No.460

A Combined Test of Normality and Its Comparison to Other Test Statistics

by

Tsunemasa Shiba

May 1991

A Combined Test of Normality and Its Comparison to Other Test Statistics

Tsunemasa Shiba*
Institute of Socio-Economic Planning, University of Tsukuba
Tennoudai 1-1-1, Tsukuba-shi 305 Japan.

May 1991

1 Introduction

Various different tests of normality have been proposed, and their powers in Monte Carlo experiments have shown that the skewness and kurtosis statistics ($\sqrt{b_1}$ and b_2), D'Agostino-Pearson K^2 statistic that combines $\sqrt{b_1}$ and b_2 (D'Agostino and Pearson (1973)) and the Shapiro-Wilk test (Shapiro and Wilk (1965)) tend to dominate the other statistics. Recently D'Agostino *et al.* (1990) have demonstrated that the $\sqrt{b_1}$, b_2 and K^2 tests are readily programmable on popular statistical software such as SAS, and thus their use can become prevalent.

Given the prolifiration of computers and statistical softwares, the use of a computer ought to be encouraged to compute a test statistic or to derive an exact size of the test. In testing normality, there are many directions of the alternative hypotheses, and thus there is a need for combining some tests which are known to be powerful in different directions of alternatives. The K^2 test is such a combined test.

In this paper I propose a test that is combined in the way Cox and Hinkley (1977) suggest, and provide an accurate significance level which is easily derivable. I combine the Shapiro-Wilk test and the K^2 test, and I call it the combined test, CB test hereafter in this paper. The reason I combine the K^2 and Shapiro-Wilk tests is as follows. Although other tests, such as the Kolmogorov-Smirnov test, may have reasonable powers occasionally with certain alternative distributions, there seems to be a general agreement[1] that the SW and K^2 test statistics are the best performing tests of all. I shall compare the CB test's power to each of the above tests, and also to some other tests such as the Kolmogorov-Smirnov test, Geary's test (Geary (1947)), Anderson and Darling's test (Anderson and Darling (1954)), and $\sqrt{b_1}$ and b_2 tests (Pearson, D'Agostino and Bowman (1977)). I have selected these statistics to compare powers, because of the omnibus property, and the fact that they are easy to compute. The Kolmogorov-Smirnov (KS hereafter) test has been widely quoted as *the* test statistic for testing normality, along with the chi-square goodness of fit test. I am not taking up the chi-square goodness of fit test in this paper, because it is not recommended as the statistic to be used in testing normality at all (see D'Agostino (1986)). Fujino (1976) says that Geary's test is not bad when compared to the Shapiro-Wilk (SW hereafter) test. D'Agostino (1986) reviews many tests of normality based on

*I am grateful to Professors Hiroki Tsurumi and Taku Yamamoto for helpful discussions.
[1]See for example D'Agostino (1986).

empirical distribution function, and concludes that the Anderson-Darling (*AD* hereafter) test is the best, even compared to the more popular *KS*. Mardia (1980) among others rate the *SW* as the best performing test overall.

In all the Monte Carlo power comparisons conducted so far, different tests were compared under *pure* alternatives only. That is, none of the studies compared powers under distributions local to null or moderately distant from the null distribution of normality. In practice, the distribution under question may not be *quite* different from the normal distribution but may be close to it in the sense that it may be predominently normal. In this paper, I have traced out powers curves of different statistics under various alternative distributions by gradually changing the linear combination coefficient between the null normal distribution and the alternative. I have, thus, created a continuous space given an alternative nonnormal distribution. As a by-product to these experiments, I have discovered that known critical values of many test statistics are sometimes inaccurate. I therefore computed exact exmpirical critical values of the above statistics by generating 30,000 samples.

The plan of this paper is as follows. In section 2, I introduce the *CB* (combined) test statistic. Power curve comparisons are given in section 3. This section also contains my interpretation of the results, and these should serve as my concluding remarks to this paper.

2 A Combined Test Statistic

D'Agostino (1986, pp.493–404) surveys the power of various tests of normality. As to the K^2 and *SW* tests, he concludes:

- For $\beta_1 \neq 0$ nonsymmetric alternative distributions, the *SW* is "clearly the most powerful."
- For $\beta_1 = 0$ and $\beta_2 < 3$ platykurtic alternative distributions, the *SW* is "often more powerful than" the K^2 test.
- For $\beta_1 = 0$ and $\beta_2 > 3$ leptokurtic alternative distributions, the K^2 "is often most powerful."

I, thus, would want to combine these tests[2] and obtain a test that has better power than that of the smaller power of either K^2 or *SW* in all three different types of alternative distributions cited in above.

The *CB* statistic that I propose rejects the null hypothesis of normality, when either one of the K^2 or *SW* rejects the null. Letting $\phi_i(x)$ be a test function of ith test statistic given data x, $\phi_{CB}(x)$ becomes

$$\phi_{CB}(x) = \begin{cases} 1 & \phi_{K^2}(x) = 1 \text{ or } \phi_{SW}(x) = 1 \\ 0 & \text{otherwise.} \end{cases}$$

Such a test is expected to perform well against union of the alternative hypotheses that each of K^2 and *SW* has higher power in. The remaining question is: how to control the size of the *CB* test? Let $\alpha_i \equiv E(\phi_i(x) \mid H_o)$, the size of ith test, then

$$\alpha_{CB} \geq max(\alpha^*_{K^2}, \alpha^*_{SW}),$$

[2]See Cox and Hinkley (1974, pp.77–) for some discussions on combining tests. In p.77, they specifically talk about choosing a test that is "most significant" from k different tests. As seen below, my method collapses to their method when $k = 2$, but my method provides a way of computing critical values of mixing tests by simulation given desired level of significance of the combined test.

Table 1: Exact Critical Value

Significance Level	$\sqrt{b_1}$	b_2	K^2	KS	AD	GE	SW
.01	2.612	2.711	11.398	1.531	3.945	2.547	0.932
.05	1.971	1.991	6.363	1.273	2.518	1.937	0.948
.10	1.662	1.653	4.586	1.134	1.954	1.631	0.955

Notes: number of replications= 30,000, used sample size=50,
K^2 =D'Agostino-Pearson statistic, KS =Kolmogorov-Smirnov statistic,
AD =Anderson-Darling statistic, GE =Geary's statistic,
SW =Shapiro-Wilk statistic.

where α_i^* ($i = K^2, SW$) denotes the adopted significance level of the ith test. In other words, since critical regions of the K^2 statistic and the SW statistic are not disjoint under the null of normality, I cannot control the size of the CB statistic at a given level easily. Moreover, there are multiple sets of $\alpha_{K^2}^*$ and α_{SW}^*, for a fixed level of α_{CB}. Since there is no prior information as to the uniform dominance of K^2 or SW over class of alternative distributions, I shall place a sort of non-informative or diffuse prior distribution in a Bayesian sense to the two statistics, *i.e.*, I let $\alpha_{K^2} = \alpha_{SW}$. I am, thus, assigning an equal weight to the two statistics. It remains to determine what the *size* should be. Suppose that I want to control the size of the CB statistic, α_{CB}, to be 5 percent. Then I would compute α_{CB}, for grids of equal significance levels of the K^2 statistic and the SW statistic (for example, 5.0 percent, 4.9 percent, 4.8 percent,*etc.*). I shall choose the significance level that gives the size of the CB statistic to be 5.0 percent exactly. For the sake of simplicity, I shall confine myself to the significance level of 5 percent[3] and the sample size of 50 throughout in this paper. Obviously, these figures may be changed easily when needed. The sample size of "50" is selected because it is a moderate sample size, and also Shapiro and Wilk (1965) published the coefficients needed for the SW statistic only up to the sample size of 50. To compute the CB statistic for the sample size, n, greater than 50, I may replace the SW statistic by the one introduced by Shapiro and Francia (1972).

Table 1, gives critical values of the various different test statistics for the significance level of 5 percent, determined from 30,000 samples. The samples were drawn from $N(0,1)$. The K^2 and AD test statistics use empirical distribution functions and they need to be standardized so that they are comparable to the $N(0,1)$ distribution function. I assumed $\mu = 0$ and $\sigma = 1$ in calculating the K^2 and AD test statistics. For the rest of the test statistics, I used the sample moments instead of assuming $\mu = 0$ and $\sigma = 1$. The figures reported in this table vary to different degrees from the published available critical values. For instance, if I used the 5 percent chi-squared 2 degrees of freedom critical value, 5.79, then since this figure is smaller than 6.363 in the table, the size of the resultant K^2 test would be larger than the 5 percent that I want. It would be easy to compute the exact critical values for significance levels and sample sizes other than in Table 1. The sample size that was used to compute the figures in Table 1 is 50. It might be necesary to see if larger sample size might give systematically different critical values.

[3]Except for Table 1 below, in which I present 1, 5 and 10 percent exact empirical critical values of various statistics.

Critical values of the K^2 and SW statistics associated with significance levels less than 5 percent, were computed empirically, this time using 20,000 samples of standard normal variates. Using these critical values, I can now control each test's significance level. For each significance level, the size of the CB statistic was calculated. I am not presenting the numbers from this exercise. In computing the two mixing statistics, I used sample mean and variance instead of the parameter values assigned in data generation, *i.e.*, zero and one. If I let the two test statistics' significance level to be 3.5 percent, say, then the resultant size of the CB test exceeded 5.0 percent. By reducing it, the size of the CB statistic approached 5.0 percent. When the significance levels of the two statistics were controlled at 3.25 percent and 3.0 percent, respective size of the CB statistic was 5.107 percent and 4.763 percent. I, therefore, interpolated the numbers in such a way that the CB statistic has its size 5 percent. In other words, I have found $\alpha^*_{K^2} = \alpha^*_{SW} = 3.172$ percent to give the size of the CB statistic to be 5 percent exactly, and associated critical values of the K^2 and SW statistics to be 7.692 and 0.944, respectively.

Some comments about the CB test may be in order. First, since the K^2 test and SW test are easily computed, so is the CB test. Secondly, with the CB test I may control its significance level easily. In this paper, I am presenting its critical value for 5 percent significance level determined from its empirical distribution under the null. There is no need to compute K^2's and SW's $p-value$, for instance, to combine tests. Third, the above critical values of the K^2 and the SW apply whenever I want to control the significance level of the CB statistic to be 5.0 percent. It does not matter whether the data at hand is suspected to have come from the standard normal distribution or not. Fourth, the CB test is expected to perform well under the union of alternative distributions that the K^2 or the SW have high powers in. In summary, the CB test is expected to be a test that has reasonably high power in many directions of alternatives, and yet easily implemented.

3 Powers Compared Empirically

Table 2 presents the alternative distributions employed in power comparison. I have selected these distributions mainly in view of prior power comparison studies such as Pearson, D'Agostino and Bowman (1977), Fujino (1976) and Shapiro, Wilk and Chen (1968). For the reader's convenience, I have put the population distribution number of Pearson, D'Agostino and Bowman (1977, p.240) in the last column of the table. The distributions in the table may be divided into two groups: symmetric $\beta_1 = 0$ distribution that are numbered #1 to 7, and skewed $\beta_1 \neq 0$ distributions that are numbered #8 to 14.

In tracing out a power curve of a given test statistic, I generated 4,000 samples of size 50, using the following equation:

$$x = (1-\lambda)N + \lambda A,$$

where x is the vector of generated random variables, N is the vector of independent normal random variates with its mean and standard deviation given in Table 2 for each A, the vector of random variates of an alternative distribution, and λ is the scalar linear combination coefficient. Note that

- Distributions considered in many previous Monte Carlo studies of normality tests, *e.g.*, Shapiro, Wilk and Chen (1968), Fujino (1976) and Pearson, D'Agostino and Bowman (1977), among others, are obtained only for $\lambda = 1$. My Monte Carlo design allows me to generate a continuous alternative space given each H_1 distribution. Comparing powers

Table 2: Alternative Distributions Used

Sequence Number	Distribution	Expected Value	Standard Deviation	$\sqrt{\beta_1}$	β_2	PDB pop. no.
#1	Uniform(0, 1)	.500	.289	0	1.800	3
#2	Johnson SB(0, 1)	.500	.208	0	2.140	n.a.
#3	Logistic(1, 1)	0	1.814	0	4.200	15
#4	Cauchy	(0)	(4.266)	n.a.	n.a.	32
#5	Student t(2)	0	(2.344)	n.a.	n.a.	31
#6	Student t(4)	0	1.410	0	n.a.	30
#7	Student t(7)	0	1.183	0	5.000	n.a.
#8	Johnson SB(1, 2)	.384	.112	.288	2.774	39
#9	Weibull	.886	.463	.631	3.245	44
#10	Johnson SB(1, 1)	.303	.183	.727	2.904	40
#11	$\chi^2(4)$	4	2.828	1.410	6.000	50
#12	Exponential	1	1	2.000	9.000	53
#13	Log-normal(1, 1)	1.649	2.161	6.186	113.900	58
#14	Extreme Value(0, 1)	-.580	1.280	-1.140	5.400	n.a.

Notes: Above distributions are generated as follows:
Uniform(0, 1): $x = R$, Log-normal(1, 1): $x = e^Z$,
Exponential: $x = -log(R)$, Weibull: $x = \sqrt{-log(R)}$,
Johnson SB(0, 1): $x = e^Z/(1 + e^Z)$, Johnson SB(1, 1): $x = e^{Z-1}/(1 + e^{Z-1})$,
Johnson SB(1, 2): $x = e^{\frac{Z-1}{2}}/(1 + e^{\frac{Z-1}{2}})$, Cauchy: $x = Z_i/Z_j$,
Student t(2): $x = Z_i/\sqrt{\sum_{k=1}^{2} Z_k^2/2}$, Student t(4): $x = Z_i/\sqrt{\sum_{k=1}^{4} Z_k^2/4}$,
Student t(7): $x = Z_i/\sqrt{\sum_{k=1}^{7} Z_k^2/7}$, $\chi^2(4)$: $x = \sum_{k=1}^{4} Z_k^2$,
Logistic(1, 1): $x = log(\frac{R}{1-R})$, Extreme Value(0, 1): $x = log(-log(1 - R))$,
where "x" is the generated random variable,
$\sqrt{\beta_1}$ = skewness measure, β_2 = kurtosis measure,
Z and $Z_h^\prime$s are independently distributed standard normal variates.
$(\cdot)$ in Student t distributions are computed by definite integration using wide enough integration limits. "PDB pop. no." corresponds to the distribution numbering in Pearson, D'Agostino and Bowman (1977).
"n.a." denotes that figure is not available.

in such space would be helpful for assessing local null and moderately distant from null behavior of the tests.

- I have changed the mean and the standard deviation of H_o, the null normal distribution comformably to the mean and the standard deviation of each H_1 distribution.

- For each λ, 4,000 different samples are generated; this means that the total of 4000Λ samples are generated for a power curve, where Λ = number of λ points between zero and one. x is generated under the null distribution of normality when $\lambda = 0$, and under H_1, the alternative, when $\lambda = 1$. To be precise, by changing λ from zero to one, I am *not* tracing out the power curve of a given statistic. For λ values between zero and one, generated x is simply a mixture of normal and H_1 distribution[4]. Thus for $0 < \lambda < 1$, x could be thought of as a H_1 distributed random vector. Instead of such *deterministic* mixtures, I could have dealt with compound distributions,[5] in which different parameters are probabilistically mixed.

Tables 3 to 5 give experiment results. Note that all figures in $\lambda = 0$ rows should be ".050." Due to the fact that the number of replications, 4,000, is not large enough, the figures are not necessarily ".050." However, only in 5 out of 112 cases, figures in $\lambda = 0$ rows fell outside of a 99 percent confidence interval of 4.11 percent and 5.89 percent, thus I may conclude that they are not significantly different from .05. The number of replications adopted in this paper is larger than that of Pearson, D'Agostino and Bowman (1977), in which they replicated 200 samples[6]. A quick calculation will show that 200 samples' most conservative[7] one standard error (or one sigma) of estimated power is 3.55 percent, whereas 4000 samples' is 0.79 percent.

As expected the two directional tests, $\sqrt{b_1}$ and b_2, perform well when they are supposed to, and perform very poorly, when distributions are symmetric or mesokurtic. For instance, $\sqrt{b_1}$ has its power less than 0.2 in cases #1,2,3, and 8, in which, except for #8, the H_1 distribution is symmetric. Cases #8 to 10 have $\beta_2 \approx 3$, and b_2 performs very poorly in these cases. As to the local powers of the tests, except for the Cauchy (case #4) alternative distribution, none of the tests had powers higher than 0.5 when $\lambda \leq 0.5$. This means that all the tests have low local power, and thus cannot be trusted to make distinction of a given sample whether it is from

[4]Letting $f_m(x)$ =probability density function (pdf) of mixed distribution of a random variable (rv) x, $f_i(x)$ = pdf of ith mixing distribution, Stuart and Ord (1987, p.171) call $f_m(x) = \sum_{i=1}^{m} \pi_i f_i(x)$, *finite mixture* of distributions, where π_i = mixing proportion for the ith distribution and $\sum_{i=1}^{m} \pi_i = 1$. They note that the case when $m = 2$ and $f_i(x)$ = different $N(\mu_i, \sigma_i^2)$ pdf's, was first proposed by Karl Pearson in 1894. Special cases of the above mixture of two normals, when $f_1(x)$ = standard normal pdf and $f_2(x)$ = a $N(0, \sigma^2)$ pdf and $\sigma \neq 1$, were taken up by Murota and Takeuchi (1981) and Hall and Welsh (1983), in their power comparison studies and they denoted the resultant $f_m(x)$ distribution, $CN(0, \sigma^2; \pi)$ and called it the contaminated normal distribution. Fujino (1976) took up more general $CN(\mu, \sigma^2; \pi)$. In their power comparison study, Pearson, D'Agostino and Bowman (1977) generated $CN(0, \sigma^2; \pi)$ and $CN(\mu, 1; \pi)$, and they call these scale contaminated (normal) distribution and location contaminated (normal) distribution, respectively. Obviously, $CN(\mu, \sigma^2; \pi)$ is a finite mixture distribution. In the finite mixture distribution, given a sample, resultant rv x has $f_i(x)$ pdf $100\pi_i$ percent of the time. My linear combination is different from the CN type combination in that it is an amalgam of two differently distributed rv's. Mixed rv x is a linear combination of two rv's, and hence its pdf never coincides with the mixing pdf $f_i(x)$ unless $\lambda = 1$ or $\lambda = 0$.

[5]See Stuart and Ord (1987, p.173) for compounding $f(x \mid \theta)$ pdf using weighting function (or prior pdf in a Bayesian sense) $p(\theta)$, as $f(x) = \int f(x \mid \theta) p(\theta) d\theta$.

[6]In one of their experiments, they generate 1000 samples, however.

[7]Here, I am referring to a standard error supposing that the true proportion, p, is .50. I should now use $p = .50$ instead of $p = .05$ that I used earlier. When I obtained a 99 percent confidence interval earlier, I assumed $p = .05$ since $\lambda = 0$.

normal or not, when it is very close to the normal. Let me turn to the tests other than $\sqrt{b_1}$ and b_2. The GE, KS and AD tests perform poorly most of the times. It seems that they may be used only when the alternative distribution is *quite* different from the normal distribution, such as the log-normal, where $\sqrt{\beta_1} = 6$ and $\beta_2 = 114$. In lower degrees of freedom Student t distribution cases (#4 to 6), the GE performed well in comparison to the KS and the AD, and in some cases significantly outperformed the K^2 and the SW.

Let me turn to the CB and its related tests, the K^2 and the SW. They are the best performing tests among the tests that are considered in this paper. I shall first examine if D'Agostino (1986)'s characterization of the K^2 and SW tests, presented earlier in section 2, apply to Tables 3 to 5. For symmetric H_1 distribution with $\beta_2 < 3$, *i.e.*, cases #1 and 2, the SW indeed marginally outperforms the K^2. For symmetric and $\beta_2 > 3$ alternatives, such as cases #3 and 7, the K^2 outperforms the SW as surveyed in D'Agostino (1986). Finally, in all non-symmetric H_1 cases, *i.e.*, cases #8 to 14, the SW has better power than the K^2. For low degrees of freedom Student t alternatives, where β_2 in some cases is not defined, the K^2 tend to outperform the SW, however. Effects of the SW and the K^2 are strong on CB's performance. In particular, the CB had power close to but slightly lower than one of the SW or K^2 in all but one case, *i.e.*, case #2, where the SW and the K^2 had .248 and .224, respectively and the CB had .218, when $\lambda = 1$. I may thus conclude that on the whole, the CB performs close to the better one of the K^2 or the SW. This is what I had expected the CB to be. The SW is relatively weak in H_1 distributions that are symmetric and $\beta_2 > 3$, and the CB outperforms it in such cases. On the other hand, the SW dominates the K^2 in distributions with $\beta_2 < 3$, and the CB outperforms the K^2 in such cases. These imply that, although the CB is outperformed by the SW marginally in most non-symmetric alternative distributions, the CB does outperform the SW in distributions with $\beta_2 < 3$, and thus it is better to use the CB instead of the SW or K^2 alone.

As pointed out earlier, however, the special set of coefficients needed for computation of the SW is available only up to $n = 50$, and thus the SW may not be a good test to be combined with. Hence, for the CB statistic to have more applicability, the K^2 and the GE test may be combined.

Table 3: Power Curves #1 to #6

λ	$\sqrt{b_1}$	b_2	K^2	GE	KS	AD	SW	CB
H_1 Distribution #1: Uniform(0, 1)								
0	.048	.052	.050	.049	.055	.051	.045	.048
.25	.049	.043	.048	.045	.073	.078	.050	.050
.50	.024	.066	.045	.068	.117	.189	.059	.049
.75	.004	.571	.408	.437	.049	.057	.434	.406
1.0	.001	.877	.777	.656	.158	.151	.875	.840
H_1 Distribution #2: Johnson SB(1, 0)								
0	.051	.054	.051	.054	.052	.055	.050	.054
.25	.042	.048	.048	.053	.072	.080	.050	.050
.50	.027	.065	.039	.054	.136	.208	.052	.046
.75	.005	.238	.118	.177	.047	.053	.142	.123
1.0	.003	.392	.224	.251	.086	.082	.248	.218
H_1 Distribution #3: Logistic(1, 1)								
0	.047	.048	.048	.049	.049	.054	.046	.047
.25	.045	.048	.049	.048	.078	.079	.049	.049
.50	.089	.085	.096	.070	.186	.274	.059	.081
.75	.162	.160	.193	.144	.133	.145	.115	.167
1.0	.178	.181	.215	.177	.054	.051	.133	.184
H_1 Distribution #4: Cauchy								
0	.041	.048	.044	.045	.052	.053	.044	.048
.25	.504	.521	.531	.465	.071	.098	.474	.520
.50	.772	.834	.839	.804	.336	.635	.792	.830
.75	.887	.976	.975	.974	.854	.981	.966	.972
1.0	.909	.994	.993	.998	.916	.989	.995	.996
H_1 Distribution #5: Student t(2)								
0	.047	.051	.051	.059	.051	.051	.050	.050
.25	.124	.123	.130	.103	.093	.098	.105	.124
.50	.374	.414	.429	.356	.377	.588	.337	.405
.75	.628	.752	.756	.746	.705	.873	.677	.731
1.0	.699	.839	.842	.870	.551	.708	.796	.829
H_1 Distribution #6: Student t(4)								
0	.043	.052	.047	.052	.047	.046	.048	.048
.25	.065	.058	.067	.057	.087	.086	.057	.062
.50	.183	.190	.210	.150	.236	.352	.143	.196
.75	.344	.403	.427	.370	.230	.307	.311	.389
1.0	.389	.476	.483	.463	.088	.090	.374	.451

References

[1] Anderson, T.W., and D.A.Darling, 1954, "A test of goodness of fit," *Journal of the American Statistical Association,* 49, 765–769.

[2] Cox, D.R. and D.V. Hinkley, 1974, *Theoretical statistics,* Chapman and Hall, London.

[3] D'Agostino, R.B., 1986, "Tests for the normal distribution," in 367–419 of *Goodness-of-fit Techniques,* R.B. D'Agostino and M.A. Stephens, eds., Marcel Dekker, New York.

[4] D'Agostino, R.B., A. Belanger, and R.B. D'Agostino, Jr., 1990, "A suggetion for using powerful and informative tests of normality," *American Statistician,* 44, 4, 316–321.

[5] D'Agostino, R.B., and E.S. Pearson, 1973, "Tests for departure from normality. Fuller empirical results for the distribution of b_2 and $\sqrt{b_1}$," *Biometrika,* 60, 613–622.

[6] Fujino, K., 1976, "Bunpu no seikisei ni kansuru hikakukenkyu," (in Japanese) *Hinshitsu,* 6,4,18–26.

[7] Geary, R.C., 1947, "Testing for normality," *Biometrika,* 34, 209–242.

[8] Hall, P., and A.H.Welsh, 1983, "A test for normality based on the empirical characteristic function," *Biometrika,* 70, 2, 485–489.

[9] Mardia, K.V., 1980, "Tests of univariate and multivariate normality," in 279–320 of *Handbook of Statistics,* P.R.Krishnaiah, ed., North-Holland, Amsterdam.

[10] Murota, K., and K. Takeuchi, 1981, "The studentized empirical characteristic function and its application to test for the shape of distribution," *Biometrika,* 68, 1, 55–65.

[11] Pearson, E.S., R.B. D'Agostino, and K.O.Bowman, 1977, "Tests for departure from normality: Comparison of powers," *Biometrika,* 64,2, 231–246.

[12] Shapiro, S.S., and R.S.Francia, 1972, "Approximate analysis of variance test for normality," *Journal of the American Statistical Association,* 67, 215–216.

[13] Shapiro, S.S., and M.B. Wilk, 1965, "An analysis of variance test for normality (complete samples)," *Biometrika,* 52, 591–611.

[14] Shapiro, S.S., M.B. Wilk, and H.J.Chen, 1968, "A comparative study of various tests for normality," *Journal of the American Statistical Association,* 63, 1343–1372.

[15] Stuart, A. and J.K.Ord, 1987, *Kendall's Advanced theory of statistics, Volume 1 Distribution theory,* 5th edition, Charles Griffin, London.

Table 4: Power Curves #7 to #12

H_1 Distribution #7: Student t(7)								
λ	$\sqrt{b_1}$	b_2	K^2	GE	KS	AD	SW	CB
0	.047	.042	.044	.048	.047	.043	.048	.046
.25	.048	.046	.048	.046	.083	.081	.046	.049
.50	.096	.090	.105	.069	.193	.275	.070	.096
.75	.185	.194	.218	.170	.137	.152	.143	.198
1.0	.231	.238	.275	.209	.054	.055	.173	.238
H_1 Distribution #8: Johnson SB(1,2)								
0	.050	.051	.050	.046	.051	.051	.043	.051
.25	.052	.051	.051	.050	.082	.090	.054	.053
.50	.048	.051	.052	.046	.143	.223	.053	.054
.75	.070	.060	.057	.053	.084	.070	.101	.087
1.0	.085	.069	.069	.056	.071	.061	.120	.106
H_1 Distribution #9: Weibull								
0	.046	.047	.049	.049	.049	.052	.049	.048
.25	.046	.052	.053	.046	.084	.087	.047	.050
.50	.084	.059	.078	.046	.163	.237	.081	.082
.75	.283	.095	.202	.062	.116	.111	.279	.263
1.0	.372	.132	.271	.083	.115	.094	.450	.414
H_1 Distribution #10: Johnson SB(1,1)								
0	.051	.052	.047	.043	.053	.052	.045	.049
.25	.046	.053	.050	.052	.076	.077	.046	.051
.50	.090	.053	.068	.046	.165	.230	.091	.083
.75	.381	.076	.214	.074	.137	.134	.529	.480
1.0	.561	.094	.339	.095	.206	.176	.833	.786
H_1 Distribution #11: $\chi^2(4)$								
0	.045	.051	.051	.047	.052	.053	.044	.046
.25	.056	.062	.060	.053	.082	.080	.052	.058
.50	.260	.142	.218	.094	.212	.320	.193	.223
.75	.756	.347	.628	.199	.260	.323	.755	.743
1.0	.892	.417	.780	.207	.286	.282	.952	.943
H_1 Distribution #12: Exponential								
0	.045	.049	.048	.051	.052	.053	.047	.050
.25	.060	.057	.060	.051	.086	.085	.049	.060
.50	.381	.217	.340	.148	.289	.416	.301	.336
.75	.928	.558	.858	.351	.456	.617	.931	.929
1.0	.989	.651	.948	.363	.498	.661	.999	.999

Table 5: Power Curves #13 and #14

H_1 Distribution #13: Log-normal(1, 1)								
λ	$\sqrt{b_1}$	b_2	K^2	GE	KS	AD	SW	CB
0	.052	.046	.052	.047	.052	.052	.051	.051
.25	.092	.091	.101	.074	.096	.099	.074	.092
.50	.492	.392	.488	.306	.451	.679	.424	.472
.75	.957	.806	.929	.717	.930	.992	.942	.942
1.0	.999	.909	.995	.770	1.0	.999	1.0	1.0
H_1 Distribution #14: Extreme Value(0, 1)								
0	.041	.045	.040	.044	.051	.055	.042	.041
.25	.053	.055	.053	.052	.082	.082	.052	.054
.50	.202	.127	.183	.091	.202	.309	.154	.177
.75	.589	.273	.489	.160	.162	.231	.524	.537
1.0	.700	.331	.580	.180	.122	.157	.666	.662