.9%

**Table 2.** Ranking of the normality tests ( $|\beta_1| < 0.3$ ).

#### 4.1.2. Performance of the Regression and Correlation Tests

When considering the regression and correlation-based group of normality tests, for small and large sample sizes (n = 25, 75), COIN, W, and BCMR are better choices for the slightly skewed alternatives. Overall, for slightly skewed distributions, COIN and W tests exhibit the same power properties (Figures A5 and A6), whereas Wsf and D statistics do not match the standards set by other members of the group (Figures A7 and A8), with maximum power losses of over 50% (Table 2). Overall, the CS outperforms its competitors in the said group, with maximum power loss ranges within 34.8%–39.8% for slightly skewed alternatives. This result strengthens the findings in [21].

# 4.1.3. Performance of the ECDF Tests

Among the ECDF class of normality tests, for slightly skewed alternatives, the AD statistic shares the second rank with COIN and CS, third rank with CS, and first rank with Tw and Rsj tests of normality for sample sizes of 25, 50, and 75, respectively (Table 2).

When considering all the selected normality tests for the slightly skewed alternative distributions, KS shares the third rank (maximum loss of power is 38.1%) with W and Zc and sixth rank (maximum loss of power is 49.9%) with Zc and BCMR for sample sizes of 25 and 50, respectively. For a sample size of 75, the KS test again holds the third rank with a 45.1% maximum loss of power, while Za and Zc tests hold the fourth rank with a maximum loss of powers slightly above 50% (Table 2). On balance, when considering the maximum deviations from the power envelope, KS has a slight edge over Za and Zc statistics. In terms of maximum deviations from the power envelope, Zc has a slight edge over Za, but it does not corroborate with the findings in [13] due to the absence of an invariant benchmark, the power envelope, in their comparison.

## 4.1.4. Performance of the Special Test

This category only includes the Rsj test of normality. The performance of the Rsj test increases with the increase in sample size for the slightly skewed alternatives. It holds the third, second, and first rank for sample sizes of 25, 50, and 75, respectively (Table 2). On balance, Rsj performed well (Figure A10), especially for medium (n = 50) to large (n = 75) sample sizes, against slightly skewed distributions.

Finally, when considering all normality tests for slightly skewed alternatives, Tw is the most stringent test, with Rsj, AD, and CS following closely behind, whereas RJB, JB, and D are the least favorable options.

#### 4.2. Moderately Skewed Alternatives

For moderately skewed alternatives, for a smaller sample size, CS, W, AD, and BCMR are the best choices and the COIN test is the least favorable option (Table 3). For a medium sample size, AD is ranked first and the COIN and Tw tests are at the bottom of the ranking table. For a larger sample size, AD, CS, W, and BCMR appear to be the best options, whereas the COIN and Tw tests are the worst options.

#### 4.2.1. Performance of the Moments-Based Tests

In general, for moderately skewed alternatives, moments-based normality tests perform poorly for all sample sizes. For a smaller sample size, Bowman & Shenton [22] K2-test occupies the fourth rank (with 46.7% maximum power loss) by outperforming the other group members. For a medium sample size, JB and RJB (with power losses above 50.0%) move to fourth place by pushing K2 down to fifth place, whereas Tw shares the seventh rank (maximum power loss is 78.4%) with the COIN test.

With the increase in sample size, both JB and RJB show an improvement in power and ranking, but their maximum power losses are still above 50% (Table 3). Both JB and RJB are good at discriminating the FAR group of distributions (where the power of the NP test is between 90–100%), with JB having a slight edge over RJB, but both suffer when the distributions are from the INTERMEDIATE group of alternatives (Figure A11).

## 4.2.2. Performance of the Regression and Correlation Tests

Among the regression and correlation-based normality tests, for a smaller sample size, CS, W, and BCMR are the best tests for moderately skewed alternatives, with a loss range of 28.5%–29.8% (Table 3), whereas the COIN test is at the bottom, with a loss range of 68.8–88.7%.

For a medium to large sample size (n = 50, 75), W, BCMR, and CS are the better options, with Wsf following closely behind. The D and COIN tests are the least favorable regression and correlation-based normality statistics for moderately skewed alternatives, which is in line with the findings in [4,22]. It is evident from Figure A12 that Tw and COIN both suffer against the INTERMEDIATE and FAR group of alternative distributions.

	n = 25			n = 50			n = 75	
Tests	Rank	Gap	Test	Rank	Gap	Test	Rank	Gap
CS	1	28.5%	AD	1	25.0%	AD	1	26.7%
W	1	29.0%	W	2	28.3%	CS	1	28.9%
AD	1	29.5%	BCMR	2	28.7%	W	1	29.5%
BCMR	1	29.8%	CS	2	29.8%	BCMR	1	31.4%
Za	2	32.8%	Wsf	3	34.9%	Wsf	2	35.8%
Wsf	2	33.5%	KS	3	35.2%	Za	2	36.2%
Zc	2	33.5%	Za	3	36.5%	Zc	2	38.2%
KS	3	42.2%	Zc	3	38.3%	KS	2	40.4%
K2	4	46.7%	JB	4	59.8%	JB	3	50.6%
D	5	49.8%	RJB	4	61.9%	K2	4	57.9%
Rsj	6	55.5%	K2	5	64.6%	RJB	4	58.0%
Tw	6	55.7%	D	6	74.6%	D	5	81.3%
JB	7	59.0%	Rsj	6	75.6%	Rsj	5	83.9%
RJB	8	64.4%	Tw	7	78.4%	Tw	6	88.0%
COIN	9	68.8%	COIN	7	79.8%	COIN	6	88.7%

**Table 3.** Ranking of the normality tests  $(0.3 < |\beta_1| \le 1.5)$ .

#### 4.2.3. Performance of the ECDF Tests

For moderately skewed alternatives, among the ECDF class of normality tests, AD exhibits superior power properties for all sample sizes. When considering all the selected normality tests for moderately skewed alternatives, AD holds the first rank for all sample sizes.

For a smaller and larger sample size, the Za and Zc statistics share the second rank. For a medium sample size, these tests occupy the third rank. For a smaller and medium sample size, the KS test holds the third rank, whereas its position improves to second rank for a larger sample size. The W test turns out to be a better test than KS (Figure A13), which corroborates the findings in Shapiro, Wilk, and Chen [18]. While evaluating the stringencies of the normality statistics for moderately skewed alternatives, we produce the same conclusion, but through a superior and reliable procedure.

## 4.2.4. Performance of the Other Tests

In general, for moderately skewed alternative distributions, the Rsj test performs poorly, exhibiting more than 50.0% maximum deviation from the power curve for all sample sizes. On balance, the worst performance of the Rsj test is against the INTERMEDIATE and FAR group of alternatives, but it performed well against the NEAR group of alternatives.

Overall, AD, CS, W, and BCMR happen to be the best and JB, RJB, Tw, Rsj, and COIN are the least favorable options for moderately skewed alternatives when considering all the selected normality tests.

#### 4.3. Highly Skewed Alternatives

This group comprises the alternatives from the FAR group only where the most powerful NP test has 100% power. As both skewness and kurtosis are high for this group of alternatives, they are palpable. All normality tests other than the COIN and Tw statistics performed well against highly skewed alternatives (Table 4).

	n = 25			n = 50			n = 75	
Test	Rank	Gap	Test	Rank	Gap	Test	Rank	Gap
Wsf	1	8.8%	Wsf	1	0.6%	RJB	1	0.0%
BCMR	1	9.3%	BCMR	1	0.7%	Zc	1	0.0%
W	1	10.1%	Zc	1	0.7%	JB	1	0.0%
CS	1	10.4%	W	1	0.7%	Wsf	1	0.0%
Za	1	10.9%	JB	1	0.7%	W	1	0.1%
Zc	1	11.0%	RJB	1	0.7%	CS	1	0.1%
AD	1	11.9%	ĊS	1	0.8%	D	1	0.1%
RJB	1	12.5%	Za	1	0.9%	BCMR	1	0.1%
JB	1	13.9%	D	1	1.0%	K2	1	0.1%
D	2	16.1%	K2	1	1.2%	Za	1	0.1%
K2	3	20.4%	AD	1	1.3%	AD	1	0.2%
KS	3	21.2%	Rsj	1	2.1%	Rsj	1	0.2%
Rsj	3	21.5%	КŚ	1	3.6%	КŚ	1	0.5%
Tw	4	46.9%	Tw	2	45.3%	Tw	2	42.5%
COIN	5	61.4%	COIN	3	69.1%	COIN	3	72.0%

**Table 4.** Normality tests for highly skewed alternatives ( $|\beta_1| > 1.5$ ).

For a smaller sample size, the Wsf, BCMR, W, CS, Za, Zc, AD, RJB, and JB tests performed well, with the maximum power loss ranging between 8.8%–13.9%, followed by the D statistic with maximum power loss of 16.1% (Table 4), while the performance of the COIN and Tw tests was below the mark.

As the sample size increases, it becomes harder to differentiate among the selected tests of normality, excluding Tw and COIN. The results clearly show that the power loss of these statistics decreases with the increase in (i) sample size and (ii) skewness and kurtosis. For all sample sizes, JB and RJB yield good powers for the highly skewed alternatives.

Overall, the performance of the normality tests against the highly skewed and heavy-tailed alternatives is very good. However, the COIN and Tw tests performed poorly compared to other normality statistics. The poor performance of the COIN test is understandable as it is only meant for perfect symmetric cases [21,22]. Bonett and Seier [4] also recommend a standard skewness test along with the Tw statistic when the alternative distribution is skewed. Therefore, the COIN and Tw tests are not recommended for highly skewed alternative distributions.

# 5. Conclusions

This study shed light on the performance of fifteen normality tests against the three different groups of alternatives. For slightly skewed alternative distributions, Tw is the best test, with COIN, AD, CS, and Rsj following closely behind. On balance, D, JB, RJB, K2, Wsf, and Za did not perform well for the slightly skewed alternatives, especially from medium (n = 50) to large (n = 75) sample sizes, with more than 50% maximum power losses.

When considering all the selected normality tests for the moderately skewed alternatives, AD, CS, W, and BCMR turn out to be the best options for testing the hypothesis of normality of data distribution. In general, JB, RJB, Tw, COIN, Rsj, D, and K2 tests perform poorly against moderately skewed distributions. The performance of JB and RJB increases with the increase in sample size, but their maximum loss, in terms of their deviations from the power envelope, is greater than 50%, even for large sample sizes (n = 75).

On balance, all normality tests except Tw and COIN performed exceptionally well against the highly skewed alternatives, especially from medium to large sample sizes.

The above findings confirm our argument that a comparison of tests against different alternatives yields different statistics as the best tests. The COIN [23] and Tw tests are the best options for slightly skewed alternatives, but these statistics perform poorly for moderately and highly skewed alternative distributions. Therefore, the comparison and ranking of normality tests do not make sense in the absence of an invariant benchmark: the power envelope.

**Author Contributions:** I.T-U. conceived the idea, conceptualized the methodology and performed the numerical simulations. The author discussed the results with field expert and incorporated the valuable suggestions while writing the original draft. The author reviewed the manuscript under the light of reviewer's comments.

Funding: This research received no external funding.

Acknowledgments: I would like to thank Asad Zaman for his valuable comments and guidance.

Conflicts of Interest: The authors declare no conflict of interest.

#### Appendix A

	9	Student t I	Distribut	tion	Mixture Distribution						
Sr. No	Sr. No t1		t2		Mixture Distribution						
	d.f	Mean	d.f	Mean	Alpha	Mean	SD	$\beta_1$	$\beta_2$		
1	8	2.0	12	5.0	0.50	3.50	1.88	-0.05	2.33		
2	100	4.0	75	6.0	0.50	5.00	1.42	0.00	2.53		
3	10	0.0			1.00	0.00	1.12	0.00	4.00		
4	100	-1.5	75	1.5	0.50	0.00	1.81	0.00	2.06		
5	10	3.0	5	50.0	0.50	26.50	23.53	0.00	1.01		
6	100	-4.0	75	4.0	0.50	0.00	4.13	0.00	1.23		
7	50	-1.2	25	1.2	0.50	0.00	1.58	0.02	2.38		
8	8	5.0	10	3.0	0.50	4.00	1.51	0.04	3.02		
9	5	2.0	7	4.0	0.70	2.60	1.56	0.09	4.95		
10	5	10.0	6	12.0	0.95	10.10	1.36	0.12	7.84		

Table A1. Distributions.

	9	Student t I	Distribut	ion		Mixtur	e Distrih	ution	
Sr. No	ł	t1		t2	-			ution	
	d.f	Mean	d.f	Mean	Alpha	Mean	SD	$\beta_1$	$\beta_2$
11	5	10.0	7	12.0	0.90	10.20	1.41	0.15	6.90
12	10	5.0	5	7.0	0.50	6.00	1.57	0.16	4.20
13	100	4.0	75	6.0	0.70	4.60	1.36	0.27	2.77
14	8	5.0	10	3.0	0.10	3.20	1.27	0.30	3.95
15	100	-1.0	75	1.0	0.75	-0.50	1.33	0.32	2.91
16	8	5.0	10	3.0	0.20	3.40	1.38	0.32	3.57
17	10	5.0	5	7.0	0.90	5.20	1.29	0.38	4.65
18	100	-1.2	75	1.2	0.75	-0.60	1.45	0.43	2.85
19	8	-1.0	10	2.0	0.95	-0.85	1.33	0.48	4.68
20	8	-1.0	12	2.0	0.85	-0.55	1.57	0.59	3.70
21	100	-1.5	75	1.5	0.77	-0.81	1.62	0.61	2.88
22	100	-4.0	75	4.0	0.70	-1.60	3.80	0.78	1.93
23	5	10.0	7	25.0	0.70	14.50	6.99	0.82	1.83
24	10	3.0	5	50.0	0.70	17.10	21.57	0.87	1.77
25	100	-4.0	75	4.0	0.75	-2.00	3.61	1.02	2.44
26	8	-10.0	12	5.0	0.78	-6.70	6.32	1.28	2.83
27	8	0.0	12	5.0	0.90	0.50	1.89	1.31	5.11
28	8	0.0	12	5.0	0.95	0.25	1.59	1.32	6.63
29	8	-10.0	12	5.0	0.80	-7.00	6.11	1.42	3.22
30	8	-10.0	12	5.0	0.82	-7.30	5.88	1.57	3.71
31	8	-1.0	12	5.0	0.90	-0.40	2.14	1.58	5.60
32	5	5.0	7	15.0	0.85	6.50	3.79	1.62	4.45
33	5	5.0	6	15.0	0.90	6.00	3.26	2.06	6.73
34	100	-4.0	75	4.0	0.90	-3.20	2.60	2.09	6.69
35	5	10.0	7	25.0	0.90	11.50	4.68	2.36	7.35
36	8	-10.0	12	5.0	0.90	-8.50	4.64	2.42	7.48
37	10	3.0	5	50.0	0.90	7.70	14.15	2.64	8.06

Table A1. Cont.

Note:  $\beta_1$  and  $\beta_2$  represent the skewness and kurtosis of the distribution.

**Table A2.** Power comparison of symmetric short-tailed alternatives (n = 25,  $\alpha$  = 0.05).

Distribution	Skew	Kurt	JB	RJB	Best Test
D(5,0.00,1.01)	0.00	1.01	0.27	0.04	1.00
D(6,0.00,1.23)	0.00	1.23	0.03	0.02	1.00
Beta(0.5,0.5)	0.00	1.50	0.00	0.00	0.91
Beta(1,1)	0.00	1.80	0.00	0.00	0.44
Tukey(2)	0.00	1.80	0.00	0.00	0.44
D(4,0.00,2.06)	0.00	2.06	0.01	0.00	0.54
Tukey(0.5)	0.00	2.08	0.00	0.00	0.14
Beta (2,2)	0.00	2.14	0.00	0.00	0.11
D(2,0.00,2.53)	0.00	2.53	0.02	0.01	0.16
Tukey(5)	0.00	2.90	0.03	0.07	0.14

# Appendix B



**Figure A1.** Power Comparison of NP and Tw Tests ( $|\beta_1| < 0.3 \& n = 75$ ).



**Figure A2.** Power Comparison of NP, Tw, RJB, and JB Tests ( $|\beta_1| < 0.3 \& n = 75$ ).



**Figure A3.** Power Comparison of NP, Tw, and K2 Tests ( $|\beta_1| < 0.3 \& n = 75$ ).



**Figure A4.** Power Comparison of NP, Tw, and CS Tests ( $|\beta_1| < 0.3 \& n = 75$ ).



**Figure A5.** Power Comparison of NP, Tw, and COIN Tests ( $|\beta_1| < 0.3 \& n = 75$ ).



**Figure A6.** Power Comparison of NP, Tw, and W Tests ( $|\beta_1| < 0.3 \& n = 75$ ).



**Figure A7.** Power Comparison of NP, Tw, and Wsf Tests ( $|\beta_1| < 0.3 \& n = 75$ ).



**Figure A8.** Power Comparison of NP, Tw, and D Tests ( $|\beta_1| < 0.3 \& n = 75$ ).



**Figure A9.** Power Comparison of NP, Tw, and BCMR Tests ( $|\beta_1| < 0.3 \& n = 75$ ).







**Figure A11.** Power Comparison of NP, Tw, and COIN Tests ( $0.3 < |\beta_1| \le 1.5 \& n = 75$ ).



**Figure A12.** Power Comparison of NP, JB, and RJB Tests ( $0.3 < |\beta_1| \le 1.5 \& n = 75$ ).





**Figure A13.** Power Comparison of NP, *W*, and KS Tests ( $0.3 < |\beta_1| \le 1.5 \& n = 75$ ).

# References

- 1. Delong, J.B.; Summers, L.H. *Are Business Cycle Symmetrical in American Business Cycle: Continuity and Change;* University of Chicago Press: Chicago, IL, USA, 1985; pp. 166–178.
- 2. Henderson, A.R. Testing experimental data for univariate normality. *Clin. Chim. Acta* 2006, 366, 112–129. [CrossRef]
- 3. Blanca, M.J.; Arnau, J.; López-Montiel, D.; Bono, R.; Bendayan, R. Skewness and Kurtosis in Real Data Samples. *Methodology* **2013**, *9*, 78–84. [CrossRef]
- 4. Bonett, D.G.; Seier, E. A test of normality with high uniform power. *Comput. Stat. Data Anal.* 2002, 40, 435–445. [CrossRef]
- 5. D'Agostino, R.; Pearson, E.S. Tests for departure from normality. Empirical results for the distributions of b2 and b1. *Biometrika* **1973**, *60*, 613–622. [CrossRef]
- 6. Gel, Y.R.; Gastwirth, J.L. A robust modification of the Jarque–Bera test of normality. *Econ. Lett.* **2008**, *99*, 30–32. [CrossRef]
- Jarque, C.M.; Bera, A.K. A Test for Normality of Observations and Regression Residuals. *Int. Stat. Rev.* 1987, 55, 163–172. [CrossRef]
- 8. Bispo, R.; Marques, T.A.; Pestana, D. Statistical power of goodness-of-fit tests based on the empirical distribution function for type\_I right-censored data. *J. Stat. Comput. Simul.* **2012**, *82*, 21–38. [CrossRef]
- 9. Shapiro, S.S.; Francia, R.S. An approximate analysis of variance test for normality. *J. Am. Stat. Assoc.* **1972**, 67, 215–216. [CrossRef]
- 10. Shapiro, S.S.; Wilk, M.B. An analysis of variance test for the exponential distribution (Complete samples). *Biometrika* **1965**, *54*, 591–611. [CrossRef]
- 11. Anderson, T.W.; Darling, D.A. A test of goodness of fit. J. Am. Stat. Assoc. 1954, 49, 765–769. [CrossRef]
- 12. Yazici, B.; Yolacan, S. A comparison of various tests of normality. *J. Stat. Comput. Simul.* **2007**, 77, 175–183. [CrossRef]
- 13. Zhang, J.; Wu, Y. Likelihood-ration tests for normality. Comput. Stat. Data Anal. 2005, 49, 709–721. [CrossRef]
- 14. Gel, Y.R.; Miao, W.; Gastwirth, J.L. Robust directed tests of normality against heavy-tailed alternatives. *Comput. Stat. Data Anal.* **2007**, *51*, 2734–2746. [CrossRef]
- 15. Önder, A.Ö.; Zaman, A. Robust tests for normality of errors in regression models. *Econ. Lett.* **2005**, *86*, 63–68. [CrossRef]
- 16. Islam, T.U. Stringency-based ranking of normality tests. *Commun. Stat. Simul. Comput.* **2017**, *46*, 655–668. [CrossRef]
- 17. Lehmann, E.L.; Stein, C. On the Theory of Some Non-Parametric Hypotheses. *Ann. Math. Stat.* **1949**, *20*, 28–45. [CrossRef]
- Shapiro, S.S.; Wilk, M.B.; Chen, H.J. A comparative study of various tests for normality. *J. Am. Stat. Assoc.* 1968, 63, 1343–1372. [CrossRef]
- 19. Thorsten, T.; Buning, H. Jarque-Bera Test and its Competitors for Testing Normality-A Power Comparison. *J. Appl. Stat.* **2007**, *34*, 87–105.

- 20. Yap, B.W.; Sim, C.H. Comparisons of various types of normality tests. J. Stat. Comput. Simul. 2011, 81, 1–15. [CrossRef]
- 21. Romao, X.; Delgado, R.; Costa, A. An empirical power comparison of univariate goodness-of-fit tests for normality. *J. Stat. Comput. Simul.* **2010**, *80*, 1–47. [CrossRef]
- 22. Bowman, K.O.; Shenton, L.R. Omnibus test contours for departures from normality based on b1 and b2. *Biometrika* **1975**, *62*, 243–250. [CrossRef]
- 23. Coin, D. A goodness-of-fit test for normality based on polynomial regression. *Comput. Stat. Data Anal.* 2008, 52, 2185–2198. [CrossRef]



© 2019 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).