Article

Ranking of Normality Tests - An Appraisal through Skewed Alternative Space

Tanweer Islam[1]

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

Abstract: In social & 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 on the stringency framework for the skewed alternative space. Stringency concept allows us to rank the tests uniquely. Bonett & Seier test (T_w) turns out to be the best statistics for slightly skewed alternatives and the Anderson-Darling (*AD*), Chen-Shapiro (*CS*), Shapiro-Wilk (*W*) and Bispo, Marques, & Pestana, (*BCMR*) statistics are the best choices for moderately skewed alternative distributions. 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 *COIN* 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. Normal distribution has the values of skewness and kurtosis as 0 & 3 respectively. If the values of skewness and kurtosis significantly deviate from 0 & 3; it is assumed that the data in hand is not distributed as normal. Macroeconomists are always concerned whether the economic variables exhibit similar behavior during recessions and booms. Delong & Summers [8] apply the skewness measure to GDP, unemployment rate and industrial production to study whether the business cycles are symmetric or not. The experimental data sets generated in clinical chemistry require the use of skewness & kurtosis statistics to determine its shape and normality [11]. Blanca, Arnau, López-Montiel, Bono, & Bendayan [3] analyze 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 are close to normality assumption.

Keeping this in view, the literature has produced few normality tests which are based on skewness and kurtosis [4, 7, 9, & 13]. Other than the moment based tests, normality literature also provides tests based on correlation & regression [2, 6, 18 & 19], empirical distribution [1, 23 & 24] and special tests [10 & 15].

Peer-reviewed version available at Symmetry 2019, 11, 872; doi:10.3390/sym11070872

This study is devoted to analyze the impact of change in skewness and kurtosis respectively on the power of normality tests. Normality tests are developed based on the different characteristics of normal distribution and the power of normality statistics varies depending upon the nature of non-normality [4]. Thus, comparisons of normality tests yield ambiguous results since all normality statistics critically depend on alternative distributions which cannot be specified [12]. Fifteen normality tests are selected for comparison of power based on stringency concept proposed by Islam [12]. The stringency concept allows you to rank the normality tests in a uniquely 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 having minimum deviation from the power curve among the maximum deviations of all the tests.

2. Stringency Framework

Islam [12] proposes a new framework to evaluate the performance of normality tests based on the stringency concept introduced by Lehmann and Stein [14].

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

For any test of size α, maximum achievable power is defined as:

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

where, $\beta(h, \varphi)$ is the power of $h(y)$ and $\emptyset_a$ represents the alternative parameter space. For different values of φ 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. Stringency of a test is defined as the maximum deviation from the power envelope when evaluated over the entire alternative space.

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

Only the uniformly most powerful test can have zero stringency which are rarely found however, slightly compromising on it can give us a test which is as good as the uniformly most powerful test [12]. Evaluating the normality tests based on their stringencies allows us to rank them in a uniquely manner and helps researcher to find the best test.

3. Tests & Alternative Distributions

Normality tests are based on different characteristics like empirical distribution, moments, correlation and regression and based on special characteristics of the data distribution. Fifteen normality tests are selected (Table 1) which are the most representative of their respective class. Departures from normality (first & second order) depends on skewness & kurtosis parameters. Mixture of t-distributions allows you to vary these two statistics in a wide range. It also covers the distributions used in literature in terms of skewness & kurtosis (for details see [12]).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 May 2019 doi:10.20944/preprints201905.0190.v1

Peer-reviewed version available at Symmetry 2019, 11, 872; doi:10.3390/sym11070872

Table 1: Normality tests

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

This study uses the mixture of t-distributions as alternative distributional space (Appendix: Table 1). The alternative distributional space is generated by the following rule.

$$\alpha \, . \, t(v_1, \mu_1) + (1-\alpha) . \, t(v_2, \mu_2) \qquad (1)$$

where v_1, v_2, μ_1, μ_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; (i) slightly skewed (ii) moderately skewed and, (iii) highly skewed. In each group, skewness remains within the bounds and we allow kurtosis to vary.

$$\text{Group I:} \quad \sqrt{\beta_1} \le 0.3$$

$$\text{Group II:} \quad 0.3 < \sqrt{\beta_1} \le 1.5$$

$$\text{Group III:} \quad \sqrt{\beta_1} > 1.5$$

Neyman-Pearson (NP) tests are computed against each alternative distribution in each group 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 having minimum deviation from the power curve among the maximum deviations of all the tests.

4. Discussion of Results

Monte Carlo procedures are called in to investigate the powers of fifteen selected normality tests for samples of sizes, 25, 50 & 75, at 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 (fig. 1-3 & table 2) for all sample sizes (n=25, 50, & 75) whereas the performance of JB & RJB tests is very poor with 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 & Fig. 1). 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% & 44.7% respectively (Fig. 3). For all sample sizes, the JB & RJB tests are the least favorable options in terms of their maximum deviations (gaps) from the power curve (Fig. 2). The worst distributions for JB and RJB statistics belong to symmetric and short-tailed class of alternatives (Fig. 2 & Appendix Table 2). These results corroborate with the findings in [21-23]. 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

doi:10.20944/preprints201905.0190.v1

Peer-reviewed version available at Symmetry 2019, 11, 872; doi:10.3390/sym11070872

been evaluated by using an arbitrary reference (e.g W & AD) however, we compute the power curve by using the most powerful NP-test which yield the exact deviations of JB test from the power curve.

Table 2: Ranking of the normality tests ($\sqrt{\beta_1}$<0.3)

Slightly Skewed								
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%
W	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%

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, & BCMR are better choices for the slightly skewed alternatives. Overall, for slightly skewed distributions, COIN & W tests exhibit same power properties (Fig. 5 & 6) whereas Wsf & D statistics are not matching the standards set by other members of the group (Fig. 7 & 8) with maximum power losses 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 alternative. This result strengthens the findings in [17].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 May 2019 doi:10.20944/preprints201905.0190.v1

Peer-reviewed version available at Symmetry 2019, 11, 872; doi:10.3390/sym11070872

4.1.3. Performance of the ECDF tests

Among the ECDF class of normality tests, for slightly skewed alternatives, AD statistic is sharing the second rank with COIN & CS, third rank with CS and first rank with Tw and Rsj tests of normality for samples of size 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 & Zc and sixth rank (maximum loss of power is 49.9%) with Zc & BCMR for samples of size 25 & 50 respectively. For samples of size 75, KS test again holds the third rank with 45.1% maximum loss of power while Za & Zc tests are at the fourth rank with 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 & 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 [24] due to the absence of invariant benchmark-power envelope in their comparison.

4.1.4. Performance of the special test

This category includes only the Rsj test of normality. The performance of Rsj test increases with the increase in sample size for the slightly skewed alternatives. It holds the third, second and first rank for samples of size 25, 50 & 75 respectively (table 2). On balance, Rsj performed well (Fig. 10), especially from 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 & CS following closely whereas RJB, JB & D are the least favorable options.

4.2. Moderately Skewed Alternatives

For moderately skewed alternatives, for smaller sample size, CS, W, AD and BCMR are the best choices and the COIN test is the least favorable option (Table 3). For medium sample size, AD is the ranked one statistic and the COIN, & Tw tests are at the bottom of the ranking table. For larger sample size, AD, CS, W & BCMR appear to be the best options whereas the COIN & 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 smaller sample size, K2 occupies the fourth rank (with 46.7% maximum power loss) by outperforming the other group members. For medium sample size, JB and RJB (with power losses above 50.0%) move to the fourth rank by pushing K2 down to fifth rank whereas Tw shares the seventh rank (maximum power loss is 78.4%) with the COIN test.

With the increase in sample size, both JB & RJB show improvement in power and ranking but their maximum power losses are still above 50% (table 3). Both JB & RJB are good at discriminating the FAR group of distributions (where the power of NP-test is between 90-100%) with JB having a slight edge over RJB but both suffers when the distributions are from INTERMEDIATE[1] group of alternatives (Fig. 12).

---

[1] Following Islam [12] we group alternative space into three categories based on the power of NP-test: FAR, INTERMEDIATE & NEAR. The alternative distributions where the power of NP-test is between 90-100%, 40-90% & 5-40% are categorized as FAR, INTERMEDIATE & NEAR group of alternatives respectively.

Table 3: Ranking of the normality tests ($0.3 < \sqrt{\beta_1} \leq 1.5$)

Moderately Skewed								
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%

4.2.2. Performance of the regression & correlation tests

Among the regression and correlation based normality tests, for 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 medium up to large sample size (n=50 & 75), W, BCMR, & CS are the better options, with Wsf following closely. The D & COIN tests are the least favorable regression and correlation based normality statistics for moderately skewed alternatives which is in line with the findings in Coin, (2008) and Bonett & Seier, (2002). It is evident from figure 13; Tw & COIN both suffers against INTERMEDIATE & FAR group of alternative distributions.

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.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 May 2019 doi:10.20944/preprints201905.0190.v1

Peer-reviewed version available at Symmetry 2019, 11, 872; doi:10.3390/sym11070872

For smaller and larger sample size, the Za, & Zc statistics share the second rank. For medium sample size, these tests occupy the third rank. For smaller and medium sample size, the KS test holds the third rank whereas its position improves to second rank for larger sample size. The W test turns out to be a better test than KS (Fig. 11) which corroborates with the findings in Shapiro, Wilk, & Chen [20]. While evaluating the stringencies of the normality statistics for moderately skewed alternatives, we conclude the same but through a superior & reliable procedure.

4.2.4. Performance of the other tests

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

Overall, AD, CS, W & BCMR happen to be the best and JB, RJB, Tw, Rsj & 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 of the alternatives from FAR group only where the most powerful NP-test has 100% power. As both skewness and kurtosis are high for this group of alternatives so they are palpable. All normality tests other than the COIN & Tw statistics performed well against highly skewed alternatives (Table 4).

For smaller sample size, the Wsf, BCMR, W, CS, Za, Zc, AD, RJB & JB tests performed well with the maximum power loss ranges 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 & Tw tests is below the mark.

As the sample size increases, it becomes harder to differentiate among the selected tests of normality excluding Tw & COIN. The results are evident that the power loss of these statistics decreases with the increase in (i) sample size and (ii) the 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 as compared to other normality statistics. The poor performance of the COIN test is understandable as it is meant only for perfect symmetric cases [6 & 17]. 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.

Table 4: Ranking of normality tests for highly skewed alternatives ($\sqrt{\beta_1} > 1.5$)

Highly Skewed								
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%

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 May 2019 doi:10.20944/preprints201905.0190.v1

Peer-reviewed version available at Symmetry 2019, 11, 872; doi:10.3390/sym11070872

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%	CS	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%	KS	1	3.6%	KS	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%

5. Conclusion

This study shed light on the performance of the selected fifteen normality tests against the three different groups of alternatives. For slightly skewed alternative distributions, Tw is the best test with COIN, AD, CS & Rsj following closely. On balance, D, JB, RJB, K2, Wsf & Za did not perform well for the slighty skewed alternatives especially from medium (n=50) up 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, & 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 & K2 tests perform poorly against moderately skewed distributions. The performance of JB & 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 smaple sizes (n=75).

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

The above findings confirm our argument that comparison of tests against different alternatives yields different statistics as best tests. The COIN & Tw are 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-power envelope.

Funding: This research received no external funding.

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

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

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 May 2019 doi:10.20944/preprints201905.0190.v1

Peer-reviewed version available at Symmetry 2019, 11, 872; doi:10.3390/sym11070872

Appendix A

Table 1: Alternative Distributions

Sr. No	Student t Distribution				Mixture Distribution				
	t1		t2						
	d.f	Mean	d.f	Mean	Alpha	Mean	SD	$\sqrt{\beta_1}$	β_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
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

doi:10.20944/preprints201905.0190.v1

Peer-reviewed version available at *Symmetry* **2019**, *11*, 872; doi:10.3390/sym11070872

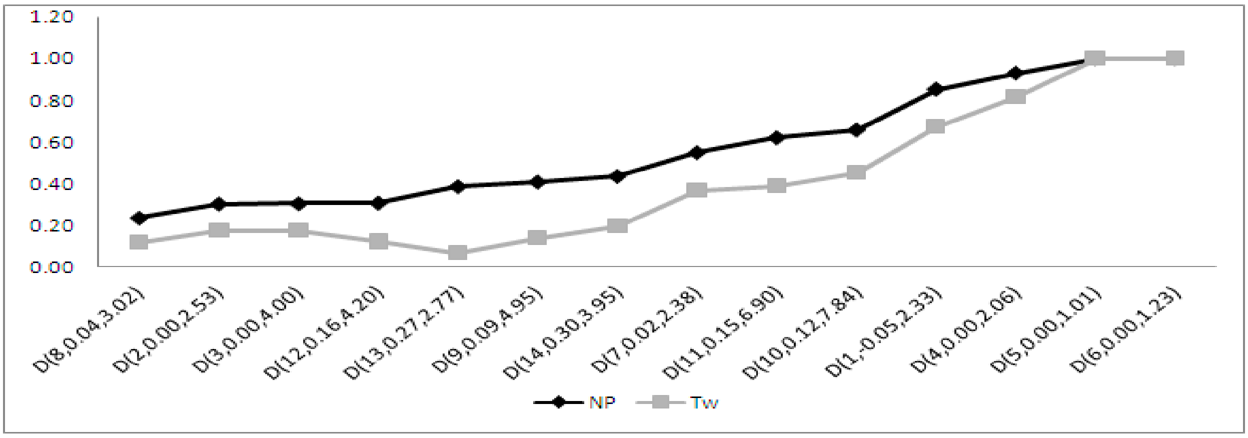
35	5	10.0	7	25.0	0.90	11.50	4.68	2.36	7.35
36	8	-10.0	12	5.0	0.90	-8.50	4.64	2.42	7.48
37	10	3.0	5	50.0	0.90	7.70	14.15	2.64	8.06

Table 2: Power comparison for 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

Fig. 1[2]: Power Comparison of Normality Tests ($\sqrt{\beta_1} < 0.3$ & $n = 75$)



[2] The convention used to refer to any distribution from table 1 is D(Sr. No., Skewness, Kurtosis). For example; D(17, 0.38, 4.65) means a distribution from table 1 with serial number 17 has a skewness and kurtosis equal to 0.38 and 4.65 respectively.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 May 2019 doi:10.20944/preprints201905.0190.v1

Peer-reviewed version available at *Symmetry* **2019**, *11*, 872; doi:10.3390/sym11070872

Fig. 2: Power Comparison of Normality Tests ($\sqrt{\beta_1} < 0.3$ & $n = 75$)

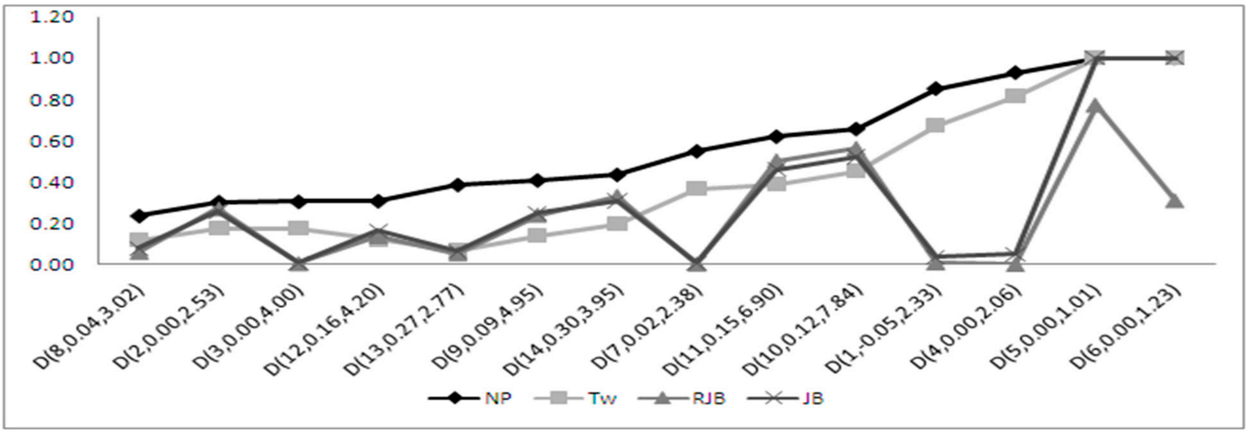


Fig. 3: Power Comparison of Normality Tests ($\sqrt{\beta_1} < 0.3$ & $n = 75$)

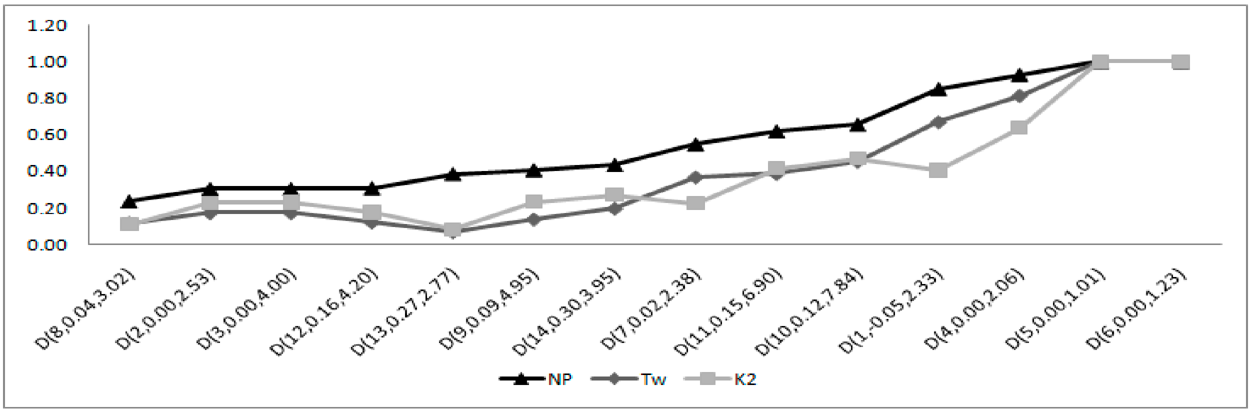
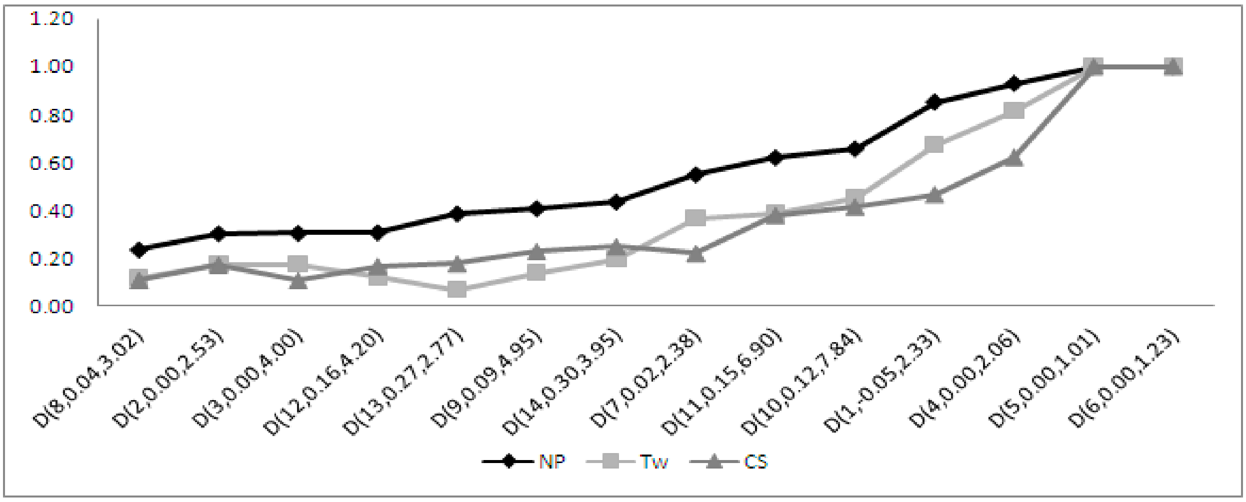


Fig. 4: Power Comparison of Normality Tests ($\sqrt{\beta_1} < 0.3$ & $n = 75$)



Peer-reviewed version available at Symmetry 2019, 11, 872; doi:10.3390/sym11070872

Fig. 5: Power Comparison of Normality Tests ($\sqrt{\beta_1} < 0.3$ & $n = 75$)

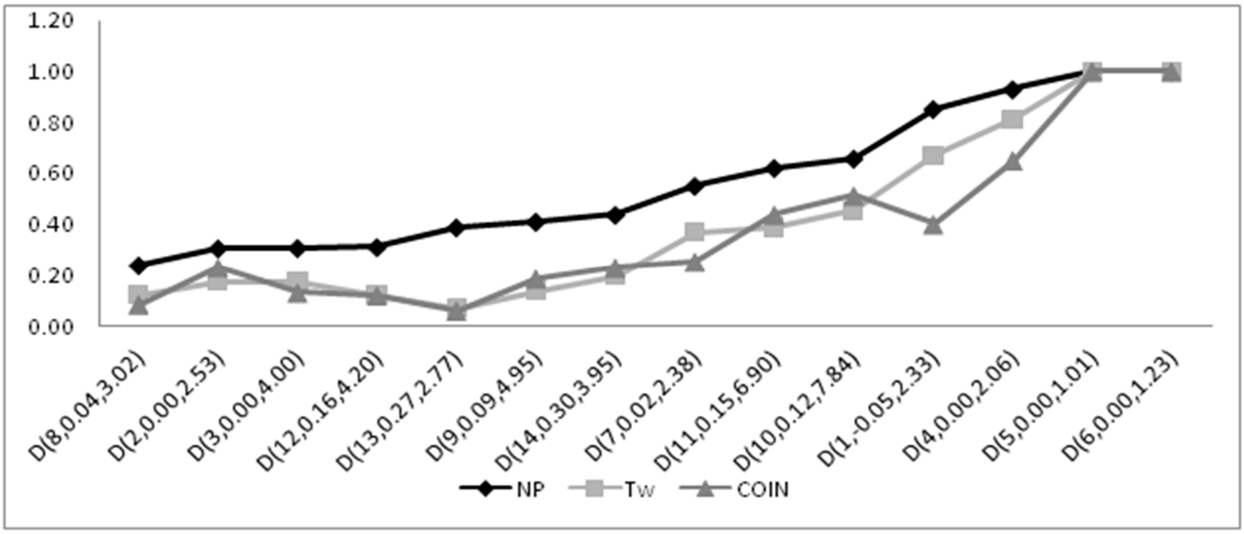
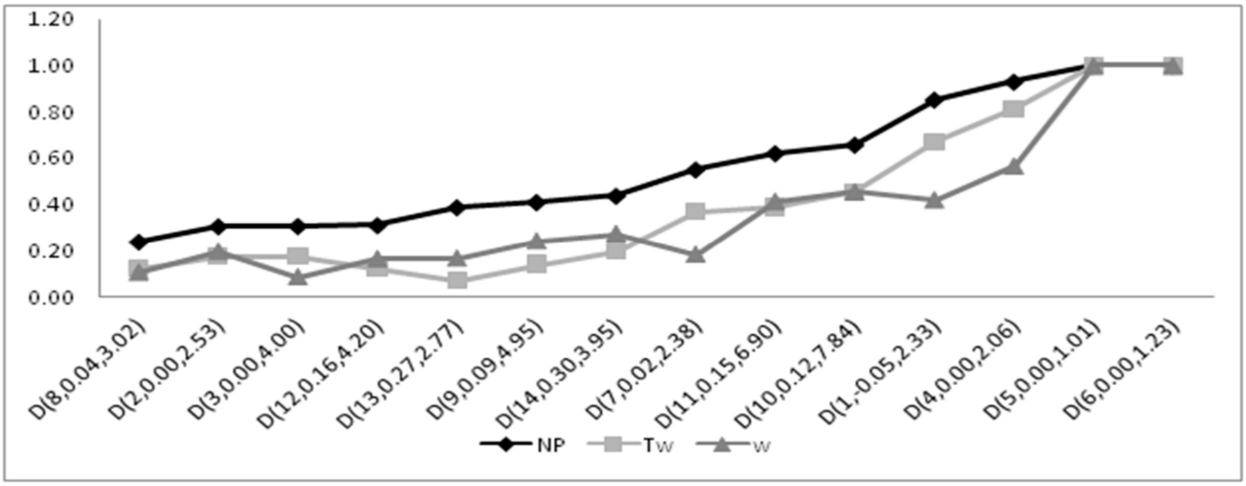


Fig. 6: Power Comparison of Normality Tests ($\sqrt{\beta_1} < 0.3$ & $n = 75$)



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 May 2019 doi:10.20944/preprints201905.0190.v1

Peer-reviewed version available at Symmetry 2019, 11, 872; doi:10.3390/sym11070872

Fig. 7: Power Comparison of Normality Tests ($\sqrt{\beta_1} < 0.3$ & $n = 75$)

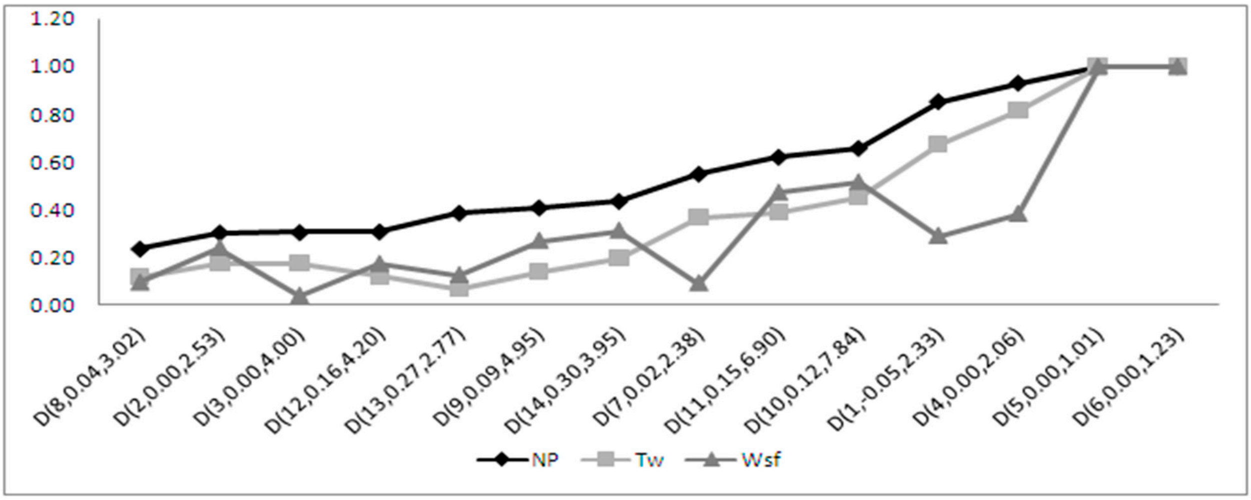


Fig. 8: Power Comparison of Normality Tests ($\sqrt{\beta_1} < 0.3$ & $n = 75$)

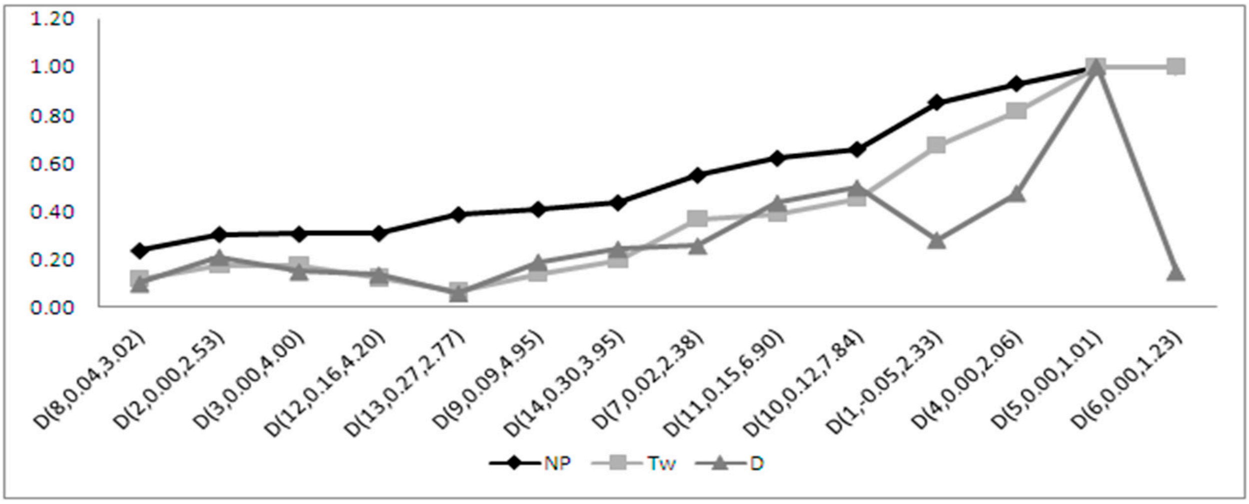
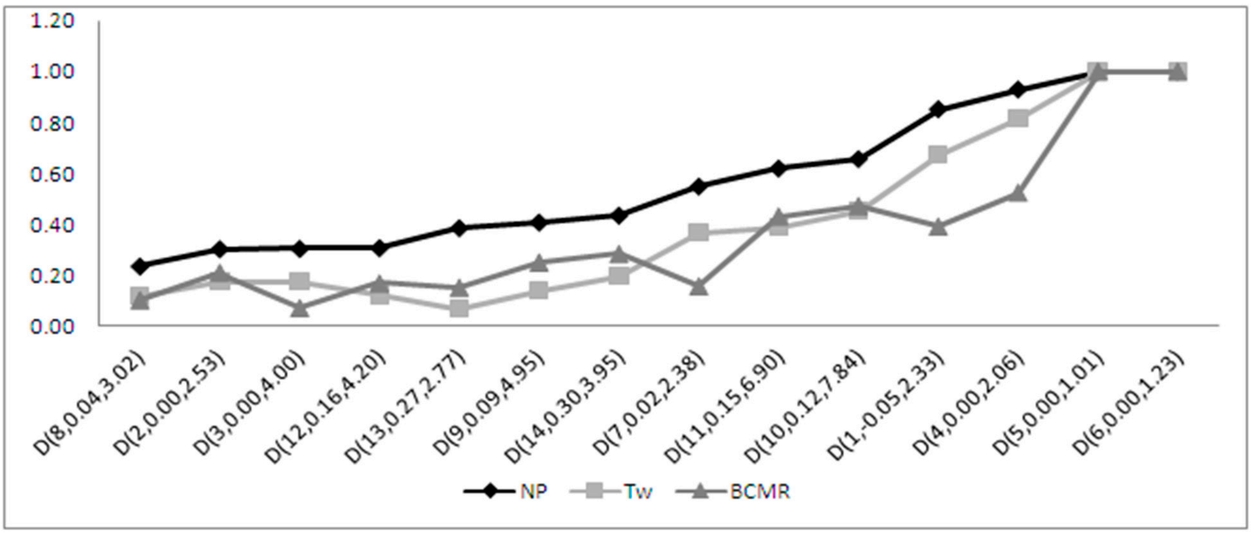


Fig. 9: Power Comparison of Normality Tests ($\sqrt{\beta_1} < 0.3$ & $n = 75$)



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 May 2019 doi:10.20944/preprints201905.0190.v1

Peer-reviewed version available at *Symmetry* **2019**, *11*, 872; doi:10.3390/sym11070872

Fig. 10: Power Comparison of Normality Tests ($\sqrt{\beta_1} < 0.3$ & $n = 75$)

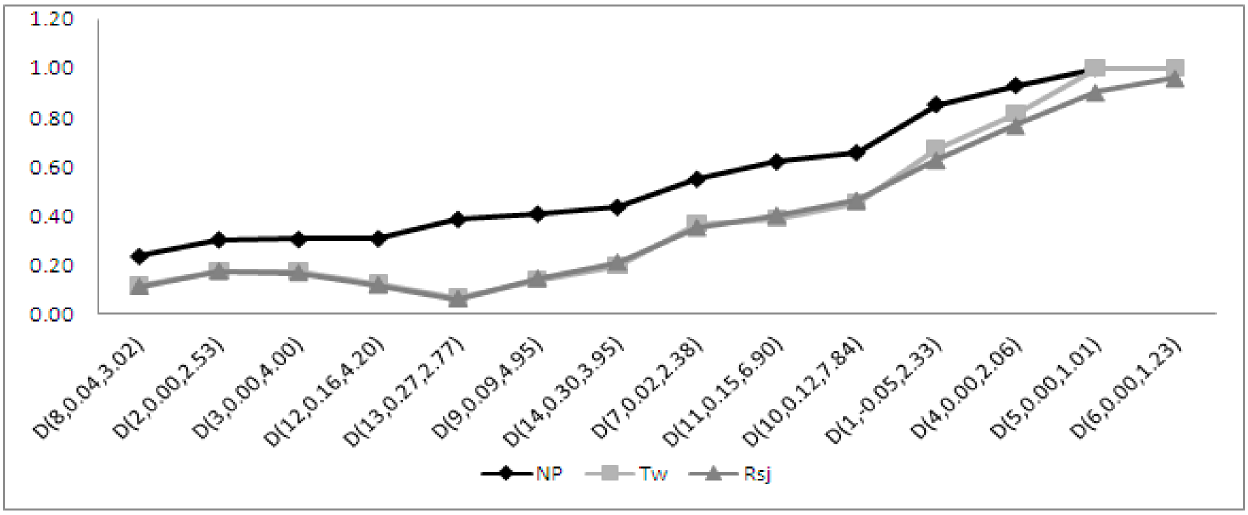


Fig. 11: Power Comparison of Normality Tests ($0.3 < \sqrt{\beta_1} \leq 1.5$ & n = 75)

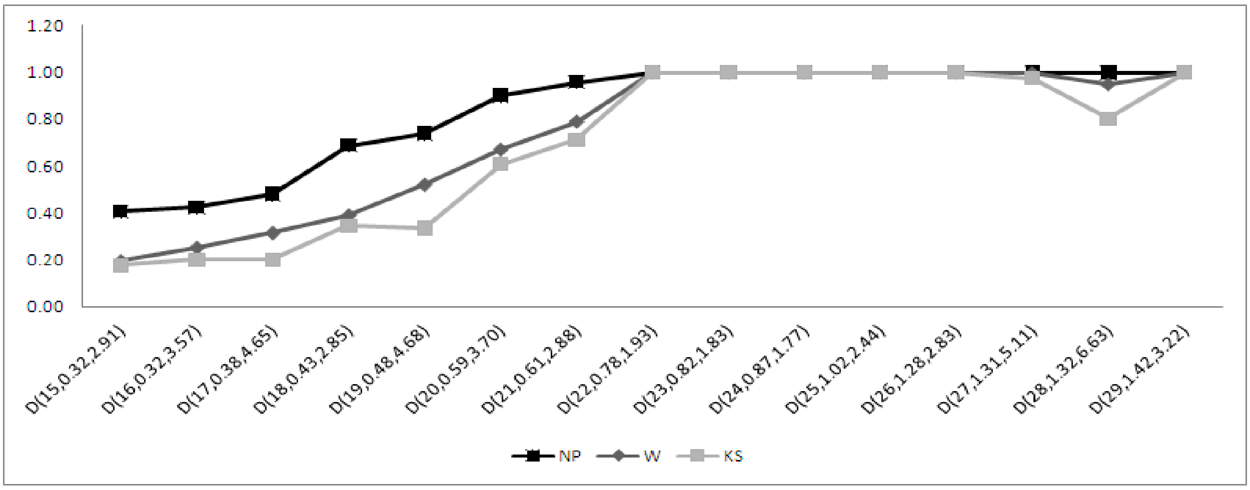
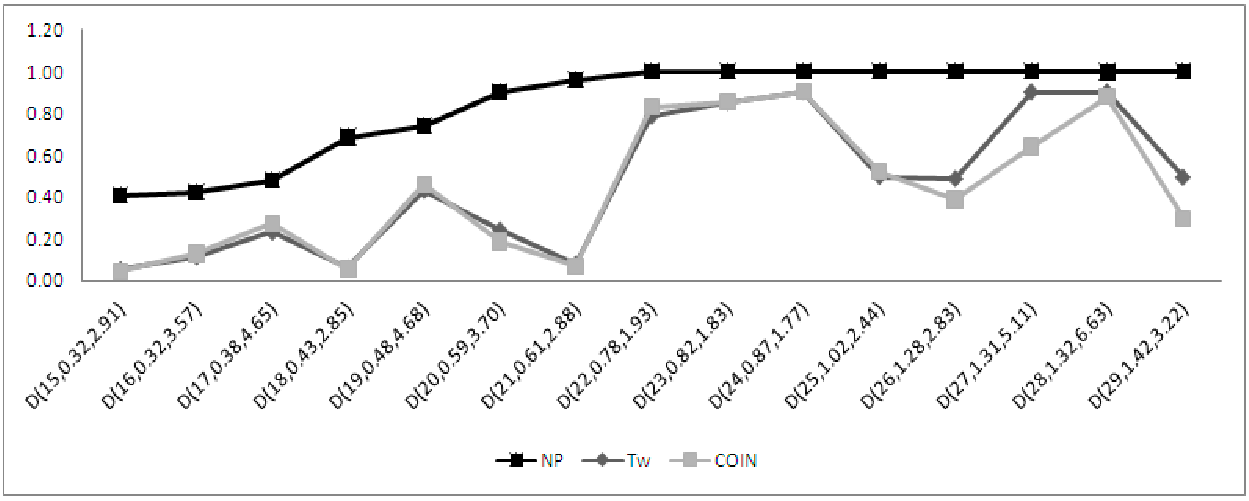
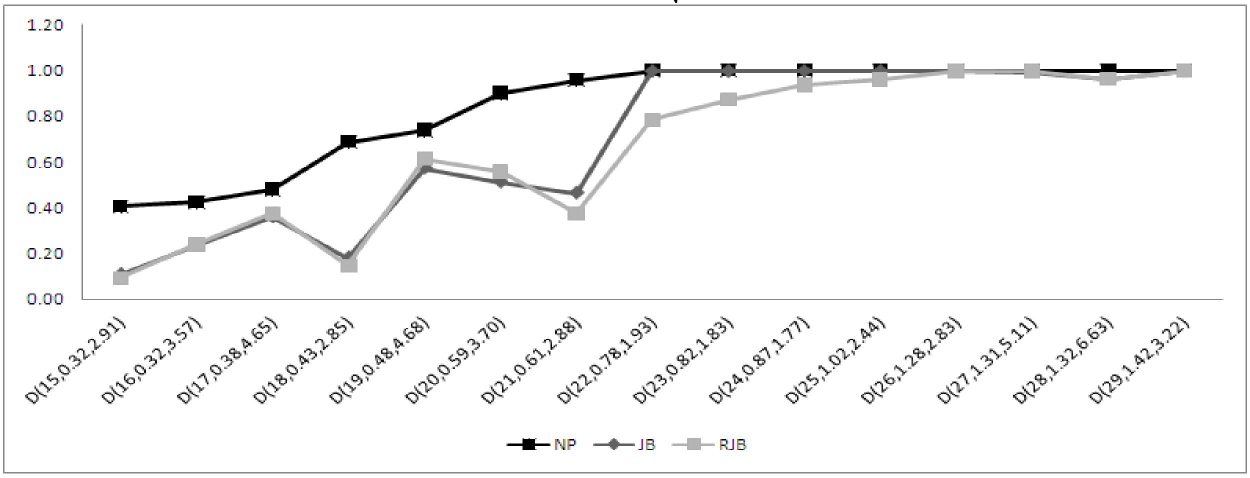


Fig. 12: Power Comparison of Normality Tests ($0.3 < \sqrt{\beta_1} \leq 1.5$ & n = 75)



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 May 2019 doi:10.20944/preprints201905.0190.v1
Peer-reviewed version available at Symmetry 2019, 11, 872; doi:10.3390/sym11070872

263 Fig. 13: Power Comparison of Normality Tests ($0.3 < \sqrt{\beta_1} \leq 1.5$ & n = 75)



264

References

Anderson, T. W., Darling, D. A. A test of goodness of fit. Journal of the American Statistical Association 1954; 49(268): 765–769.

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. Journal of Statistical Computation and Simulations. 2012: 21-38.

Blanca, M. J., Arnau, J., López-Montiel, D., Bono, R., & Bendayan, R. Skewness and Kurtosis in Real Data Samples. Methodology. 2013; 9: 78-84.

Bonett, D. G., Seier, E. A test of normality with high uniform power. Computational Statistics & Data Analysis. 2002; 40: 435-445.

Bowman, K. O., Shenton, L. R. Omnibus test contours for departures from normality based on $\sqrt{b_1}$ and b2. Biometrika. 1975; 62(2): 243–250.

Coin, D. A goodness-of-fit test for normality based on polynomial regression. Computational Statistics & Data Analysis. 2008; 52: 2185-2198.

D'Agostino, R., Pearson, E. S. Tests for departure from normality. Empirical results for the distributions of b2 and $\sqrt{b_1}$. Biometrika. 1973; 60(3): 613–622.

Delong, J. B., Summers, L. H. "Are Business Cycle Symmetrical," in American Business Cycle: Continuity and Change. University of Chicago Press. 1985: 166–178.

Gel, Y. R., Gastwirth, J. L. A robust modification of the Jarque–Bera test of normality. Economics Letters. 2008; 99 (1): 30-32.

Gel, Y. R., Miao, W., Gastwirth, J. L. Robust directed tests of normality against heavy-tailed alternatives. Computational Statistics & Data Analysis. 2007; 51: 2734–2746.

Henderson, A. R. Testing experimental data for univariate normality. Clinica Chimica Acta. 2006; 366:112-129.

Islam, T. U. Stringency-based ranking of normality tests. Communications in Statistics - Simulation and Computation. 2017; 46(1): 655-668.

Jarque, C. M., Bera, A. K. A Test for Normality of Observations and Regression Residuals. International Statistical Review. 1987; 55(2): 163-172.

Lehmann, E. L., Stein, C. On the Theory of Some Non-Parametric Hypotheses. Ann. Math. Statist. 1949; 20(1):28–45.

Önder, A. Ö., Zaman, A. Robust tests for normality of errors in regression models. Economics Letters. 2005; 86(1): 63-68.

Pearson, E. S., D' Agostino, R. B., Bowman, K. O. Tests for Departure from Normality: Comparison of Power. Biometrika. 1977; 64(02): 231-246.

Romao, X., Delgado, R., Costa, A. An empirical power comparison of univariate goodness-of-fit tests for normality. Journal of Statistical Computation and Simulation. 2010; 80(5): 1-47.

Peer-reviewed version available at *Symmetry* **2019**, *11*, 872; doi:10.3390/sym11070872

Shapiro, S. S., Francia, R. S. An approximate analysis of variance test for normality. Journal of American Statistical Association. 1972; 67: 215–216.

Shapiro, S. S.,Wilk, M. B. An analysis of variance test for the exponential distribution (Complete samples). Biometrika. 1965; 54(3/4): 591–611.

Shapiro, S. S.,Wilk, M. B., Chen, H. J. A comparative study of various tests for normality. Journal of the American Statistical Association. 1968; 63(324):1343–1372.

Thorsten, T., Buning, H. Jarque-Bera Test and its Competitors for Testing Normality- A Power Comparison. Journal of Applied Statistics. 2007; 34(1): 87- 105.

Yap, B. W., Sim, C. H. Comparisons of various types of normality tests. Journal of Statistical Computation and Simulation. 2011: 1-15.

Yazici, B., Yolacan, S. A comparison of various tests of normality. Journal of Statistical Computation and Simulation. 2007; 77(02): 175-183.

Zhang, J., & Wu, Y. (2005). Likelihood-ration tests for normality. Computational Statistics & Data Analysis. 2005; 49: 709-721.