5-1-2014

Exponentiality Test Using a Modified Lilliefors Test

Achut Prasad Adhikari
University of Northern Colorado

UNIVERSITY OF NORTHERN COLORADO

Greeley, Colorado

The Graduate School

AN EXPONENTIALITY TEST USING A MODIFIED LILLIEFORS TEST

A Dissertation Submitted in Partial Fulfillment
Of the Requirements for the Degree of
Doctor of Philosophy

Achut Prasad Adhikari

College of Education and Behavioral Sciences
Department of Applied Statistics and Research Methods

May 2014

ABSTRACT

Adhikari, Achut Prasad *An Exponentiality Test Using a Modified Lilliefors Test.* Published Doctor of Philosophy dissertation, University of Northern Colorado, 2014.

A new exponentiality test was developed by modifying the Lilliefors test of exponentiality for the purpose of improving the power of the test it directly modified. Lilliefors has considered the maximum absolute differences between the sample empirical distribution function (EDF) and the exponential cumulative distribution function (CDF). The proposed test considered the sum of all the absolute differences between the CDF and EDF. By considering the sum of all the absolute differences rather than only a point difference of each observation, the proposed test would expect to be less affected by individual extreme (too low or too high) observations and capable of detecting smaller, but consistent, differences between the distributions. The proposed test statistic is not only easy to understand but also very simple and easy to compute. The proposed test was compared directly to the Lilliefors test (LF-test), the Cramer-Von Mises test (CVM-test), Finkelstein & Schafers test (S-test) and the $\widetilde{D}_{\mathrm{n}}$ test (D-test).

The critical values were developed and the accuracy of the intended significance levels was verified for the proposed test. The results showed that all five tests of exponentiality worked very well in terms of controlling the intended significance levels. The proposed test performed very closely to the other four tests of exponentiality in terms of the accuracy of the intended significance levels across all considered sample sizes.

The proposed exponentiality test (PML-test) did successfully improve upon the power of the test it directly modified (i.e. LF-test). The actual method employed in the development of the test statistic in this study, achieved its primary goal of improving the power of the LF-test of exponentiality. This study showed that the proposed exponentiality test (PML-test) demonstrated consistently superior power over the S-test, LF-test, CVM-test, and D-test for most of the alternative distributions presented in this study. The D-test, CVM-test, and S-test exhibited similar power for a fixed sample size and significance level. The LF-test consistently showed the lowest power among five exponentiality tests. So, practically speaking the proposed test can hope to replace the other four exponentiality tests discussed throughout this study while maintaining a very simple form for computation and easy to understand for those people who have limited knowledge of statistics.

This study has shown that using the sum of all the absolute differences between the two functions (CDF and EDF) will have more power than just using the maximum differences between these two functions (like LF-test) or using the sum of squared differences between these two functions (like Cramer-Von Mises type test). The research presented here has the potential to modify many other tests and / or to develop tests for distributional assumption.

ACKNOWLEDGEMENTS

I would like to thank several people for their professional and personal assistance. I pay tribute to the Department of Applied Statistics and Research Methods (ASRM) of the University of Northern Colorado (UNC) for providing me opportunities to use its resources. My gratitude goes to Dr. Jay, my research advisor, who has helped me since I was admitted to this university. He made extremely valuable suggestions and has therefore contributed to enhancing the standard of this study. In this endeavor, Dr. Jay deserves my special thanks for his guidance and efforts to make it more useful to the users and bring this study to its final form. I am equally grateful to Dr. Lalonde, Dr. Shafie, and Dr. Gilliam whose thorough guidance strengthened my study.

I am also very grateful to my parents Dev Nath Upadhyaya Adhikari and Bhim Kumari Upadhyaya for their enthusiastic encouragements in every walk of my life. I would like to thank my beloved wife Sita Sharma-Adhikari, daughter Kyasurina Adhikari and son Kshitiz Adhikari for their valuable assistance; their attendance made me fresh in every walk of my movement. My educational journey at this point was not possible without the continuous support from my wife. She shared both my joy and sorrow, which I expect throughout my life. She fills my life with all the colors that I didn't know existed. She always brings a smile to my face and makes me forget all my problems. To my family, I am so sorry for the time spent away from you and look forward to a better future where we can share and enjoy throughout our life. I love you, and I dedicate this dissertation to you.

Although my first school is my home where I learned to speak the first word like mama, I started my formal education from Shree Prembasti Higher Secondary School, Bharatpur, Chitwan, Nepal. All of my previous teachers from that school inspired me for a better future. I can never forget my previous teachers whose guidance made me a good citizen, a good educator and a good researcher.

The shaping of my career goals were further enhanced by the Institute of Agriculture and Animal Science (IAAS), Rampur, Chitwan, Nepal of the Tribhuvan University (TU). I learned the experimental designs for the first time from this institute along with several course materials especially in the biological field. I am also indebted to this institute.

I cannot overlook the significant efforts from Dr. Steiner and Dr. Mundfrom of the New Mexico State University (NMSU) whose advice and guidance helped me to pursue the doctoral degree at the University of Northern Colorado. Dr. Steiner was my research advisor in my Master’s program at the New Mexico State University.

Finally, my wife’s parents, Narayan Dutta Upadhyaya Paudel and Subhadra Devi Paudel, also deserve my special thanks; their blessings are on my head and their love is in my heart.

I shall feel amply rewarded if this study proves helpful in the development of genuine research. I look forward to suggestions from all readers for the further improvements on the subject matter of this study.

TABLE OF CONTENTS

CHAPTER

LIST OF FIGURES

FIGURE

LIST OF TABLES

TABLE

CHAPTER I

INTRODUCTION

This chapter purveys the background information, the importance of assumptions in any statistical tests, some notations being used, the purpose and significance of the study, the research questions to be addressed, the limitations of the study, and some definitions necessary to understand the related theories behind this study.

Importance of Assumptions in Statistical Tests

Testing equality of the means is a very common task encountered by researchers and statistical consultants (Yan, 2009). Exponential distributions are quite often used in duration models and survival analysis, including several applications in macroeconomics, finance and labor economics (optimal insurance policy, duration of unemployment spell, retirement behavior, etc.). Quite often the data-generating process for estimating these types of models is assumed to behave as exponential. This calls for developing tests for distributional assumptions in order to avoid misspecification of the model (Acosta & Rojas, 2009).

The exponential distribution is often concerned with the amount of time until some specific event occurs. Also, the exponential distribution can be used to model situations where certain events occur with a constant probability per unit length (Thongteeraparp & Chodjuntug, 2011).

"The validity of estimates and tests of hypotheses for analyses derived from linear models rests on the merits of several key assumptions. The analysis of variance can lead to erroneous inferences if certain assumptions regarding the data are not satisfied" (Kuehl, 2000, p. 123).

As statistical consultants we should always consider the validity of the assumptions, be doubtful, and conduct analyses to examine the adequacy of the model. "Gross violations of the assumptions may yield an unstable model in the sense that different samples could lead to a totally different model with opposite conclusions" (Montgomery, Peck, & Vining, 2006, p. 122).

Background Information

This study developed a new test of exponentiality. In order to assess the exponentiality assumptions, several techniques have been developed in the field of statistics, ranging from descriptive statistics including plots to the inferential statistics.

The Chi-Square Goodness-of-Fit Test is quite general and can be applied for any distribution (Conover, 1999) but this test requires large sample size, and the formation of intervals for a continuous distribution is arbitrary (Agresti, 1996). The Kolmogorov-Smirnov Goodness-of-Fit Test appears to be more powerful than the Chi-Square Goodness-of-Fit Test for any sample size (Lilliefors, 1967). Seier (2002) found that the power of Anderson-Darling Goodness-of-Fit Test outperforms the Kolmogorov-Smirnov Goodness-of-Fit Test. Among the Shapiro-Wilk, Anderson-Darling, Lilliefors and Kolmogorov-Smirnov Goodness-of-Fit Tests, the Shapiro-Wilk Goodness-of-Fit Test is the most powerful (Razali & Wah, 2011).

Notations and Assumptions

Due to the frequent use and the lengthy names of some commonly used Goodness-of-Fit Tests which were compared in this study, it was chosen to use the full name of any test only for the first time in each chapter. This study used the short (abbreviated) form of these tests thereafter in each chapter (e.g. χ^2-test for Chi-Square Goodness-of-Fit Test, KS-test for Kolmogorov-Smirnov Goodness-of-Fit Test, AD-test for Anderson-Darling Goodness-of-Fit Test, SW-test for Shapiro-Wilk Goodness-of-Fit Test, LF-test for Lilliefors Test for exponentiality, PML-test for proposed modified Lilliefors exponentiality test, CVM-test for Cramer-Von Mises test of exponentiality, GOFT for Goodness-of-Fit Test, etc.).

Purpose of this Study

The purpose of this study was to develop a new Goodness-of-Fit Test (GOFT) of exponentiality and compare it with four other existing GOFTs in terms of computation and performance. The LF-test considered the supremum difference between the sample empirical distribution function (EDF) and the cumulative distribution function (CDF) of the exponential distribution (Lilliefors, 1969). The proposed test considered the sum of all the absolute differences between the EDF and the exponential CDF.

This study approximated the alpha levels by using the corresponding percentile of the ordered observed test statistics from the proposed test. By considering the sum of all the absolute differences rather than only a point difference of each observation, the proposed test was expected to be less affected by individual extreme (too low or too high) observations and capable of detecting smaller, but consistent, differences between the distributions. It is relevant to point out that the sample size and / or the outlier(s) can have a striking effect on the GOFT.

Significance of this Study

LF-test has proven to have low power among the commonly used exponentiality tests in many power studies (Schafer, Finkelstein & Collins, 1972, etc.).

Overholt & Schaffer (2013) proposed a modified Lilliefors normality test by using the sum of all absolute differences between the normal CDF and EDF. The authors compared their test with the AD-test, LF-test, and the SW-test in terms of significance levels and the power under ten different sample sizes and four different significance levels. Their study showed that the proposed test statistic had similar accuracy in regards to the significance levels when compared to the other three tests. The authors claimed that their test method showed some improvement in terms of power over the original Lilliefors test in their sets of parameters used in the study. They also argued that the increase in power was due to incorporating more information in their test statistic. This study will extend this idea for testing the exponentiality of the distribution.

Shaw (1994) introduced the test II statistic (horizontal distance test statistic) that uses the sum of all differences between two step functions for testing the null hypothesis that two randomly selected independent samples of equal size come from population having the same cumulative distribution function. Shaw showed that the power of the KS-test was found to be lower than the test II statistic.

Combining and extending the ideas of Overholt & Schaffer (2013), and Shaw (1994) in the context of exponentiality test constitutes a natural modification and / or extension of original LF-test in which this study used the sum of all the absolute differences between the EDF and exponential CDF as the test statistic for exponentiality test. Articles using the sum of all the absolute differences between the EDF and the exponential CDF as a test statistic are almost non-existent. It was expected prior to this investigation that an increase in power was resulted due to incorporating more information in the LF-test. It was also expected this proposed test would be less affected by extreme observation(s) because this test does not depend only on a single observation. If the proposed exponentiality test exhibits meaningful increases in power over the other four existing exponentiality tests, there would be a more powerful alternative available for researchers and a consulting statistician *may* be able to test for exponentiality using the proposed exponentiality test. Even if the proposed test demonstrates comparable power over the existing commonly used exponentiality GOFTs, the proposed test would be easier to understand for people who have a limited knowledge of statistics.

Research Questions to be Resolved

The following questions were addressed in this study:

Q1 How will the proposed test be designed to assure reliable critical values and their corresponding significance levels?

Q2 For specified significance levels, how will the proposed test perform in terms of detecting departures from exponentiality for data simulated from 12 alternative distributions?

Q3 For specified significance levels, how will the proposed test compare in terms of power with the four other exponentiality tests (Cramer-Von Mises test (CVM-test), Lilliefors test (LF-test), Finkelstein & Schafers statistics (S-test) and $\widetilde{D}_n$-test (D-test) as shown in 60, 61, 62, and 63 respectively?

Limitations of this Study

Power comparisons were examined for the proposed test and the other four exponentiality tests using 12 alternative distributions (11 right skewed and 1 symmetric distributions). Only three significance levels were examined in this study. It is possible that the findings of this study may be limited only to these sets of parameters.

Definitions

Power (1-β): Power of a test is the probability of rejecting null hypothesis when the null hypothesis is false.

Level of significance (α): α is the allowed maximum probability of rejecting null hypothesis when the null hypothesis is true.

Test statistic: It is the numerical value obtained from a statistical test. The test statistic summarizes how far that estimate falls from the parameter value in the null hypothesis.

p-value (*p*): *p* is the probability of getting a sample statistic or a more extreme sample statistic in the direction of the alternative hypothesis when the null hypothesis is true.

z-score: *z*-score represents the number of standard deviations that a data value falls above or below the mean.

Critical value (C.V.): This separates the critical region from the non-critical region.

Critical region or rejection region: It is the range of values of the test statistic that indicates that there is a significance difference and that the null hypothesis should be rejected.

Non-critical or non-rejection region: It is the range of values of the test statistic that indicates that the difference was probably due to chance and that the null hypothesis should not be rejected.

Random variable (R.V.): Variable whose values are determined by chance is called a random variable.

Null hypothesis (H0): Null hypothesis is a statistical hypothesis that states that there is no difference between a parameter and a specific value.

Alternative hypothesis (H1): Alternative hypothesis is a statistical hypothesis that states the existence of difference between a parameter and a specific value.

Right-tailed test: A one-tailed test which indicates that the null hypothesis should be rejected when the test statistic is in the critical region on the right side of the population parameter being tested.

Parametric methods: Any hypothesis test or confidence interval that is based on the assumption that the population distribution function is known, or known except for some unknown parameters, is called a parametric method.

Nonparametric methods: Any statistical methods which do not assume a particular population probability distribution, and are therefore valid for data from any population with any probability distribution, which can remain unknown.

Monte Carlo simulation: This is a broad class of computational algorithms that rely on repeated random sampling to obtain numerical results.

Goodness-of-Fit Test (GOFT): A test of conformity between an experimental results and theoretical expectations.

Normality assumption: It is the supposition that the underlying random variable of interest is distributed normally.

Exponentiality assumption: It is the supposition that the underlying random variable of interest is distributed exponentially.

Probability density function (PDF, $f(x)$): A function $f(x)$ is a PDF for some random variable X if and only if it satisfies the properties:

$$f(x) \geq 0 \tag{1}$$

for all real *x*, and

$$\int_{-\infty}^{\infty} f(x)dx = 1 \tag{2}$$

Cumulative distribution function (CDF, $F(x)$): The cumulative distribution function of a random variable X is defined for any real *x* by

$$F(x) = p[X \leq x] \tag{3}$$

Empirical distribution function (EDF): The cumulative distribution of the observed data values is called the empirical distribution function.

Supremum of x (SUPx): For SUPx(f(*x*)); the supremum is the smallest value of x within f(x), which is greater than or equal to all other values of x in f(*x*).

Infimum of x (INFx): For INFx(f(*x*)); the infimum is the largest value of x within f(x), which is less than or equal to all other values of x in f(*x*).

Unbiased estimator: An estimator T is said to be an unbiased estimator of $\tau(\theta)$ if $E(T) = \tau(\theta)$, for all $\theta \in \Omega$, otherwise we say T is a biased estimator.

Asymptotic relative efficiency (ARE): Let n_1 and n_2 be the sample sizes required for two tests T_1 and T_2 to have the same power under the same level of significance. If α (probability of type I error) and β (probability of type II error) remain fixed, the limit of n_2/n_1 as $n_1 \rightarrow \infty$ is called the ARE of the first test to the second test if that limit is independent of α and β.

Uniformly minimum variance unbiased estimator (UMVUE): Let $X_1, X_2, \ldots, X_n$ be a random sample of size *n* from $f(X ; \theta)$, an estimator T^* of $\tau(\theta)$ is called UMVUE of $\tau(\theta)$ if following holds:

i. T^* is an unbiased estimator for $\tau(\theta)$ i.e. $E(T^*) = \tau(\theta)$

ii. For any other unbiased estimator T of $\tau(\theta)$, variance(T^*) $\leq$ Variance(T) for all $\theta \in \Omega$

Parameter: A parameter can be defined as the numerical summary of a population.

Skewness: Skewness is the third moment around the mean and characterizes whether the distribution is symmetric (skewness = 0).

Kurtosis: Kurtosis is a function of the fourth central moment and characterizes peakedness, where the normal distribution has a value of three and small values corresponding to thinner tails (less peakedness).

CHAPTER II

LITERATURE REVIEW

This chapter provides a synopsis of the theories necessary to understand this research project. The summary and synthesis of a variety of nonparametric Goodness-of-fit-tests (GOFT) are presented, explained and compared.

Categories of Goodness of Fit Tests

Dufour, Farhat, & Gardiol (1998) reported 40 different tests that can be used to test normality. These tests were grouped into three categories: empirical distribution function (EDF) tests which pertain to the location-scale model; the skewness and kurtosis-based moment tests; and correlation tests which are based on the ratio of two estimates of scale obtained from order statistics.

Arshad, Rasol, & Ahmad (2003) evaluated the Anderson Darling and modified Anderson Darling test statistics for testing the goodness of fit. They modified the completely specified generalized Pareto distribution by using their probability weighted moment estimates. The authors divided the goodness of fit techniques into four categories: tests of Chi-Square types, moment ratio techniques, tests based on correlation, and tests based on empirical distribution function (EDF). Researchers argued that the test of Chi-Square type have less power due to loss of information caused by grouping. Similarly, use of EDF tests has been difficult due to lack of readily available tables of significance point for the case where the parameters of the assumed distribution have to be estimated from the sample data.

Seier (2002) categorized the GOFTs into four groups: tests based on skewness and kurtosis, EDF tests, regression and correlation tests, other tests of normality (e.g. empirical characteristic function based tests, U-statistics based tests, etc.).

Oztuna, Elhan, & Tuccar (2006) divided the GOFTs into two broad groups: Graphical methods (e.g. Histogram, Stem and Leaf Plot, Boxplot, Normal Quantile Quantile plot, Normal Probability Plot, etc.) and tests methods (e.g. Kolmogorov-Smirnov test, Lilliefors corrected Kolmogorov-Smirnov test, Shapiro-Wilk test, D' Agostino-Pearson omnibus test, Jarqua-Bera test, etc.).

Most of the GOFTs for testing exponentiality of the distribution are based on the GOFTs originally developed to test normality in the nineteenth century. In order to better understand the GOFTs developed for testing exponentiality of the distribution, it is relevant to review the GOFTs developed for testing normality as well. An explanation of the different types of GOFTs follows.

The Chi-Squared Goodness-of-Fit Test

According to Conover (1999) the oldest and best-known Goodness-of-Fit Test is the Chi-Squared Goodness-of-Fit Test (χ^2-test), first presented by Pearson (1900). The test assumes that the sample is a random sample whose measurement scale is at least nominal. Pearson wanted to test the hypotheses:

H0 P (X is in class j) = p_j^* for j = 1, … , c (i.e. the sample has been drawn from a population that follows a specified distribution)

H1 P (X is in class j) $\neq p_j^*$ for at least one class (i.e. the sample has not been drawn from a population that follows the specified distribution)

The data consist of *N* independent observations of a random variable X. The *N* observations are grouped into c classes (in most of the cases these c categories are the natural classes or defined by the researcher), and the number of observations in each class are presented in the form of a 1 X c contingency table as shown in Figure 1.

	Class				
	1	2	…	c	Total
Observed Frequencies	O_1	O_2	…	O_c	*N*

Figure 1. Observations in a Contingency Table

Let O_j denotes the number of observations in class j, for j = 1, 2, … , c and p_j^* be the probability of a random observation of X being in class j, assuming that the null hypothesis is true. The expected number of observations in class j is denoted by E_j assuming the null hypothesis is true is defined as:

$$E_j = p_j * N, \qquad j = 1, 2, \dots, c \tag{4}$$

The test statistic, $\chi 2$, is then given by:

$$\chi 2 = \sum_{j=1}^{c} \frac{(O_j - E_j)^2}{E_j} \tag{5}$$

To find the critical value, it is necessary to know the null distribution of the test statistic. However, the exact distribution of χ^2 is difficult to find. It can be approximated with the Chi-Squared distribution with c-1 degrees of freedom. The critical values can be found in many nonparametric statistics books. The null hypothesis is rejected if the test statistic, χ^2, is greater than the 1-α quantile from the Chi-Squared distribution with c-1 degrees of freedom. This test will always be a right-tailed test.

The χ^2-test was designed for nominal data. However, it can also be used in continuous data:

> The Chi-Squared Goodness of Fit Test is not limited to discrete random variables. It can also be used to test whether the data come from a specified continuous distribution, where some of the unknown parameters may be estimated from the data. The first step is to "discretize" the continuous random variable by forming intervals, which then become the classes described in the test. The number of observations in each interval O_j is compared with the expected number in each interval
>
> $$E_j = N * P \text{ (X is in interval j)} \quad (6)$$
>
> when the null hypothesis is true. (Conover, 1999, pp 245)

If some of the E_j's are small, the χ^2-test may not be accurate. Several studies have examined the E_j's and the Chi-Squared approximation and suggested the minimum values for the expected counts in each cell. Cochran (1952) suggested that none of the E_j's should be less than 1 and no more than 20 % of the cells should be smaller than 5. Yarnold (1970) proposed that if the number of classes under consideration, s, is 3 or more, and if r denotes the number of expectations less than 5, then the minimum expectation may be as small as 5r/s. Koehler and Larntz (1980) argued that "for the null hypothesis of symmetry, the chi- squared approximation for the Pearson statistic is quite adequate at the 0.05 and 0.01 nominal levels for expected frequencies as low as 0.25 when k (number of categories) > 3, n > 10, $n^2/k > 10$" (p. 343). From these discussions, it can be inferred that the researchers could combine some of the cells if many of the E_j's are small.

Liu (2012) conducted Monte Carlo simulations to investigate what sample sizes are required to obtain the desired power for the χ^2-test. The author listed the sample sizes and power of the test under different non-central Chi-Squared distributions.

Agresti (1996) pointed out some limitations of the χ^2-test. The test requires large samples. If the data were given in raw form and intervals for the classes had to be determined, the formation of these intervals is somewhat arbitrary and therefore a weakness in applying the χ^2-test to any continuous distribution.

The primary advantage of the χ^2-test is that it is quite general. It can be applied for any distribution, either discrete or continuous, for which the cumulative distribution function (CDF) can be computed.

Empirical Distribution Function Tests of Normality

> The idea of the empirical distribution function (EDF) tests in testing normality of data is to compare the EDF 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. The most popular EDF tests are the ones developed by Kolmogorov–Smirnov, Cramer–von Mises, and Anderson–Darling. (Yap & Sim, 2011, p. 5)

The function $F(x)$ often is referred to simply as the distribution function of X, and the subscripted notation,$F_X(x)$, sometimes is used. The EDF is the observed CDF of the data denoted by S_X. Bain & Engelhardt (1992) argued that “the EDF tests generally are considered to be more powerful than the χ^2-test, because they make more direct use of the individual observations. Of course, then they are not applicable if the data are available only as grouped data” (p. 457).

The first known EDF Goodness of Fit Test was introduced by Kolmogorov (1933). The Kolmogorov-Smirnov test was first proposed by Kolmogorov (1933) and then developed by Smirnov (1939) (Mendes & Pala, 2003). The Kolmogorov-Smirnov

GOFT (referred to as KS-test here forth) belongs to the supremum class of EDF statistic and this class of statistics is based on the largest vertical difference between the hypothesized and empirical distribution (Conover, 1999). Conover (1999) and Yap & Sim (2011) presented the precise description of this test. Unlike the χ^2-test, the Kolmogorov Goodness of Fit Tests (KS-test) was designed for ordinal data. The KS-test statistic enables the readers to form a confidence band. The KS-test assumes that the sample is a random sample and the data consist of observations $X_1, X_2, \ldots, X_n$ of sample size, *n*, associated with some unknown distribution function, denoted by $F(x)$. The test statistics is based on the largest vertical difference between the hypothesized and empirical distribution which actually measures the discrepancy between the empirical distribution function (S_X) and the hypothesized distribution function ($F^*(x)$). This test requires that the null distribution ($F^*(x)$) be completely specified with known parameters. In the KS-test, $F^*(x)$ is taken from a normal distribution with known parameters mean, μ, and standard deviation, σ. Depending upon the researchers' interest, the hypothesis could be one-sided or two sided. The hypotheses and test statistic for KS-test is also defined differently for three different types of hypotheses.

Two sided Test

H0 $F(x) = F^*(x)$ for all x from $-\infty$ to $+\infty$

H1 $F(x) \neq F^*(x)$ for at least one value of x

The test statistic, *D*, be the greatest absolute vertical distance between and $F^*(x)$ and $S(x)$ is given by:

$$D = \sup_{x} |F^*(x) - S(x)| \quad (7)$$

Schoder, Himmelmann, & Sim. (2006) ran Monte Carlo simulations to assess the performance of the KS-test, depending on sample size and severity of violations of normality. This test performs badly (cannot detect non-normal) on data with single outliers, 10 % outliers and skewed data at sample sizes < 100, whereas normality was rejected to an acceptable degree for likert-type data. From this study, it can be inferred that the KS-test with the Lilliefors correction cannot be recommended as a tool to identify reliably deviations from normality. Similar results were obtained by Yap & Sim (2011). They studied and compared the power of eight selected normality tests. Results showed that the KS-test performed poor in terms of power. Seier (2002) also reported that the KS-test relatively has lower power as compared to other GOFTs.

Lilliefors (1967) demonstrated that KS-test can be used with small sample sizes (at least four) where the validity of the χ^2-test would be questionable. Lilliefors further explained that the KS-test appears to be a more powerful test than the χ^2-test for any sample size. Mendes & Pala (2003) studied the Shapiro-Wilk, Lilliefors, and Kolmogorov-Smirnov GOFTs under various sample sizes and distributions. The study found that the Shapiro-Wilk test gave the most powerful results followed by the Lilliefors test. The KS-test results were the weakest in power among all three tests.

In implementing the KS-test, most statistical software packages use the sample mean and sample variance as the parameters of the normal distribution. However, the sample mean and sample variance do not necessarily provide the closest fit to the empirical distribution of the data. Drezner, Tuyrel, & Zerom (2009) proposed a modified KS-test in which they optimally choose the mean and variance of the normal distribution by minimizing the KS statistics. Drezner et al. demonstrated that the power of the proposed test indicated that the test is able to discriminate between the normal distribution and distributions such as uniform, bi-modal, beta, exponential and log-normal which are different in shape, but has a relatively lower power against the student's *t*-distribution that is similar in shape to the normal distribution.

Breton, Devore, & Brown (2008) estimated the power of a test for normality for any mean, variance, skewness, and kurtosis. They suggested that if samples are of size less than 20, the KS-test can be expected to yield greater power than the χ^2-test; otherwise the χ^2-test is preferred. The χ^2-test is to be generally preferred over KS-test if the sample size is between $n = 18$ and $n = 330$.

Although the KS-test is originally designed to handle continuous data, it can also be applied with non-continuous distribution. Conover (1972) derived a method for finding the exact critical level for the KS-test for all completely specified distribution functions, whether continuous or non-continuous. Pettitt & Stephens (1977) proposed a modified KS-test that can handle the discrete and grouped data. They found identical power between the modified KS-test and the χ^2-test.

Razali & Wah (2011) explored the four most commonly used GOFTs of normality for the purpose of comparing power. Among Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors, and Anderson-Darling tests; the KS-test yield the least power while the Shapiro-Wilk test yielded the most power for all types of distribution and sample sizes under the study.

From these arguments, it can be concluded that the KS-test may be preferred over the χ^2-test for small samples. When certain parameters of the distribution must be estimated from the sample, the KS-test no longer should be employed at least not using the commonly tabulated critical points. Among the most common GOFTs of normality (e.g. Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors, and Anderson-Darling tests), the KS-test has the smallest power. Since, the KS-test was the first EDF type GOFT, many subsequent tests (including the tests for exponentiality) are based on some form of the modification of this test. So, this test is still a valuable resource in the foundation of many GOFTs.

To test the hypothesis that the sample has been drawn from a population with a completely specified density function, Anderson and Darling (1954) proposed a distribution-free Goodness-of-Fit Test. This procedure may also be used if one wishes to reject the hypothesis whenever the true distribution differs materially from the hypothetical and especially when it differs in the tails. The Anderson-Darling test (AD-test) is a modification of the Cramer-Von Mises test (CVM-test). It differs from the CVM- test in such a way that it gives more weight to the tails of the distribution (Farrel & Stewart, 2006).

Denote the specified cumulative distribution function by $F(x)$ and the empirical cumulative distribution function by $F_n(x)$. The AD-test statistic, W_n^2, belongs to the quadratic class of EDF statistic which is an average of the squared discrepancy, $[F_n(x) - F(x)]^2$, weighted by $\Psi[F(x)]$ and the increase in $F(x)$ (and the normalized *n*).

$$W_n^2 = n \int_{-\infty}^{\infty} [F_n(x) - F(x)]^2 \psi\big(F(x)\big) dF(x) \tag{8}$$

The weight function is some non-negative function as shown in 9.

$$\psi\big(F(x)\big) = [F(x)(1 - F(x))]^{-1} \tag{9}$$

Substituting equation 9 into equation 10 produces

$$W_n^2 = n\int_{-\infty}^{\infty}[F_n(x) - F(x)]^2[F(x)(1 - F(x))]^{-1}dF(x) \quad (10)$$

In order to make the computation of this statistic easier, the following formula can be applied (Arshad, Rasol, & Ahmad, 2003),

$$W_n^2 = -n - \frac{1}{n}\Sigma(2i - 1)\{logF^*(x_i) + \log(1 - F^*(x_{n+1-1})\} \quad (11)$$

where, $F^*(x_i)$ is the cumulative distribution function of the specified distribution and x_i's are the ordered data. The hypothesis is to be rejected if test statistic, W_n^2, is sufficiently large:

> The weight function $\Psi(F(x))$, $0\leq F(x)\leq 1$, is to be chosen by the statistician so as to weight the deviations according to the importance attached to various portions of the distribution function. This choice depends on the power against the alternative distributions considered most important. (Anderson and Darling, 1952, pp. 194)

Seier (2002) pointed out that the tests based on skewness and kurtosis tends to have lower power than AD-test for skewed distributions when the kurtosis is low. Razali & Wah (2011) assessed both symmetric and asymmetric distributions for the purpose of comparing the power of four formal tests of normality: Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors, and Anderson-Darling. The study found that although the AD-test does not outperform the power of Shapiro-Wilk GOFT (SW-test), it outperforms the power of the other two GOFTs. Researchers also found that for sample size ≤ 30, the power at the 5 % significance level for all four tests was low (less than 40 %).

Instead of applying the Monte Carlo simulations, Henderson (2006) used four experimentally-derived data sets representing normal, positive kurtotic, positively skewed and negatively skewed distributions to testing experimental data for univariate normality. The study found that Anderson-Darling, Shapiro-Wilk, Shapiro-Francia, and Filliben tests correctly classified all four test samples. The author further explained:

> It is not easy to draw firm conclusions from the foregoing regarding the best test for normality. In general, however, the Anderson–Darling, Shapiro–Wilk, and Shapiro–Francia tests appear to be the most frequently favored tests. Certainly these three tests perform well when used on the four test samples of the type commonly encountered in clinical chemistry when studying experimentally-derived results. (Henderson, 2006, pp. 128)

Normality tests are not only used to determine whether a data set is well-modeled by a normal distribution or to compute how likely an underlying random variable is to be normally distributed, but also for evaluating the performance of the normality tests to ensure the validity of the *t*-statistic used for assessing significance of regressors in a regression model. Islam (2011) explored 40 distributions and found that Anderson-Darling statistic is the best option among the five normality tests: Jarque-Bera, Shapiro-Wilk, D’Agostino & Pearson, Anderson-Darling, and Lilliefors GOFT.

Stephens (1974) argued that:

> Even if a new statistic is proposed and claimed to have advantages only for a certain type of alternative (say very skewed, or long-tailed), for a real comparison with statistics of the Shapiro-Wilk or Anderson-Darling type, we need to see how the new statistic fares when used on other alternatives also. (p. 6)

From the above arguments, it can be concluded that the AD-test required the density function be completely specified. The power of this test outperforms the Kolmogorov-Smirnov, Lilliefors and Skewness-Kurtosis based GOFTs of normality.

Evans, Drew, & Leemis (2008) presented mathematical derivations of the distributions of the Kolmogorov-Smirnov, Cramer-Von Mises, and Anderson-Darling test statistics in the case of exponential distribution when the parameters are unknown and estimated from sample data for small sample sizes via maximum likelihood. These derivations can help the readers to understand how the maximum likelihood estimators can be used to derive the distributions of these test statistics.

A modification of the KS-test was proposed by Lilliefors (1967). The test compares the empirical distribution of X with a normal distribution where its unknown μ and σ are estimated from the given sample data. This test is suitable when the unknown parameters of the null distribution must be estimated from the sample data. The only difference between Lilliefors and KS-test statistic is that the CDF, $F^*(x)$, is obtained from the normalized sample (Z_i) while CDF, $F^*(x)$, in the KS-test used the original X_i values. The test assumes the sample is a random sample. The hypotheses of interest are:

H0 The data comes from a normal distribution with unknown mean and unknown standard deviation

H1 The distribution function of the X_i's is non-normal

The test statistic, *D*, is obtained by

$$D = \underset{x}{\text{Sup}}\ |F^*(x) - S(x)|, \qquad (12)$$

where $S(x)$ is the sample cumulative distribution function and $F^*(x)$ is the cumulative normal distribution function with $\mu = \bar{X}$, the sample mean, and $\sigma^2 = S^2$, the sample variance, defined with denominator *n*-1. The test rejects the null hypothesis that the observations are from a normal distribution, if the test statistic, *D*, exceeds the critical value.

The exact quantiles, and the exact mathematical form of null distribution, are unknown. The null distribution has been obtained approximately, by generating thousands of pseudo-random numbers on a computer, and estimating quantiles from the empirical distribution function of the thousands of subsequent values of the test statistic (Conover, 1999). Lilliefors used 1000 random samples of various sample sizes to approximate the distribution of the test statistic, *D*. The ordered $(1 - \alpha)^{th}$ percentile was used to find the critical values for the selected sample size and the desired significance level.

Lilliefors also compared the power of this test with the χ^2-test in several non-normal distributions and found this test to be more powerful in the situations reported. Yap & Sim (2011) compared the power of eight selected normality tests of sample data generated from several distributions. The study showed that Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests did not outperform the SW-test.

Razali & Wah (2011) investigated the power of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors, and AD-test. Study found that Lilliefors test outperforms the KS-test only among the four GOFTs. They also argued that even though the Lilliefors statistic is same as the Kolmogorov-Smirnov statistic, the table for critical values is different which leads to a different conclusion about the normality of data.

Dallal & Wilkinson (1986) claimed that there exist some difficulties finding an analytic approximation to Lilliefors' table. They attempted to duplicate Lilliefors simulation. In order to find the corrected table, they used SYSTAN's NONLIN procedure. The authors argued that their proposed table corrects the critical values for testing normality originally proposed by Lilliefors (1967).

To test the hypothesis that a set of data arises from a normal distribution with unknown mean and variance, Scott & Stewart (2011) presented a modified version of the one-sample Cramer-Von Mises test (CVM-test). The test statistic takes the form:

$$W^2(n) = \frac{1}{12n} + \sum_{i=1}^{n} \left[t_i - \left(\frac{i-0.5}{n}\right)\right]^2 \tag{13}$$

where, $t_i = \Phi\left(\frac{X_i - \bar{X}}{S}\right)$ for i = 1, 2, …, n. Authors demonstrated that their test was superior than the Lilliefors test in terms of power.

The Cramer-Von Mises two-sample normality test (CVM-test) is one of the best-known distribution free GOFT. This test was first introduced by Cramer (1928) and Von Mises (1931) (as cited in Xiao & Gordon, 2007). Conover (1999) gives a thorough and precise description of this test. The test assumes that the samples are random and their measurement scale is at least ordinal. In general, the random variables are assumed to be continuous. If they are discrete, then the test is likely to be conservative. Assuming there are two random samples: $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_m$ with unknown distribution functions $F(x)$ and $G(x)$, the hypotheses of interest are:

H0 $F(x) = G(x)$ for all x from $-\infty$ to $+\infty$

H1 $F(x) \neq G(x)$ for at least one value of x

Consider $S_1(x)$ and $S_2(x)$ be the empirical distribution functions of the two samples. The Cramer-Von Mises test statistic adds up the squared differences between the cumulative distribution function being compared as given in equation 14,

$$T_2 = \frac{mn}{(m+n)^2}\Sigma[S_1(x) - S_2(x)]^2, \tag{14}$$

where, m and n are the sample sizes of the first and second populations respectively.

The exact distribution of the test statistic, T_2, is found by considering all orderings of Xs and Ys to be equally likely under the null hypothesis. Quantiles for T_2 using the asymptotic distribution when $n \to \infty$ and $m \to \infty$ are given in many nonparametric statistics books. The test rejects the null hypothesis if the test statistic exceeds the $(1 - \alpha)$ quantiles of $W_{1-\alpha}$.

Sprent (1989) argued that adding the squared differences between the cumulative distribution functions being compared makes the CVM-test often more powerful than the Kolmogorov-Smirnov test.

Shaw (1994) investigated the use of horizontal distances between the two sample step functions to develop a nonparametric rank test for testing the null hypothesis that two randomly selected independent samples of equal size come from populations having the same cumulative distribution functions. The author evaluated three horizontal distance test statistics for this purpose as shown in 15, 16, and 17.

$$\text{Test I: } T = \text{Sup}|R(x) - R(y)|, \tag{15}$$

$$\text{Test II: } T = \sum|R(x) - R(y)|, \tag{16}$$

$$\text{Test III: } T = \sum[R(x) - R(y)]^2, \tag{17}$$

where R(x) and R(y) are the step functions for each of the two independent samples. All three tests were two tailed at the $\alpha = 0.05$ level. The test rejects the null hypothesis at the level of significance α if the test statistic exceeds the $1 - \alpha$ quantile. For most of the distributions under study, the power of Test II and Test III were identical but both of them outperformed the Test I statistic. The power of the Smirnov test was found to be lower than both the Test II and Test III statistic. The author explained that the test II statistic is easier to calculate then test III statistic.

Overholt & Schaffer (2013) proposed a modified Lilliefors normality test by using the sum of all the absolute differences between the normal CDF and EDF. They compared their test with the AD-test, LF-test, and the SW-test in terms of significance levels and the power under ten different sample sizes and four different significance levels. Their study showed that their test statistic had similar accuracy in regards to the significance levels when compared to other three tests. The authors claimed that their test method showed improvement in terms of power over the original Lilliefors test in their sets of parameters used in the study. They argued that the increase in power was due to incorporating more information in their test statistic.

Correlation Tests of Normality

Correlation tests are based on the ratio of two estimates of scale obtained from order statistics: a weighted least-squares estimate given that the population is normally distributed and the unbiased estimate of scale for any population, i.e. the sample variance. Correlation tests focus on the slope of the line when the order statistics of the sample are confronted with their expected value under normality and these tests focused on the strength of the linear relationship (Seier, 2002).

The most well-known of the correlation based GOFTs is defined by Shapiro & Wilk (1965), originally restricted for $n \leq 50$. The test considers that the data consist of a random sample $X_1, X_2, \ldots, X_n$ of size n whose distribution function, $F(x)$, is unknown. The hypotheses of interest are

H0 $F(x)$ is a normal distribution function with unspecified mean and variance

H1 $F(x)$ is non-normal

The denominator, D, of the test statistic is calculated such that

$$D = \sum_{i=1}^{n} (x_i - \bar{x})^2, \tag{18}$$

where $\bar{x}$ is the sample mean.

Ordering the sample from smallest to largest produces $X^{(1)} \leq X^{(2)} \leq \ldots \leq X^{(n)}$. Also denote the best linearly unbiased coefficients of the X_i's by a_i (where a_i represents what the order statistics would look like if population is normal). The Shapiro-Wilk statistic (Shapiro &Wilk, 1965), W, is given by:

$$W = \frac{1}{D}\left[\sum_{i=1}^{k} a_i\left(X^{(n-i+1)} - X^{(i)}\right)\right]^2. \quad (19)$$

The test statistic, *W*, is basically the square root of Pearson's correlation between the ordered statistics $X^{(i)}$ and the coefficients a_i. This test statistic is scale and origin invariant.

If the test statistic is close to 1, the sample behaves like a normal distribution. On the other hand, if the test statistic is too small (i.e. too far below 1), the sample looks non-normal. The test rejects the null hypothesis at the level of significance α if the test statistic is less than the α^{th} quantile. These quantile values can be found in many nonparametric statistics books.

As mentioned earlier, although the SW-test is originally restricted for $n \leq 50$, D' Agostino (1971) presented a test that may be used for *n* greater than 50. Similarly, Shapiro & Francia (1972) suggested an approximate test for *n* greater than 50.

A problem common to most of the GOFTs is sensitivity to the presence of outliers in the sample. In fact a single such observation can lead to rejecting the null hypothesis even if the majority of the data are drawn from a normal distribution (Coin, 2008). The author showed a possible extension of SW-test that is not as much affected by outlier(s). The author claimed that the proposed test was able to determine whether the majority of the data is normally distributed, and moreover, it presents an optimal capacity of outlier detection. The study concluded that the Shapiro-Wilk (when feasible) and the Shapiro-Francia approximates are among the most powerful GOFTs against practically all alternatives.

Shapiro, Wilk, & Chen (1968) studied nine statistical procedures for evaluating normality. Their study concluded that the Shapiro-Wilk statistic provides a generally superior omnibus measure of non-normality. Stephens (1974) suggested that “even if a new statistic is proposed and claimed to have advantages only for a certain type of alternative, for a real comparison with the statistics of the Shapiro-Wilk or Anderson-Darling type, we need to see how the new statistic fares when used on other alternatives” (p. 8).

Oztuna, Elhan, & Tuccar (2006) compared four GOFTs of normality to investigate the type I error rates and power of the tests. The authors found that for small sample sizes, the SW-test outperforms the power of Lilliefors corrected Kolmogorov-Smirnov, D’ Agostino Pearson, and Jarqua-Bera tests. Yap & Sim (2011) studied and compared the power of eight selected normality tests. Results of this study indicated that SW-test has the best power properties over a wide range of asymmetric distributions. Mendes & Pala (2003) compared the Shapiro-Wilk, Lilliefors, and Kolmogorov-Smirnov tests for type I error and for power. Their study found that for all different sample sizes and distributions, Shapiro-Wilk test gave the most powerful results followed by Lilliefors test. Kolmogorov-Smirnov test results were the weakest among the three tests. Razali & Wah (2011) also found that the Shapiro-Wilk test is the most powerful normality test, followed by Anderson-Darling test, Lilliefors test, and Kolmogorov-Smirnov test among the four GOFTs.

From the above discussions, it can be concluded that in recent years, the SW-test has become the preferred test of normality because of its good power properties as compared to a wide range of alternative tests. This test is very simple to compute once the table of linear coefficients is available, and the test is quite sensitive against a wide range of alternatives even for small samples ($n < 20$). A drawback of the Shapiro-Wilk test is that for large sample sizes, it may prove awkward to tabulate or approximate the necessary values of the multipliers in the numerator of the statistic. Also, it may be difficult for large sample sizes to determine percentage points of its distribution. Practically speaking, another weakness of the SW-test is the difficulty some researchers have in understanding exactly what the test does.

Descriptive Methods of Normality Tests

Moment tests derive from the recognition that the third and fourth moments of the standard normal distribution are equal to 0 and 3, respectively. Hence, deviations from normality may be assessed using the sample moments i.e. the coefficients of skewness and kurtosis (Dufour, Farhat, & Gardiol, 1998).

The simplest and perhaps the oldest graphical display for one-dimensional data is the histogram, which divides the range of the data into bins and plots bars corresponding to each bin. The height of each bar reflects the number of data points in the corresponding bin (Oztuna et al., 2006). The histogram graphically summarizes the distribution of a data set such as the center of the data, spread of the data, skewness of the data, presence of outliers, and presence of multiple modes in the data. Unfortunately, the manner in which histograms depict the distribution of the data is somewhat arbitrary, depending heavily on the choice of bins and bin widths.

A stem-and-leaf plot is a variant on histograms that combines the features of a graphic and a table in that the original data values are explicitly shown in the display as a “stem” and a “leaf” for each value. The stem determines a set of bins into which leaves are sorted, and the resulting list of leaves for each stem resembles a bar in a histogram (Oztuna et al., 2006). Turned on its side, it has the same shape as the histogram. In fact, it shows each observation, and displays information that is lost in a histogram. Stem and leaf plots are useful for quick portrayals of small data sets. Similar conclusions can be drawn from stem-and-leaf plot and the histogram about the shape of the distribution (Bluman, 2012).

A boxplot provides an excellent visual summary of many important aspects of a distribution. According to Bluman (2012), Tukey developed the boxplot display, based on the five-number summary (minimum, first quartile, median, third quartile, and maximum) of the data. Suspected outliers appear in a boxplot as individual points O or an asterisk outside the box. If these appear on both sides of the box, they suggest the possibility of a heavy-tailed distribution. If they appear on only one side, they suggest the possibility of a skewed distribution.

The normal Q-Q plot may be the single most valuable graphical aid in diagnosing how a population distribution appears to differ from a normal distribution. Normal Q-Q plots plot the quantiles of a variable's distribution against the quantiles of the normal distribution. For values sampled from a normal distribution, the normal Q-Q plot has the points all lying on or near the straight line drawn through the middle half of the points. Scattered points lying away from the line are suspected outliers that may cause the sample to fail a normality test (Oztuna et al., 2006).

The normal probability plot (P-P plot) graphs observed cumulative probabilities of occurrence of the standardized residuals on the Y axis and of expected normal probabilities of occurrence on the X axis, such that a 45-degree line will appear when the observed conforms to the normally expected (Oztuna et al., 2006).

Comparing Different Goodness of Fit Tests

Many GOFTs for normality and exponentiality are available in literature but they have different performances in different situations. Most of the researches presented in this chapter show that the criteria to compare different GOFTs are mostly based on the power, type I error rates, and the simplicity of their computation for general use. This study has presented the comparative powers and some prominent features of the commonly used GOFTs on above discussions.

Finding the correlations among the different GOFTs is also an interesting field for many researchers. Stephens (1974) investigated the correlation among various GOFT statistics. The study found fairly strong correlations between the various test statistics leading to similar conclusions for hypotheses testing.

Although simple descriptive statistics can provide some information relevant to the GOFT, more precise information can be obtained by performing one of the GOFTs of exponentiality to determine whether the sample comes from a exponentially distributed population.

Graphical displays try to answer the question of how the data are distributed by showing what the data distribution "looks like", but they do not focus on the issue of how the data distribution compares with some theoretical distributions:

> An analyst often concludes that the distribution of the data 'is normal' or 'not normal' based on the graphical exploration (*Q–Q* plot, histogram or box plot) and formal test of normality. Even though graphical methods are useful in checking the normality of a sample data, they are unable to provide formal conclusive evidence that the normal assumption holds. The graphical method is subjective as what seems like a 'normal distribution' to one may not necessarily be so to others. In addition, vast experience and good statistical knowledge are required to interpret the graph properly. Therefore, in most cases, formal statistical tests are required to confirm the conclusion from graphical methods. (Yap & Sim, 2011, p. 2)

Exponentiality Tests

In order to test the null hypothesis that the random variable X has an exponential distribution, Dahiya & Gurland (1972) presented a GOFT for the exponential distribution. They used the generalized minimum χ^2 estimators to develop a test. The test statistic, $\hat{Q}$, takes the form as shown in equation 20,

$$\hat{Q} = nh'\hat{A}h, \tag{20}$$

where $\hat{A} = \hat{\Sigma}^{-1}(I - \hat{R})$, $\hat{R} = W\left(W'\hat{\Sigma}W\right)^{-1}W'\hat{\Sigma}^{-1}$, $\hat{\Sigma}$ is an estimator of covariance matrix, and W is a matrix of known constants. The authors claimed that the power of $\hat{Q}$ test of fit for the exponential distribution is invariant with respect to the scale parameter of the alternative distribution. Although they claimed that the test is highly efficient to detect the departure from the exponential distribution, this test could be difficult to compute and difficult to understand for those who have a limited knowledge of statistics.

Statistical inference under progressive censoring has received the attention of many authors. In many life tests, it is common practice to cease testing before all units have failed. In singly censored samples, *n* units are placed on a test and as each failure occurs, the time is noted. Finally, at some pre-determined time or after a pre-determined number of failures, the test is terminated. Data obtained from such experiments are called censored data (Wang, 2008). Wang developed a test statistic to test whether the progressively type-II censored samples come from an exponential distribution. If the lifetime distribution is exponential, S_1, S_2, … , S_m are all independent and identically distributed as exponential with scale parameter θ, then the test statistic proposed by Wang takes the form,

$$\chi^2 = 2\sum_{i=1}^{m-1} log\frac{(S_1+S_2+S_3+\cdots+S_m)}{(S_1+S_2+S_3+\cdots+S_i)}, \tag{21}$$

where, the numerator and denominators are censoring scheme and remaining surviving units respectively. Author compared this statistic with the statistic proposed by Balakrishnan & Lin (2002) which takes the form,

$$T = \frac{1}{m-1}\sum_{i=1}^{m-1}\frac{(S_1+S_2+S_3+\cdots+S_i)}{(S_1+S_2+S_3+\cdots+S_m)}. \tag{22}$$

The author argued that the test statistic, χ^2, performs better than the test statistic T, except for a few cases.

Acosta & Rojas (2009) constructed a simple information matrix (IM) misspecification test for exponential distributions that can be applied in duration models. Assume a random variable u has exponential distribution with parameter, θ. The proposed IM statistic can be expressed as:

$$IM = \frac{n\left[\frac{1}{n}\sum_{i=1}^{n}(u_i^2 - \frac{2u_i}{\hat{\theta}_n})\right]^2}{\left[\frac{1}{n}\sum_{i=1}^{n}u_i^2 - \frac{2u_i}{\hat{\theta}_n}\left(\frac{1}{\hat{\theta}_n} - u_i\right)\frac{1}{n}\sum_{i=1}^{n}2u_i\right]^2}, \tag{23}$$

where, $\hat{\theta} = n/\sum_{i=1}^{n}u_i$ and the statistic follows a Chi-square distribution with one degree of freedom. Authors concluded that this test statistic exhibited good empirical size properties and good power against Weibull and gamma distributions. They further explained that the IM test procedure can also be applied to other distributions (i.e. weibull, gamma, etc.), although the interpretation of the IM statistics is less straightforward.

Instead of using the original observations for testing the exponential distribution using the Kolmogorov-type statistics, Seshadri, Csorgo, & Stephens (1969) used two techniques to transform the original observations to the random variable which will be uniformly distributed on the null hypothesis. In one of the transformation techniques, researchers transform the observations y_i to Z_j given by,

$$Z_j = \sum_{i=1}^{j} \frac{y_i}{S}, \text{ where } S = \sum_{i=1}^{n} y_i \tag{24}$$

The authors called this transformation *J* and write *Z* = *Jy* where *Z* = (z_1, z_2, …,z_n) and ***y*** = (y_1, y_2, …, y_n) and therefore the test of hypothesis is actually a test for the uniformity of *Z*.

In another transformation, they transformed the y-values differently to produce another set Z' = (z'_1, z'_2, … z'_n) as follows. Let $y_{(i)}$ ($1 \leq i \leq n$) denote the order statistics of y. Then writing $y_{(0)} = 0$, and $d_i = (n+1-i)(y_{(i)} - y_{(i-1)})$, ($1 \leq i \leq n$) gives,

$$Z'_j = \sum_{i=1}^{j} \frac{d_i}{S}, \text{ where } S = \sum_{i=1}^{n} d_i \tag{25}$$

The authors called this transformation as a K transformation and write *Z' = Ky*. According to them, this transformation used a method discussed by Durbin (1961). The Z_i's($1 \leq i \leq n-1$) are also uniformly distributed in the unit interval and therefore the test of hypothesis is actually a test for the uniformity of *Z'*. Using the Kolmogorov-smirnov type tests, authors claimed that the *K* transformation produced more powerful results as compared to *J* transformation.

Spinelli & Stephens (1987) developed five tests for testing two parameters exponentiality (a given random sample of *n* values of *x* comes from the exponential distribution) when origin (θ) and scale (η) parameters are estimated from the data. The tests developed by authors were either EDF based or regression based tests. To understand these five test statistics, assume $X \sim \exp(\theta, \eta)$. The authors estimated the parameters required for their tests as shown in the equation 26.

$$\hat{\eta} = \frac{n(\bar{x} - x_1)}{(n-1)}, \hat{\theta} = x_1 - \frac{\hat{\eta}}{n}, \bar{x} = \sum_{i=1}^{n} \frac{x_i}{n}, w_i = \frac{(x_i - \hat{\theta})}{\hat{\eta}} \text{ and } z_i = 1 - \exp(-w_i) \tag{26}$$

The five different test statistics are given by the following equations:

$$D^{+} = \max 1 \le i \le n \left[\left(\frac{i}{n}\right) - z_i\right] \tag{27}$$

$$D^{-} = \max 1 \le i \le n \left[z_i - \left(\frac{i-1}{n}\right)\right] \tag{28}$$

$$D = \max(D^{+}, D^{-}) \tag{29}$$

$$W^2 = \sum_{i=1}^{n} \left(z_i - \frac{2i-1}{2n}\right)^2 + \frac{1}{12n} \tag{30}$$

$$U^2 = W^2 - n\left(\bar{z} - \frac{1}{2}\right)^2 \tag{31}$$

$$\bar{z} = \sum_{i=1}^{n} \frac{z_i}{n} \tag{32}$$

$$A^2 = -\frac{1}{n}\sum_{i=1}^{n}(2i-1)\{\ln z_i + \ln(1 - z_{n+1-i})\} - n, \tag{33}$$

Among these test statistics, authors claimed that the A^2 test statistic is the superior in terms of the power in their sets of parameters.

Lilliefors (1969) explored a test for testing whether a set of observations is from an exponential population when the mean is not specified but must be estimated from the sample. This test used the same concept for normality test developed by Lilliefors (1967). The test statistic, *D*, takes the form as shown in equation 34,

$$D = \max |F^{*}(x) - s(x)|, \tag{34}$$

where s(*x*) is the sample CDF and $F^{*}(x)$ is the cumulative distribution function of the exponential distribution with $\bar{X} = \frac{1}{\lambda}$, where λ is the scale parameter. Author presented the critical values for using five significance levels with various sample sizes. Using the log normal and χ^2 (1) distributions as an alternative distributions, the study compared the power of this test with the χ^2 test and found that this test is more powerful than χ^2 test for testing whether a set of observations is from an exponential distribution. The author further explained that this test can be used with sample sizes which are too small for use of the χ^2 test. It is important to note that this test can further be explored with several alternative distributions (not just two).

For testing the goodness of fit of exponential distribution, Schafer, Finkelstein & Collins (1972) proposed a test and compared it with the test presented by Lilliefors (1969). The statistic proposed was,

$$\widetilde{D}_n = \max 1 \leq i \leq n \, |S_n(x_i) - \tilde{F}(x\,;\lambda)|, \tag{35}$$

where, λ is a scale parameter, $\tilde{F}(x\,;\lambda)$ = 1 - $\{1 - \frac{x_i}{(n\bar{x})}\}^{n-1}$, $S_n(x_i)$ is the empirical distribution function (EDF). According to Pugh (1963), the test statistic, $\widetilde{D}_n$, is based on the Blackwell-Rao and Lehman-Scheffe theorems which gives the best unbiased estimate. Authors compared this statistic with the statistic proposed by Lilliefors (1969) which takes the form,

$$\widehat{D}_n = \max |\hat{F}_x - S_n(x)|, \tag{36}$$

where, $\hat{F}_x$= 1 – exp$(-\frac{x}{\bar{x}})$ and $\hat{F}$ is the maximum likelihood estimate of $F(x\,;\lambda)$. Using lognormal and χ^2 (1) as alternative distributions, authors argued that their test is more powerful than the test proposed by Lilliefors (1969) for most part of the parameters, sample size, and significance levels under study.

Barry & Margolin (1976) obtained the computationally efficient approximations of the Kolmogorov-Smirnov type (such as proposed by Lilliefors (1969)) one sample statistic to test GOFT for the exponential data with unknown scale parameter.

Rogozhnikov & Lemeshko (2012) reviewed some tests for exponentiality and compared their powers. One of the tests they reviewed will also be considered in the proposed test for the purpose of power comparison which is presented in 37. Let exp(θ) be exponential distribution with the density function f(x) = exp [(-x / θ) / θ]. In test statistic, the authors used scaled observations $Y_j = (X_j / \hat{\theta}_n)$ or their transformed values $Z_j = 1 - \exp(Y_j)$, where $\hat{\theta}_n = \bar{X}_n$ and $X_1, X_2, \ldots, X_n$ be the given independent observations of nonnegative random variables. The Cramer-Von Mises exponentiality test (CVM) is given by:

$$CVM_n = \frac{1}{12n} + \sum_{i=1}^{n} \left[Z_j - \left(\frac{2j-1}{2n} \right) \right]^2 \quad (37)$$

which is basically the modification of one sample Cramer-Von Mises test for normality in the context of testing exponentiality of the distribution which replaces the normal CDF by exponential CDF. Among the all tests studied, authors could not unambiguously choose a test with the highest power with respect to every considered competing alternative distribution. Authors further explained it was as well unrealistic to place the tests in some unconditional order (i.e. descending by power).

Grouped data can often arise due to the lack of resolution of the measurement instruments. They also arise when data are deliberately rounded to certain accuracy and are presented, say, in the form of histogram (Spinelli, 2001). Spinelli used two statistics of the Cramer-Von Mises (CVM) type to test for the exponential distribution when data are grouped. Suppose a random sample of n observations of X is given, labeled $x_1, x_2, \ldots, x_n$. The observed values of X fall into one of the K groups whose lengths may be different. When the parent distribution is exponential, the probability of an observation falling in group j is $p_j = exp^{-k_{j-1}\theta - e^{-k_j\theta}}$, j = 1, 2, …, K. Let O_j be the number of observations in group j. Let $np_j = e_j$ be the estimate of the expected number of observations in group j. Also, let $Z_j = \sum_{i=1}^{j} (O_i - e_i)$ and $H_j = \sum_{i=1}^{j} p_i$, where, i,j = 1, 2,

…, K. Finally, $t_j = (p_j + p_{j+1}) / 2$ with $p_{k+1} = p_1$. The CVM type statistics are then given by:

$$W^2 = n^{-1} \sum_{j=1}^{k} (Z_j^2 t_j) \quad (38)$$

$$A^2 = \frac{n^{-1} \sum_{j=1}^{k} (Z_j^2 t_j)}{\{H_j(I - H_j)\}} \quad (39)$$

The W^2 and A^2 test statistics are the different functions of the Z_j, H_j, and t_j which are in fact the modification of the Cramer-Von Mises (CVM) type tests to test for the exponential distribution when the data are grouped. The author compared these statistics with χ^2-GOFT which is given by:

$$\chi 2 = \sum_{i=1}^{k} \frac{(O_i - e_i)^2}{e_i} \quad (40)$$

The author concluded that the proposed tests are easy to compute, the asymptotic approximations apply for small sample sizes, and the test statistics have good power in comparison to the Pearson's Chi-Square test statistic.

Pettitt (1977) presented the asymptotic distributions of two Cramer-Von Mises type statistics used to test for the exponential distribution with censored data when the scale parameter must be estimated from the sample. The author also derived the asymptotic percentage points for these statistics.

Gail & Gastwirth (1978) developed a scale-free GOFT for testing the exponential distribution based on the Gini statistic. The Gini statistic is defined as:

$$G_n = \frac{\sum_{i=1}^{n} |X_i - X_j|}{2n(n-1)\bar{X}}, \quad (41)$$

where $\bar{X}$ is the sample mean . Authors showed that, the Gini based statistic is more powerful scale-free test of exponentiality against a variety of alternatives. Compared to the maximum likelihood test, the asymptotic relative efficiency of the Gini statistic is

0.69 against gamma and 0.88 against Weibull alternatives. On the basis of good power compared to competing tests, ease of computation, availability of exact critical values and robustness to measurement error, authors recommend the Gini statistic as a scale-free goodness-of-fit test for the exponential distribution.

Chen (2008) investigated the analysis of variance tests for testing the exponentiality of two distributions. The first statistic proposed was the V-exponential statistic for complete samples which turns out to be a normalized ratio of the square of the generalized least square estimator (also the minimum variance unbiased estimator) of the common scale parameter to a pool sum of squares about the sample means. This statistic is origin and scale invariant and has a null distribution depends only on the sample size. The statistic takes the form:

$$V(n_1, n_2) = \frac{\{n_1(\bar{Y}_1 - Y_{11}) + n_2(\bar{Y}_2 - Y_{21})\}^2}{2n^*(n^*-1)S^2}, \quad (42)$$

where, $Y_1 = [Y_{11}, Y_{12}, \ldots, Y_{1n1}]^T$, $Y_2 = [Y_{21}, Y_{22}, \ldots, Y_{2n2}]^T$, $\bar{Y}_i = \sum_{j=2}^{n_i} Y_{ij}/n_i$ (for i = 1,2 and j = 1, 2, .. n_i) , $S^2 = S_1^2 + S_2^2$, $n^* = \max(n_1, n_2)$, and $S_i^2 = \sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)^2$. The second proposed statistic was the two samples V*-exponential statistic which is in fact a two sample generalization of the one-sample Shapiro-Wilk statistic (Shapiro &Wilk, 1965). The test statistic takes the form:

$$V^*(n_1, n_2) = \frac{\{n_1(\bar{Y}_1 - Y_{11}) + n_2(\bar{Y}_2 - Y_{21})\}^2}{(n_1+n_2-2)\{(n_1+n_2-1)[S_1^2+S_2^2] - [n_1(\bar{Y}_1 - Y_{11}) + n_2(\bar{Y}_2 - Y_{21})]^2\}}, \quad (43)$$

where, $S_i^2 = \sum_{j=2}^{n_i}(y_{ij} - \bar{y}_i)^2$, i = 1, 2. Author compared W-exponential test (based on Shapiro-Wilk statistic), V-exponential statistic, and V*-exponential statistic and argued that the powers were comparable which are useful additions to the current literature on testing exponentiality of two distributions.

Bakalizi (2005) presented three GOFTs for the Rayleigh distribution with grouped data. Suppose, there is a random sample of size n from the Rayleigh distribution with PDF given by:

$$f(x,\sigma) = \frac{x}{\sigma^2}\, exp^{-\frac{x^2}{2\sigma^2}}. \tag{44}$$

Assume that the inspection times are t_i, i = 1, 2, …, k-1. Also assume that $t_0 = 0$ and $t_k = \infty$. Thus, the intervals are $[0, t_1)$, $[t_1, t_2)$, …, $[t_{k-1}, \infty)$; and the i^{th} interval is $[t_{i-1}, t_i)$. Let r_i be the number of failures in the i^{th} interval. The EDF, $F_n(t_i)$ evaluated at the upper bounds of the i^{th} interval or group is given by:

$$F_n(t_i) = \frac{1}{n}\sum_{j=1}^{i} r_i \tag{45}$$

The maximum likelihood estimator (MLE) of the distribution function of the Rayleigh distribution at t_i, $F\ (t_i, \hat{\sigma})$ is given by:

$$F\ (t_i, \hat{\sigma}) = 1 - exp^{-\frac{t_i^2}{\hat{\sigma}^2}} \tag{46}$$

A natural measure of distance between the two estimators at t_i is then given by:

$$S_i = |F_n(t_i) - F(t_i, \hat{\sigma})| \tag{47}$$

The author proposed following three statistics which are the weighted distance, S_i, at all inspection times t_1, t_2, …, t_{k-1}:

$$Q1 = \sum_{i=1}^{k-1} S_i \tag{48}$$

$$Q2 = (F(t_i, \hat{\sigma})(1 - F(t_i, \hat{\sigma})^{-0.5})S_i \tag{49}$$

$$Q3 = \sum_{i=1}^{k-1}\left(\frac{k}{2} - i\right)^2 S_i \tag{50}$$

The author compared the powers of Q1, Q2, Q3, χ^2-test, and likelihood ratio test (LRT) and concluded that Q2 test and the Chi-Squared test have the best performance with the Chi -Squared test better for smaller significance levels or smaller number of inspection intervals, and the Q2 test better otherwise. Overall, the worst test in terms of power appears to be the likelihood ratio test (LRT).

Thongteeraparp & Chodjuntug (2011) compared the powers of five GOFTs for testing exponential distribution with grouped data. Of the five GOFTs, three of them were the Q1, Q2, and Q3 statistics presented by Bakalizi (2005) as were shown above. The other two statistics were the Anderson-Darling statistics (A) and the Cramer-Von Mises statistic (W) as shown below:

$$A = n^{-1}\sum_{j=1}^{k}\frac{Z_j^2 p_j}{H_j(1-H_j)}, \tag{51}$$

where, $p_j = exp^{\left(-\frac{x_{j-1}}{\lambda}\right)} - exp^{\left(-\frac{x_j}{\lambda}\right)}$ for j = 1, 2, …, K; $H_j = \sum_{i=0}^{j} p_i$, $Z_j = N_1 + N_2 + \ldots + N_j - n(p_1 + p_2 + \ldots + p_j)$, and N_j is the number of observations in the j^{th} interval.

$$W = n^{-1}\sum_{j=1}^{k} Z_j^2 p_j \tag{52}$$

The authors claimed that the empirical type I error rates at the nominal 0.05 level of significance, the statistics Q1, Q2 and Q3 can control the type I error for all sample sizes and number of inspection intervals. Statistics A and W cannot control the type I error for the number of inspection intervals equal six for sample size equal to 50 and statistic W cannot control the type I error for number of inspection intervals equal to six for all sample sizes, number of inspection intervals equal seven for sample size *n* equal 50 and 100 and number of inspection intervals equal ten at sample size *n* equal 100. The tests Q2 and Q1 have more powers than Q3, W, and A tests.

Morris & Szynal (2013) presented GOFTs for six distributions (exponential, Weibull, extreme value, logistic, normal, and Cauchy). Authors derived the GOFTs from characterization conditions of continuous distributions in terms of moments of the k^{th}

record values. In order to estimate the associated expectations, they used U-statistics. Their study mostly focused on the mathematical derivation of the associated expectations which will be very helpful to understanding the theory behind the expectations of the population parameters under considerations.

Using the Integrated Distribution Function (IDF), Klar (2001) proposed GOFTs for exponential and the normal distributions. The test is based on the IDF, $\Psi(t) = E(X-t)^+ = \int_t^\infty (1 - F(x))dx$ and the EDF, $\Psi_n(t)$, as shown below:

$$\Psi_n(t) = \int_t^\infty (1 - F_n(x))dx \quad = \frac{1}{n} \sum_{i=1}^{n} (X_i - t)\, 1\{X_i > t\}, \tag{53}$$

where, 1 denotes the indicator function, and $F_n(x) = n^{-1} \sum_{j=1}^{n} \{X_j \le x\}$ which is the EDF of $X_1, X_2, \ldots, X_n$. The proposed test statistic for testing exponentiality is scale-invariant and turned out to be:

$$T_n = \frac{n}{2} - 2\sum_{i=1}^{n} e^{-Y_i} - (3n)^{-1} \sum_{i=1}^{n} (n-i-1) Y_{(i)}^3 + n^{-1} \sum_{i<j} Y_{(J)} \tag{54}$$

The author compared the power of this test with the CVM-type test (W^2) and AD-type test (A^2) and concluded that the proposed test is a serious competitor to classical tests for exponentiality (W^2 and A^2) .

Baratpour & Rad (2012) developed a new exponentiality test based on the cumulative residual entropy. The proposed test statistic takes the form:

$$T_n = \frac{\sum_{i=1}^{n-1} \frac{n-1}{n}\left(ln\frac{n-1}{n}\right)\left(X_{(i+1)} - X_{(i)}\right) + \sum_{i=1}^{n} \frac{X_i^2}{2\sum_{i=1}^{n} X_i}}{\sum_{i=1}^{n} \frac{X_i^2}{2\sum_{i=1}^{n} X_i}} \tag{55}$$

The power of this test was compared with the S* test (Proposed by Finkelstein & Schafer (1971), Lilliefors test (1969), W^2 test (proposed by Van-Soest (1969) and the KLC_{mn} test proposed by Choi, Kim, & Song (2004). Authors argued that the power of these test were almost identical but the proposed test was claimed to be computationally easier.

The Lilliefors test was found to have low power by several authors. The Type II statistics (Shaw, 1994) were found to have higher power than the Kolmogorov-Smirnov test. Overholt & Schaffer (2013) established that their test has more power than Lilliefors test (1967). Articles reporting the sum of all the absolute differences between the exponential CDF and EDF (continuous variable) are almost non-existence. This study extended the concept of sum of all the differences from Shaw (1994) and Overholt & Schaffer (2013) in the context of exponentiality test. Hence, it is reasonable to assume the proposed test would outperform for both the KS-test and the Lilliefors test (1969). The proposed test statistic takes the form as shown in equation 56 and will be discussed further in chapter three,

$$D = \sum_{i=1}^{n} |F^*(x_i) - S(x_i)|, \tag{56}$$

where $F^*(x_i)$ is the CDF of exponential distribution using the maximum likelihood estimator for the scale parameter θ and $S(x_i)$ is the sample cumulative distribution function.

CHAPTER III

METHODOLOGY

This chapter summarizes the derivation of a proposed test statistic, data sources, relevant R syntax and the strategies to address specific research questions from chapter I.

Development of Test Statistic

The proposed study is a right tail test which considers the sum of all the absolute differences between exponential cumulative distribution function (CDF) and the sample empirical distribution function (EDF) hoping to gain more power than the Lilliefors test (1969). The proposed modified Lilliefors test statistic (PML) takes the form,

$$PML = \sum_{i=1}^{n} |F^*(x_i) - S(x_i)|, \quad (57)$$

where $F^*(x_i)$ is the CDF of exponential distribution using the maximum likelihood estimator for the scale parameter θ and $S(x_i)$ is the sample cumulative distribution function. The estimator $\hat{\theta}$ is the uniformly minimum variance unbiased estimator (UMVUE) of the scale parameter θ.

The CDF, $F^*(x_i)$, is given by 58

$$F^*(x_i) = 1 - \exp(-\frac{xi}{\bar{x}}), \quad (58)$$

where $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$. The EDF is given by 59

$$S(x_i) = i/n. \quad (59)$$

Development of Critical Values

For each sample size (4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45 and 50), 50,000 trials (replications) of size *n* were generated from an exponential distribution. The proposed test statistic was determined for each replication. Critical values (CV) were determined from these groups of replications. Three different significance levels (0.01, 0.05, and 0.10) were considered. Since the proposed test is a right tail test, the critical value for various significance levels are $50{,}000*(1-\alpha)^{th}$ ordered value of the simulated test statistics. For example, for $\alpha = 0.05$, $50{,}000*(1-0.05) = 47{,}500^{th}$ test statistic was the observation which was smaller than only 2,500 other observations. Three scale parameters ($\theta = 1, 5, 10$) were used to generate critical values. The scale parameters were arbitrarily chosen.

Power Analyses Procedures

To compare the power of the proposed test and the other four exponentiality tests, this study utilized three significance levels (0.01, 0.05 and 0.10) and 50,000 replications were drawn from each sample size (*n* = 5, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 1000, and 2000). This study compared the power of the proposed test to the Lilliefors test (LF-test), Schafer et al. test (D-test), Finkelstein and Schafers statistics (S-test) and Cramer-Von Mises test (CVM-test). A total of 12 alternative distributions were utilized (Weibull(1,0.50), Weibull(1,0.75), Gamma(4,0.25), Gamma(0.55,0.275), Gamma(0.55,0.412), Gamma(4,0.50), Gamma(4,0.75), Gamma(4,1), Chi-Square(1), Chi-Square(2), t(5), and log-normal (0,1)) to see how the proposed test statistic works. Among 12 alternative distributions, only the t(5) distribution is the symmetric distribution. The rest 11 distributions are right skewed distributions. These distributions covered a wide range of skewness and kurtosis which were arbitrarily chosen. Table 1 presents the skewness and kurtosis by distributions used in this study.

Table 1
Skewness and Kurtosis of Alternative Distributions

Distributions	Skewness	Kurtosis
Weibull(1,0.50)	6.62	87.72
Weibull(1,0.75)	3.06	18.51
Gamma(4,0.25)	4	12
Gamma(0.55,0.275	3.81	11.44
Gamma(0.55,0.412)	3.12	9.35
Gamma(4,0.50)	2.83	8.49
Gamma(4,0.75)	2.31	6.93
Gamma(4,1)	1	6.00
Chi-Square(1)	2.83	15
Chi-Square(2)	2	9
t(5)	0	9

Shapiro, Wilk, & Chen (1968) classified the continuous distributions into five major groups based on the nature of the alternative distributions. Their classifications are summarized in table 2.

Table 2
Classification of Continuous Distributions

Group	Skewness	Kurtosis	Category
1	> 0.30	> 3.00	Asymmetric, long-tailed
2	> 0.30	< 3.00	Asymmetric, short-tailed
3	≤ 0.30	> 4.50	Symmetric, long-tailed
4	≤ 0.30	< 2.50	Symmetric, short-tailed
5	≤ 0.30	2.5≤ Ku≤ 4.50	Near Normal

Number of Trials, Significance Levels and Alternative Distributions

Sample sizes and number of trials are important variables in Monte Carlo simulations for power comparisons. Of the previous studies discussed in chapter II, 35 of them directly compared powers among several GOFTs. As seen in Figure 2, researchers have used anywhere from 400 – 1,000,000 replications in their simulation studies, with 10,000 replications being the most popular choice. A Pareto chart from the articles was cited in chapter II and is shown in figure 2.

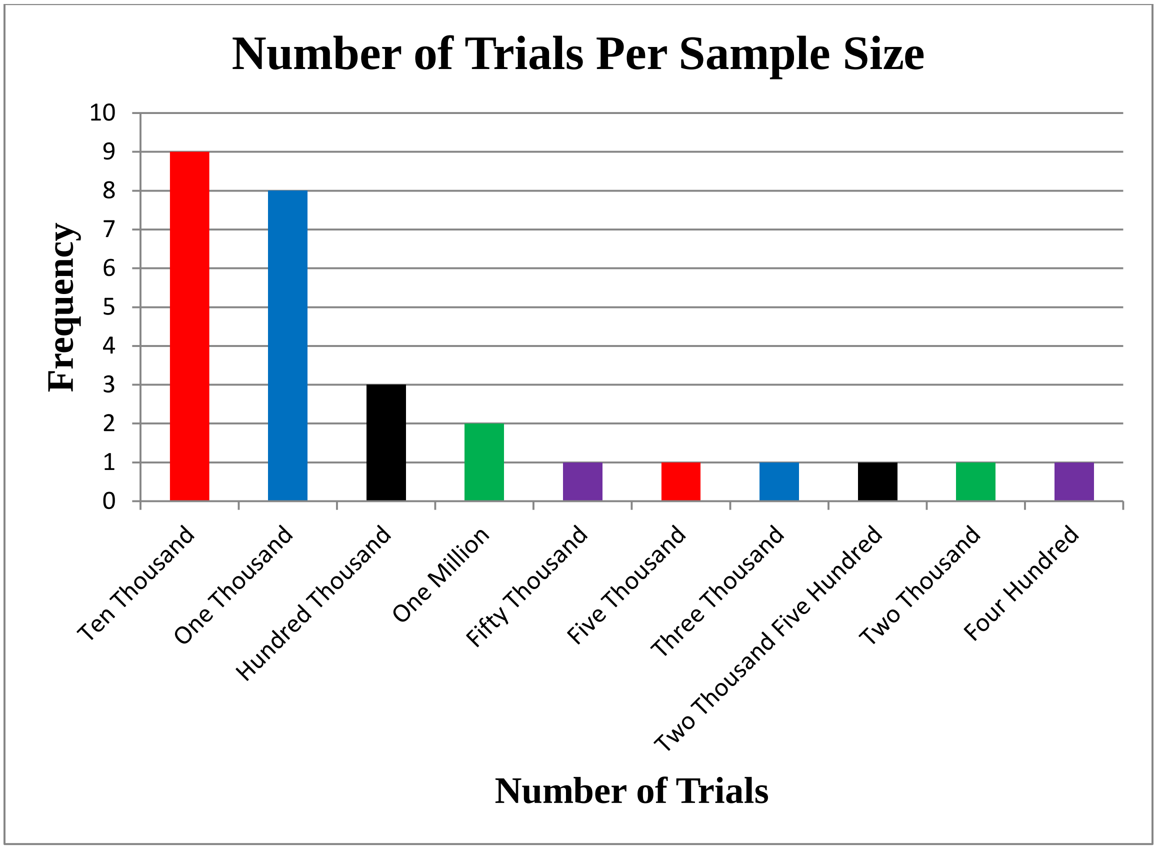

Figure 2. Number of Trials on Monte Carlo Simulations

Several studies have been conducted to approximate the optimum number of replications for given parameter settings. According to Hutchinson & Bandalos (1997):

> Once the variables and levels within variables have been selected and the design has been specified, the next decision involves selecting the number of replications, where replications essentially represent the number of times the

> analysis will be repeated with a different sample. With a large number of replications, the sampling distribution of results can be examined. With too few replications, idiosyncratic results based on a particular sample are more likely to arise. Unfortunately for simulation researchers there are no definite guidelines for selecting the appropriate number of replications. The specific number will depend on the type of phenomenon being studied, the extent to which the steps of the simulation can be automated, as well as available computer resources. Wilcox (1988) recommended 10,000 replications as a conservative choice, whereas Robey and Barcikowski (1992) suggested that in some cases over 100,000 replications might be needed to adequately detect discrepancies between nominal and actual type I error rates. However, in some areas of research such as discriminant analysis, the number of replications has varied from 2 to 5,000 (Sedek & Huberty, 1994). In structural equation modeling, it is not uncommon to see as few as 20 replications (Browne & Cudeck, 1989; MacCallum, Roznowski, & Necowitz, 1992). (Hutchinson & Bandalos, 1997, p. 238)

Schaffer & Kim (2007) studied the number of replications required in control charts and indicated that using 10,000 replications was unnecessarily large and a smaller number of replications could be used to reproduce the target average run lengths within the 2% error bands. In many cases, only 5,000 replications or fewer were required. Lilliefors (1967) used 1,000 replications. In this study, 50,000 replications (trials) were run for each sample size.

Lilliefors (1967) used 0.01, 0.05, 0.10, 0.15, and 0.20 significance levels. Of the previous studies discussed in chapter II, authors used 11 different significance levels for power comparisons. The top three significance levels used were the 5 %, 10 % and 1 %. A Pareto chart of significance level used by several authors as discussed in chapter II are shown in figure 3. The proposed study used 0.01, 0.05 and 0.10 significance levels.

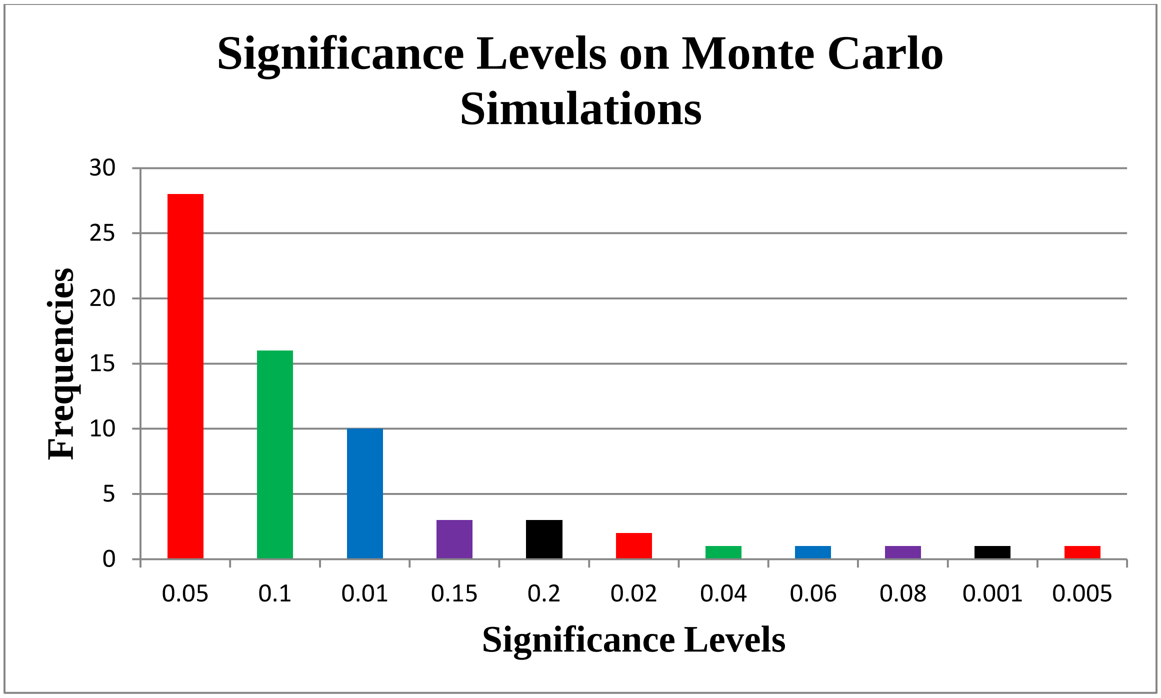

Figure 3. Significance Levels in Previous Studies

Of the studies discussed in chapter II, authors used 20 different types of alternative distributions for power comparisons. For each type of distribution there were several parameters' combinations. Among the 20 different types of alternative distributions, eight distributions were symmetric and 12 distributions were non-symmetric. These distributions are presented in figures 4 & 5. The Top five non-symmetric distributions used were: Chi-square, lognormal, Beta, Weibull, and exponential. Similarly, the top five symmetric distributions used were: t, Uniform, Cauchy, Laplace, and Logistic. Lilliefors (1967) used Chi-square (3), t (3), exponential, and Uniform (0, 1) distributions.

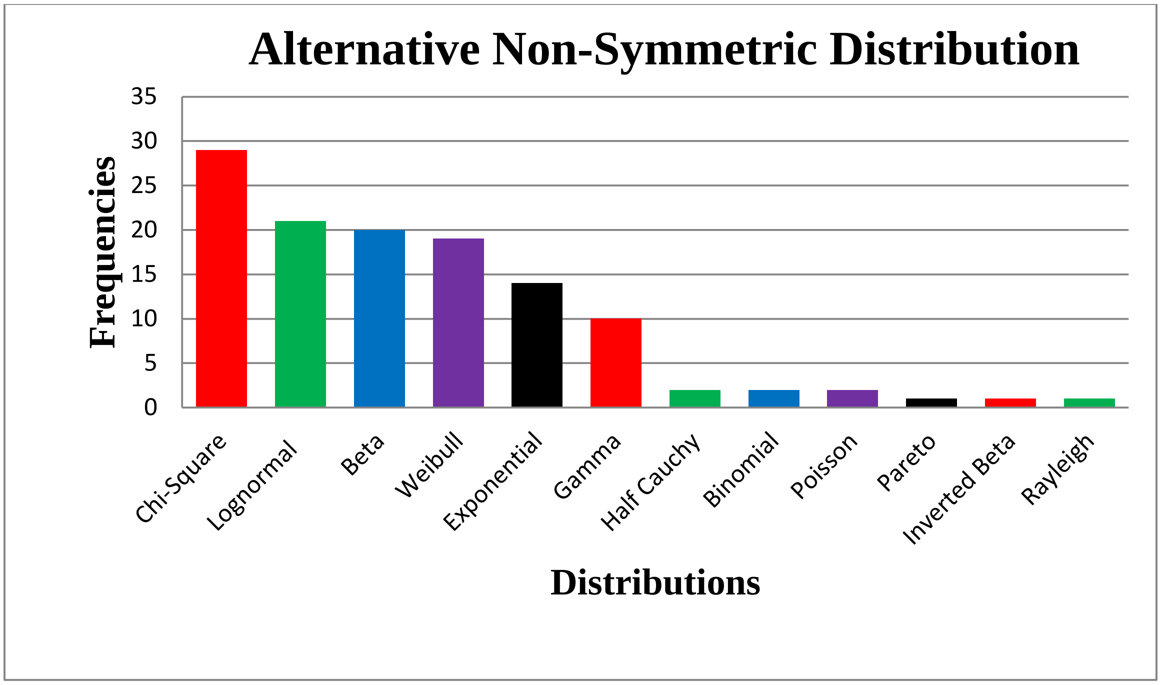

Figure 4. Alternative Non-Symmetric Distributions in Previous Studies

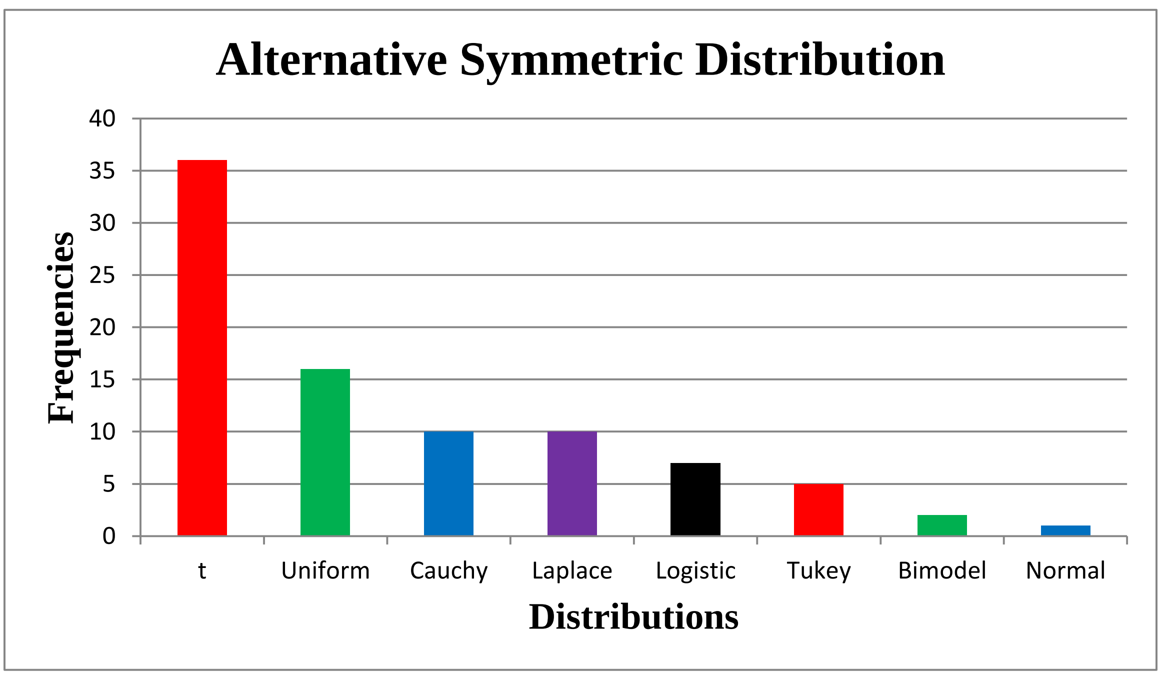

Figure 5. Alternative Symmetric Distributions in Previous Studies

The proposed study used 12 alternative distributions (Weibull(1,0.50), Weibull(1,0.75), Gamma(4,0.25), Gamma(0.55,0.275), Gamma(0.55,0.412), Gamma(4,0.50), Gamma(4,0.75), Gamma(4,1), Chi-Square(1), Chi-Square(2), t(5) and log-normal (0,1)) to see how the proposed test statistic works. Among 12 alternative distributions, only the t(5) distribution is symmetric. The rest 11 distributions are right skewed distributions.

Of the studies discussed in chapter II, authors used 19 different sample size patterns from three to 2,000. Lilliefors (1967) used four to 30 (inclusive) and over 30 sample sizes. This study used 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45 and 50 sample sizes for obtaining critical values. This will further be mentioned when addressing specific research questions.

Research Questions Revisited

Below, each research question from Chapter I is restated and addressed individually in order to describe how this study would answer each of the research question using the defined parameter settings.

Q1 How will the proposed test be designed to assure reliable critical values and their corresponding significance levels?

This study used data simulation techniques to mimic the desired parameters settings. Three different scale parameters (θ = 1, 5, and 10) were used to generate random samples from exponential distribution. Sample sizes 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45 and 50 were used. The study considered three different significance levels (α) (0.01, 0.05 and 0.10). For each sample size and significance level, 50,000 trials were run from an exponential distribution which generated 50,000 test statistics. The 50,000 test statistics were then arranged in the order from smallest to largest. The proposed test is a right tail test. If α = 0.05 is considered, the 95th percentile of the test statistic was used as the critical value for the given sample size.

> Q2 For specified significance levels, how will the proposed test perform in terms of detecting departures from exponentiality for data simulated from 12 alternative distributions?

Data were produced from varieties of 12 distributions (Weibull(1,0.50), Weibull(1,0.75), Gamma(4,0.25), Gamma(0.55,0.275), Gamma(0.55,0.412), Gamma(4,0.50), Gamma(4,0.75), Gamma(4,1), Chi-Square(1), Chi-Square(2), t(5) and log-normal (0,1)) to see how the proposed test statistic works. Fifty thousand replications were drawn from each distribution for sample sizes 5, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 1000, and 2000. For each sample size, the proposed test statistic and critical values were compared to make decisions about the null hypothesis. There were 50,000 trials for each sample size. The study tracked the number of rejections (rejection yes or no) in 50,000 trials to evaluate capacity of the proposed test to detect the departure from exponentiality.

> Q3 For specified significance levels, how will the proposed test compare in terms of power with the four other exponentiality tests (Cramer-Von Mises test (CVM-test), Lilliefors test (LF-test), Finkelstein & Schafers statistics (S-test) and $\widetilde{D}_n$-test as shown in 60, 61, 62, and 63 respectively?

This study used the distributions, sample sizes and alpha levels as mentioned above in Q1 and Q2 for this purpose.

Cramer-Von Mises test (CVM) is given by:

$$CVM_n = \frac{1}{12n} + \sum_{i=1}^{n}\left[t_i - \left(\frac{i-0.5}{n}\right)\right]^2, \tag{60}$$

where, $t_i = 1 - \exp(-\frac{xi}{\bar{x}})$, and $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$. Lilliefors test (LF-test) is given by:

$$D = \underset{x}{\text{Sup}}\, |F^*(x_i) - S(x_i)|, \tag{61}$$

where, $F^*(x_i) = 1 - \exp(-\frac{xi}{\bar{x}})$, $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$, and $S(x_i)$ is the empirical distribution function (EDF). Finkelstein & Schafers statistics (S-test) is given by:

$$S = \sum_{i=1}^{n} \max\left\{\left|F_0\left(X_{(i)}, \hat{\theta}\right) - \frac{i}{n}\right|, \left|F_0\left(X_{(i)}, \hat{\theta}\right) - \frac{i-1}{n}\right|\right\}, \quad (62)$$

where, $\hat{\theta} = \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$. Schafer et al. (1972) test ($\widetilde{D}_n$) (here after denoted by D-test) is given by:

$$\widetilde{D}_n = \max 1 \leq i \leq n \, |S_n(x_i) - \tilde{F}(x\,;\lambda)|, \quad (63)$$

where, λ is a scale parameter, $\tilde{F}(x\,;\lambda)$ = 1 - $\{1 - \frac{x_i}{(n\bar{x})}\}^{n-1}$, $S_n(x_i)$ is the EDF. According to Pugh (1963), the test statistic, D-test, is based on the Blackwell-Rao and Lehman-Scheffe theorems which gives the best unbiased estimate.

Software and Programming Considerations

The study used R 3.0.2 for most of the simulations to generate test statistics, critical values and power comparisons. The outputs from R were presented in tables and charts in chapter IV. Microsoft Excel 2010 was also used to make tables and charts. The required R syntaxes were presented in appendix A. Monte Carlo simulation techniques were used to generate random numbers which were used to approximate the distribution of critical values for each test.

CHAPTER IV

RESULTS

This chapter answers all the three research questions from chapter I based on the Monte Carlo simulations whose computational algorithms rely on repeated random sampling to obtain numerical results. This study developed a new test of exponentiality by modifying the Lilliefors test of exponentiality. Lilliefors considered the maximum differences between the empirical distribution function (EDF) and the cumulative distribution function (CDF). The proposed test considered the sum of all the absolute differences between the CDF and EDF. The proposed test statistic is not only easy to understand but also very simple and easy to compute.

Below, each research question from Chapter I is restated and addressed individually in order to describe how this study answered each of the research question using the defined parameter settings.

Development of Critical Values

Q1 How will the proposed test be designed to assure reliable critical values and their corresponding significance levels?

This study used data simulation techniques to mimic the desired parameters settings. Three different scale parameters (θ = 1, 5, and 10) were used to generate random samples from exponential distribution. Sample sizes 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45 and 50 were used. The study considered three significance levels (0.01, 0.05, and 0.10). The actual distribution of the proposed test statistic is unknown. So, this study used the data simulation techniques to approximate the critical values instead of using its asymptotic distribution.

For each sample size and significance level, 50,000 trials were run from an exponential distribution which generated 50,000 test statistics. The 50,000 test statistics were then arranged in the order from smallest to largest. The proposed test is a right tail test. The critical value for various significance levels are 50,000*(1-α)th ordered value of the simulated test statistics. So, this study used 99th, 95th, and 90th percentile of the test statistics as the critical values for the given sample size for the 0.01, 0.05, and 0.10 significance levels respectively. Table 3 shows the critical values for the proposed test. Due to space limitations, only five digits are shown on table 3.

Table 3
Critical Values for the Proposed Exponentiality Test (θ = 1)

n	α = 0.01	α = 0.05	α = 0.10
4	1.0567	0.8331	0.7409
5	1.1760	0.9315	0.8202
6	1.2703	1.0109	0.8931
7	1.3642	1.0856	0.9562
8	1.4647	1.1580	1.0189
9	1.5403	1.2209	1.0757
10	1.6274	1.2875	1.1310
15	1.9444	1.5561	1.3653
20	2.2271	1.7731	1.5636
25	2.4762	1.9682	1.7342
30	2.7097	2.1624	1.9066
35	2.9111	2.3291	2.0584
40	3.1062	2.4837	2.1904
45	3.3216	2.6331	2.3204
50	3.4557	2.7526	2.4309

The critical values from the simulated data generated for the three different values of the scale parameters (θ = 1, 5, and 10) are exactly the same for the set of parameters. It appeared that the critical values for the proposed test are the functions of the sample size (*n*) and the significance levels (α) but invariant with the choice of the scale parameter (θ).

Analyses of Significance Level

Q2 For specified significance levels, how will the proposed test perform in terms of detecting departures from exponentiality for data simulated from 12 alternative distributions?

To answer the second research question, it was relevant to verify the accuracy of the intended significance levels and to analyze the power of the proposed test. To verify the accuracy of the three intended significance levels (α = 0.01, 0.05, and 0.10), data were generated from exponential distributions (null distribution: exponential (θ = 5) and alternative distribution: exponential (θ = 10)). For sample sizes 5, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 1000, and 2000; 50,000 trials were performed and the null hypothesis of data came from an exponential distribution was tested by five exponentiality tests. To allow for a better view of the five exponentiality tests across all sample sizes and significance levels, the columns for Lilliefors test are labelled by “LF”, Cramer-Von Mises test by “CVM”, proposed modified Lilliefors test by “PML”, Shafer et al. test by “D” and Finkelstein & Schafers test by “S” for the rest of the tables presented in this study. The number of times a given test reject null hypothesis was counted and the total number of rejections were divided by 50,000 which should be pretty close to the intended significance levels. The simulated significance levels are presented on tables 4 and B.1. Due to the limitations of the space, the simulated significance levels are rounded to three digits.

Table 4
Average Simulated Significance Levels

α	LF	D	CVM	S	PML
0.01	0.010	0.010	0.010	0.010	0.010
0.05	0.051	0.051	0.051	0.051	0.051
0.10	0.100	0.100	0.101	0.101	0.101

The results showed that all five tests of exponentiality worked very well in terms of controlling the intended significance levels. The study found that the proposed test performs very closely to other four tests of exponentiality in terms of the accuracy of the intended significance levels (for each sample size and overall averages across the 19 different sample sizes).

To analyze the power of the proposed test, data were generated from 12 different alternative distributions (combination of 19 sample sizes and 3 significance levels). The results of the power analysis showed that powers were increased with increased sample sizes. Similarly, powers were also increased with the higher significance levels (higher values of α) in the set of the parameters under consideration. The detailed results of the power analysis are discussed below in answering research question 3, while comparing the powers across the five exponentiality tests.

Power Analyses

Q3 For specified significance levels, how will the proposed test compare in terms of power with the four other exponentiality tests (Cramer-Von Mises test (CVM-test), Lilliefors test (LF-test), Finkelstein & Schafers statistics (S-test) and $\widetilde{D}_n$-test as shown in 60, 61, 62, and 63 respectively?

To compare the power of the proposed test and the other four exponentiality tests, this study utilized three different significance levels (0.01, 0.05, and 0.10) and 50,000 replications were drawn from each sample size (*n* = 5, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 1000, and 2000). A total of 12 alternative distributions were utilized (Weibull(1,0.50), Weibull(1,0.75), Gamma(4,0.25), Gamma(0.55,0.275), Gamma(0.55,0.412), Gamma(4,0.50), Gamma(4,0.75), Gamma(4,1), Chi-Square(1), Chi-Square(2), t(5) and log-normal (0,1)) for power comparisons. The tables and figures of power analysis for every one of the twelve alternative distributions which are not in the body of text can be found in appendix B.

First consider the relationship between the alternative distribution, Weibull (1, 0.50) and the simulated power. Table B.2 and figure 6 summarize the power analysis for the Weibull (1, 0.50) alternative distribution. The PML-test outperformed the power for all other four exponentiality tests across all significance levels and sample sizes. The power of all four exponentiality tests exceeded the LF-test. The CMV-test, the D-test, and the S-test showed similar performance in power. It appears that for sample sizes 40 or more, the powers for all five exponentiality tests close to 1.

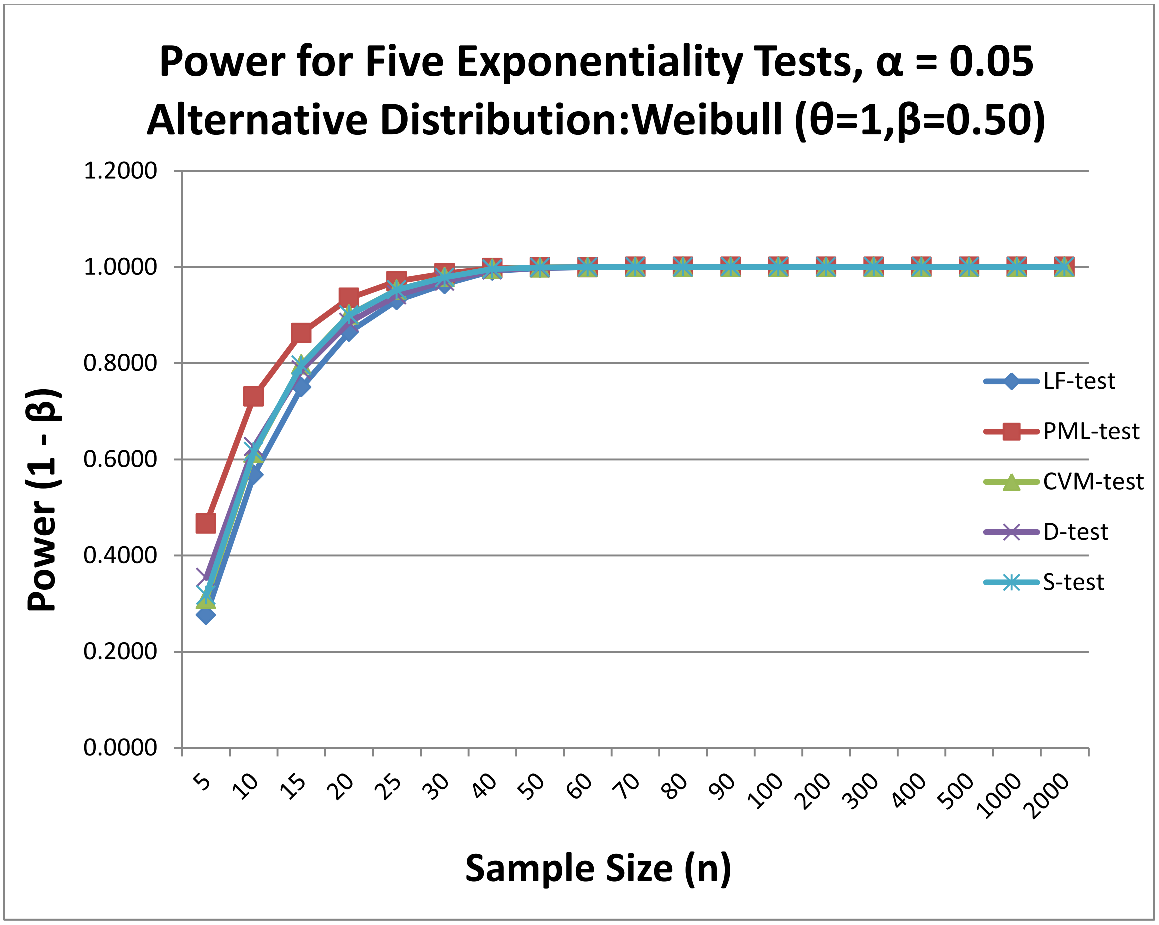

Figure 6. Power for Alternative Distribution: Weibull (1, 0.50)

Second consider the relationship between the alternative distribution, Weibull (1, 0.75) and the simulated power. Table B.3 and figure 7 summarize the power analysis for the Weibull (1, 0.75) alternative distribution. This distribution has the same scale parameter ($\theta = 1$) with the previous Weibull (1, 0.50) distribution but the shape parameter (β) is changed from 0.50 to 0.75. This caused the power to reduce substantially across all sample sizes and all significance levels under consideration.

The PML-test outperformed the power for all other four exponentiality tests across all sample sizes and significance levels. In all parameter settings under investigation, the powers for the LF-test were the lowest as compared to other four exponentiality tests. The powers of the S-test and CVM-test were almost identical across all sample sizes and significance levels. For a fixed significance level, the powers for the D-test were greater than the S-test and CVM-test for small sample sizes but this relationship was reversed for medium to large sample sizes. For all significance levels with sample sizes at least 200, the powers for all five exponentiality tests were almost equal and they approach 1.

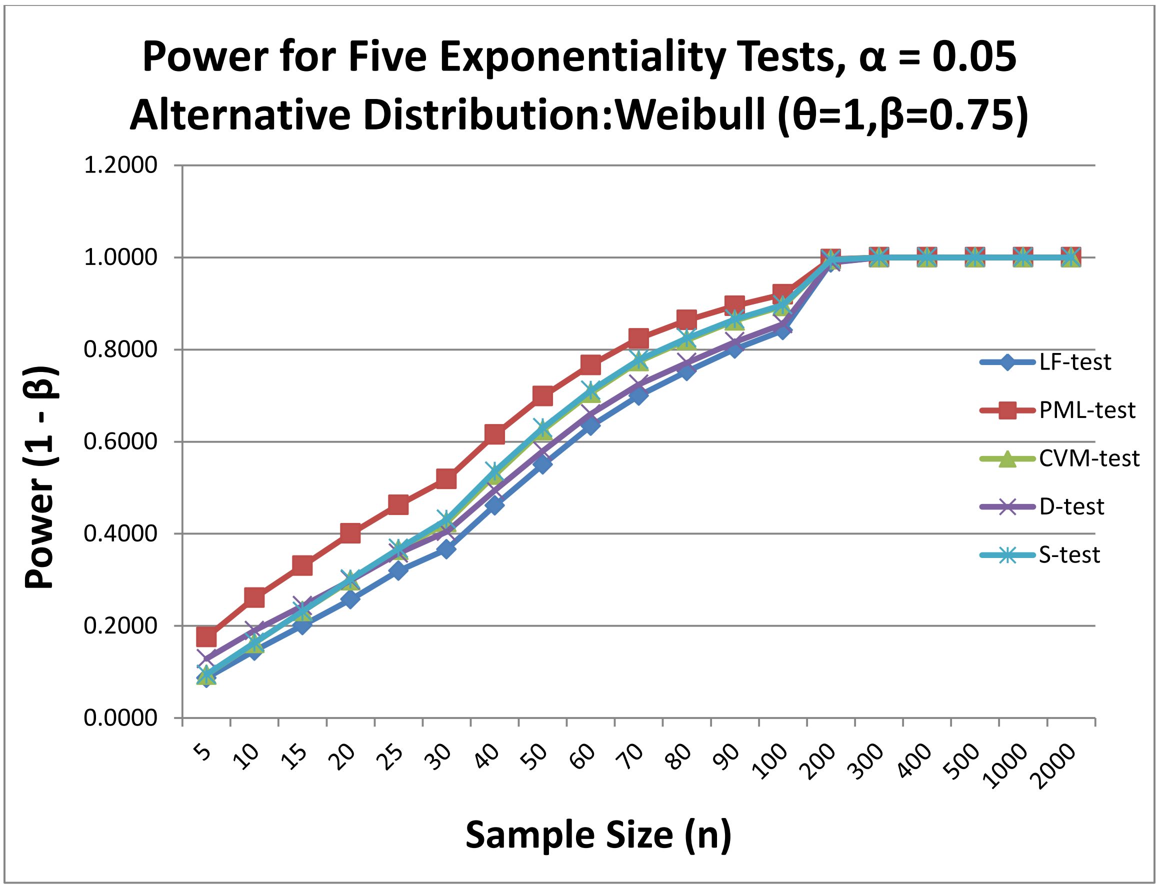

Figure 7. Power for Alternative Distribution: Weibull (1, 0.75)

Third consider the relationship between the alternative distribution, Gamma (4, 0.25) and the simulated power. Table B.4 and figure 8 summarize the power analysis for the Gamma (4, 0.25) alternative distribution. According to Bain & Engelhardt (1992), the shape parameter, *k*, in the Gamma distribution determines the basic shape of the graph of the probability distribution function (PDF). The value of the shape parameter in null distribution is 1 and the shape parameter in this alternative distribution is 0.25 which are much different. The PML-test outperformed the powers of all other four exponentiality tests across all sample sizes and all significance levels under consideration. For a fixed significance level, the powers of the D-test, CVM-test, and S-test exceeded the powers of the LF-test for small sample sizes. For medium to large sample sizes, the LF-test, D-test, S-test, and the CVM-test exhibited the identical power across all significance levels. In all parameter settings, the powers of the D-test, the CVM-test and the S-test were similar. For sample sizes at least 40, the powers of all five exponentiality tests were found almost equal which were close to 1 across all significance levels.

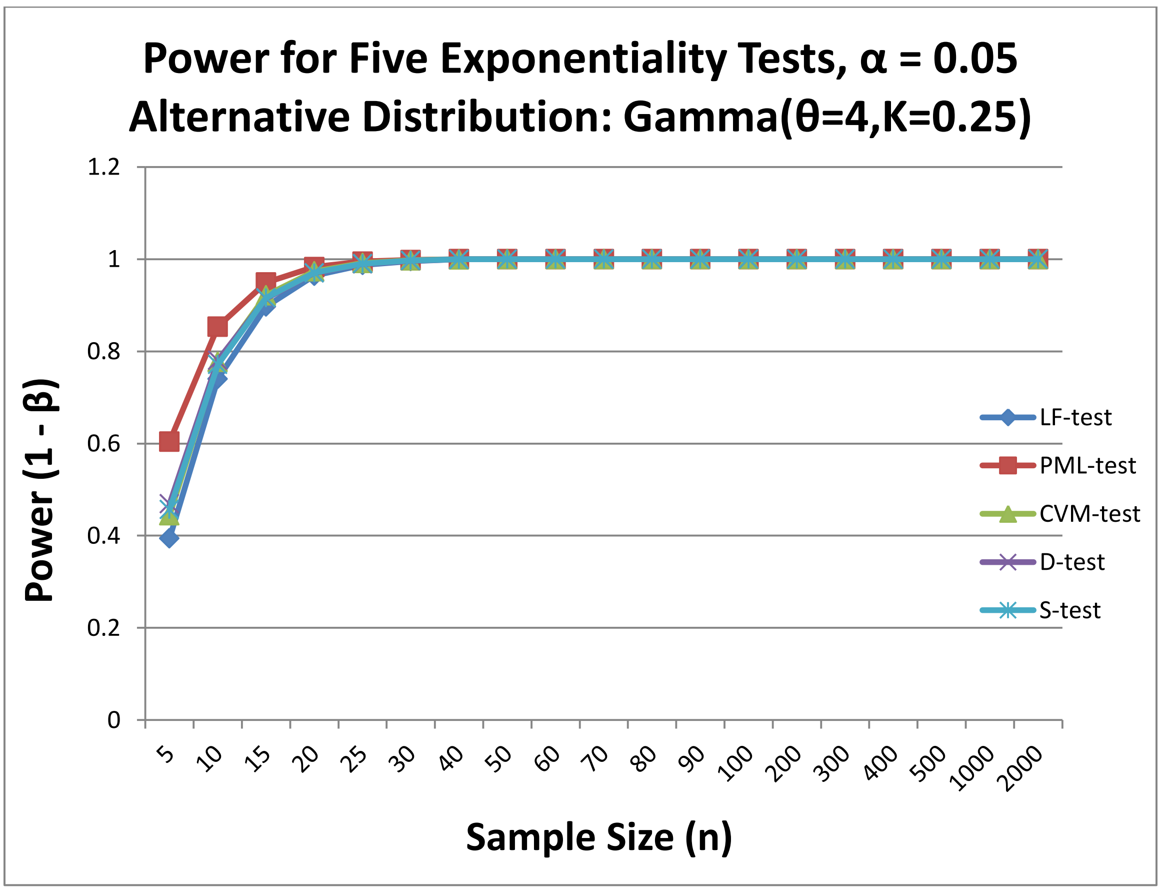

Figure 8. Power for Alternative Distribution: Gamma (4, 0.25)

Forth consider the relationship between the alternative distribution, Gamma (0.55, 0.275) and the simulated power. Table B.5 and figure 9 summarize the power analysis for the Gamma (0.55, 0.275) alternative distribution. The PML-test outperformed other four exponentiality tests across all sample sizes and significance levels. The LF-test exhibited the lowest power across all sample sizes and significance levels. For sample sizes at least 50, the powers for all five tests were found almost equal which were close to 1 across all significance levels. In all parameter settings, the powers for the CVM-test, the D-test, and the S-test were identical but all these three tests outperformed the LF-test across all sample sizes and significance levels.

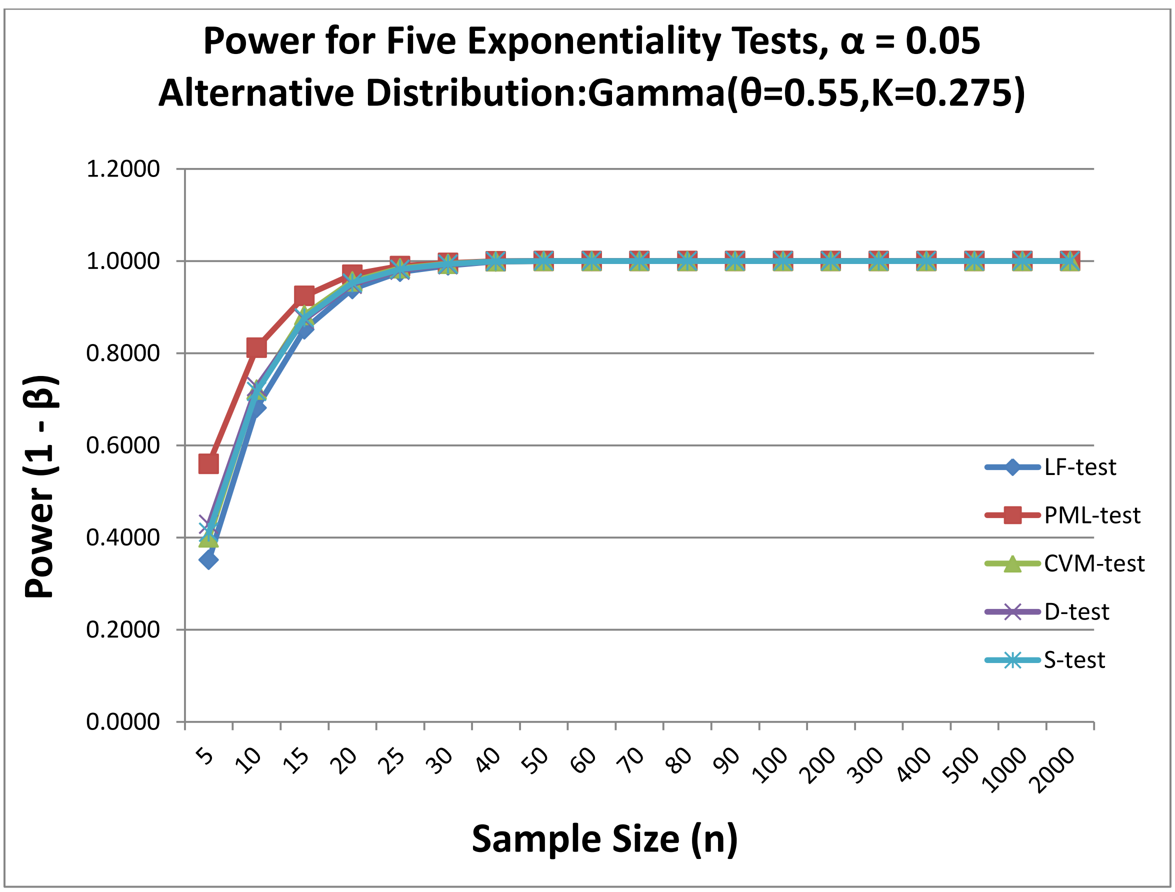

Figure 9. Power for Alternative Distribution: Gamma (0.55, 0.275)

Although the overall power trends in the previous alternative distribution (Gamma (4, 0.25)) and this distribution were similar among five exponentiality tests, the powers for this distribution was lower than the previous alternative distribution across all sample sizes and significance levels. In the previous alternative distribution, the value of the shape parameter (*K*) is 0.25 which is 0.275 in this alternative distribution.

Fifth consider the relationship between the alternative distribution, Gamma (0.55, 0.412) and the simulated power. Table B.6 and figure 10 summarize the power analysis for the Gamma (0.55, 0.412) alternative distribution. The PML-test outperformed other four exponentiality tests across all sample sizes and significance levels. The LF-test exhibited the lowest power across all sample sizes and significance levels. For sample sizes at least 80, the powers for all five tests were found almost equal which were close to 1 across all significance levels. In all parameter settings, the powers for the CVM-test, the D-test, and the S-test were identical but all these three tests outperformed the LF-test across all sample sizes and significance levels. Comparing the powers for this alternative distribution with the previous alternative distribution (Gamma (0.55, 0.275)), the powers were reduced in this alternative distribution across all sample sizes and significance levels. This is due to only the change in shape parameter (*k*) from 0.275 to 0.412. The scale parameters (θ) were the same on these two alternative distributions. It is relevant to argue that for Gamma alternative distribution, the powers for these five exponentiality tests depend only on the shape parameter (*k*). It is also important to note that the shape parameter (*k*) in the null distribution was 1. So, this study showed that as the shape parameter in the alternative distribution is close to the shape parameter of the null distribution, the simulated powers would be decreased.

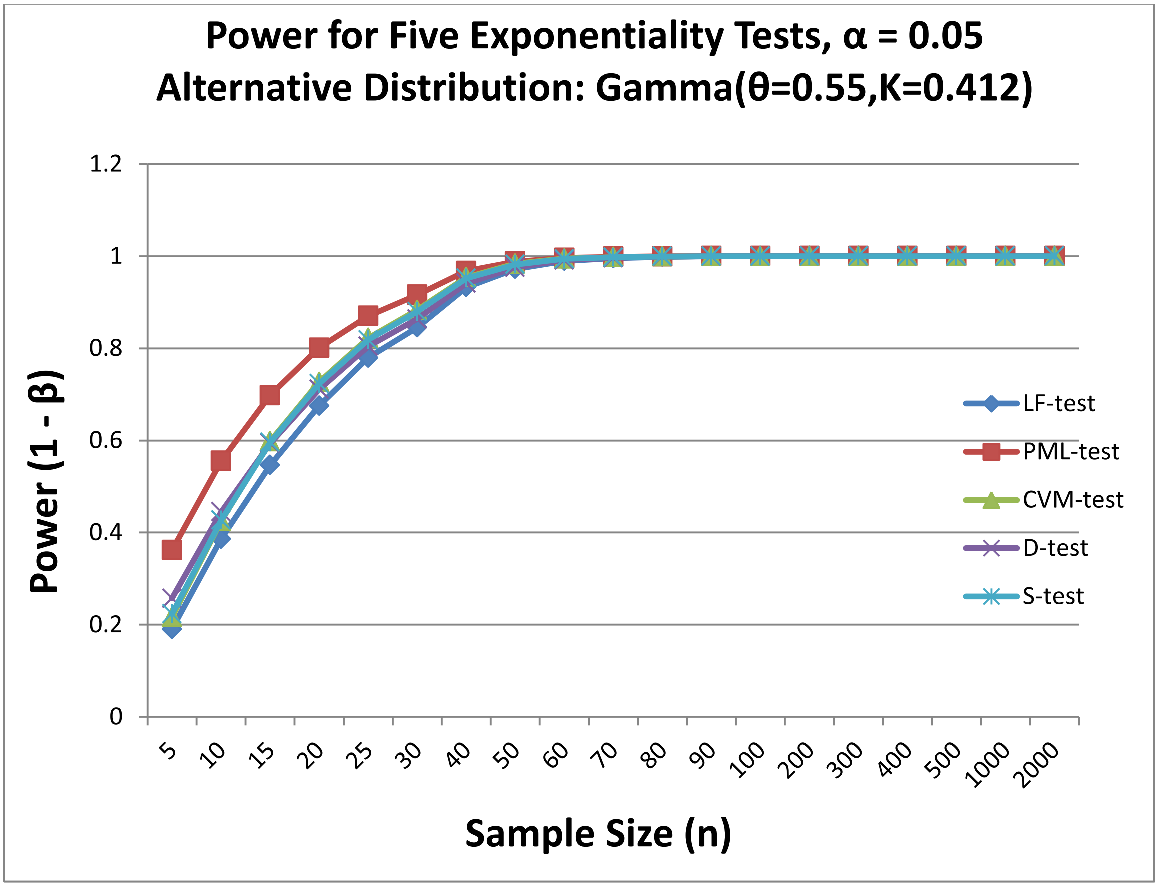

Figure 10. Power for Alternative Distribution: Gamma (0.55, 0.412)

Before considering the power for next two alternative distributions, it is imperative to discuss that the Chi-Square distribution is a special case of Gamma distribution. According to Bain and Engelhardt (1992), if a variable Y is a special Gamma distribution with scale parameter ($\theta = 2$) and shape parameter ($k = \nu/2$), the variable Y is said to follow a Chi-Square distribution with ν degrees of freedom. So, if Y ~ Gamma ($\theta = 2$, $k = \nu/2$), a special notation for this distribution can be written as:

$$Y \sim \chi^2 (\nu) \tag{64}$$

Using 64, the Gamma (4, 0.5) and the Chi-Square (1) distributions are equivalent. This study previously showed that the power for the Gamma distribution depends only on the shape parameter (*k*). So, the powers of the Gamma (4, 0.5) and Chi-Square (1) alternative distributions must be equivalent.

Sixth consider the relationship between the alternative distributions, Gamma (4, 0.5), Chi-Square (1) and the simulated power. Table B.7 and figure 11 summarize the power analysis for the Gamma (4, 0.5) and Chi-Square (1) alternative distributions. For a fixed sample size and a significance level, powers for these two alternative distributions were exactly the same. As in the previous alternative distributions, the PML-test outperformed all other four exponentiality tests across all sample sizes and significance levels. The LF-test was in the last place on the power curve. The powers for the CVM-test and S-test were identical for a fixed sample size and a significance level. The D-test demonstrated the superior power than the CVM-test and the S-test for small sample sizes across all significance levels but this relationship was reversed for medium to large sample sizes. For sample sizes at least 200, the powers for all five tests were equivalents which were close to 1. As compare with the previous alternative distribution (Gamma (0.55, 0.412)), powers for these two alternative distributions decrease across all sample sizes and significance levels. It is relevant to note that the shape parameter (*k*) was changed from 0.412 to 0.50 which caused the decrease in power. It appears that as the value of the shape parameter (*k*) approaches that of the null distribution ($k = 1$), the simulated powers decreases.

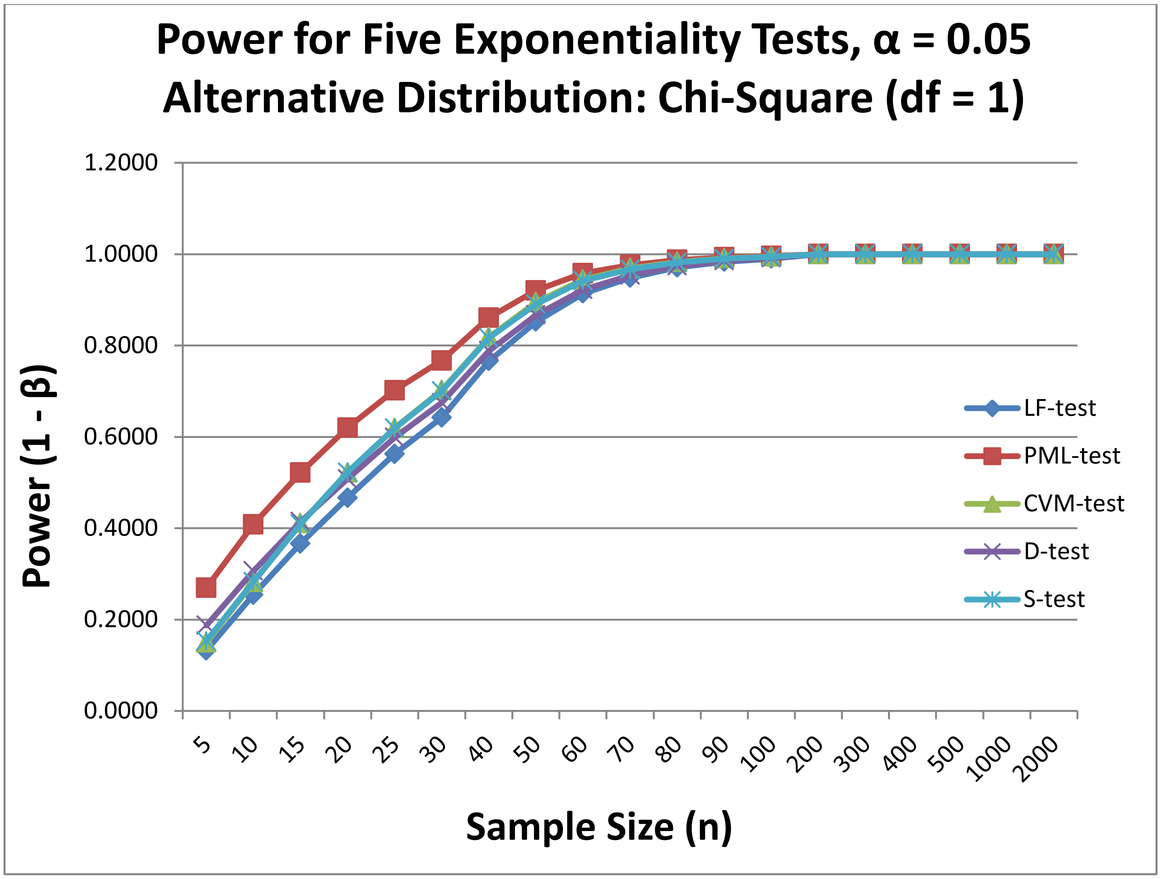

Figure 11. Power for Alternative Distribution: Chi-Square (1)

Seventh consider the relationship between the alternative distribution Gamma (4, 0.75) and the simulated power. Table B.8 and figure 12 summarize the power analysis for the Gamma (4, 0.75) alternative distribution. The PML-test outperformed all other four exponentiality tests across all sample sizes and significance levels. The LF-test was in the last place on the power curve. The powers for the CVM-test and S-test were identical for a fixed sample size and significance level. The D-test demonstrated the superior power than the CVM-test and the S-test for small sample sizes across all significance levels but this relationship was reversed for medium to large sample sizes. For sample size at least 1,000, the powers of all five tests were equivalents which were close to 1. As compare with the previous alternative distribution (Gamma (4, 0.5)), powers of this alternative distributions were significantly decrease across all sample sizes and significance levels. It is relevant to note that the shape parameter (*k*) was changed from 0.5 to 0.75 which caused the decrease in power. Among five Gamma alternative distributions discussed in this chapter, this alternative distribution exhibited the lowest power across all sample sizes and significance levels.

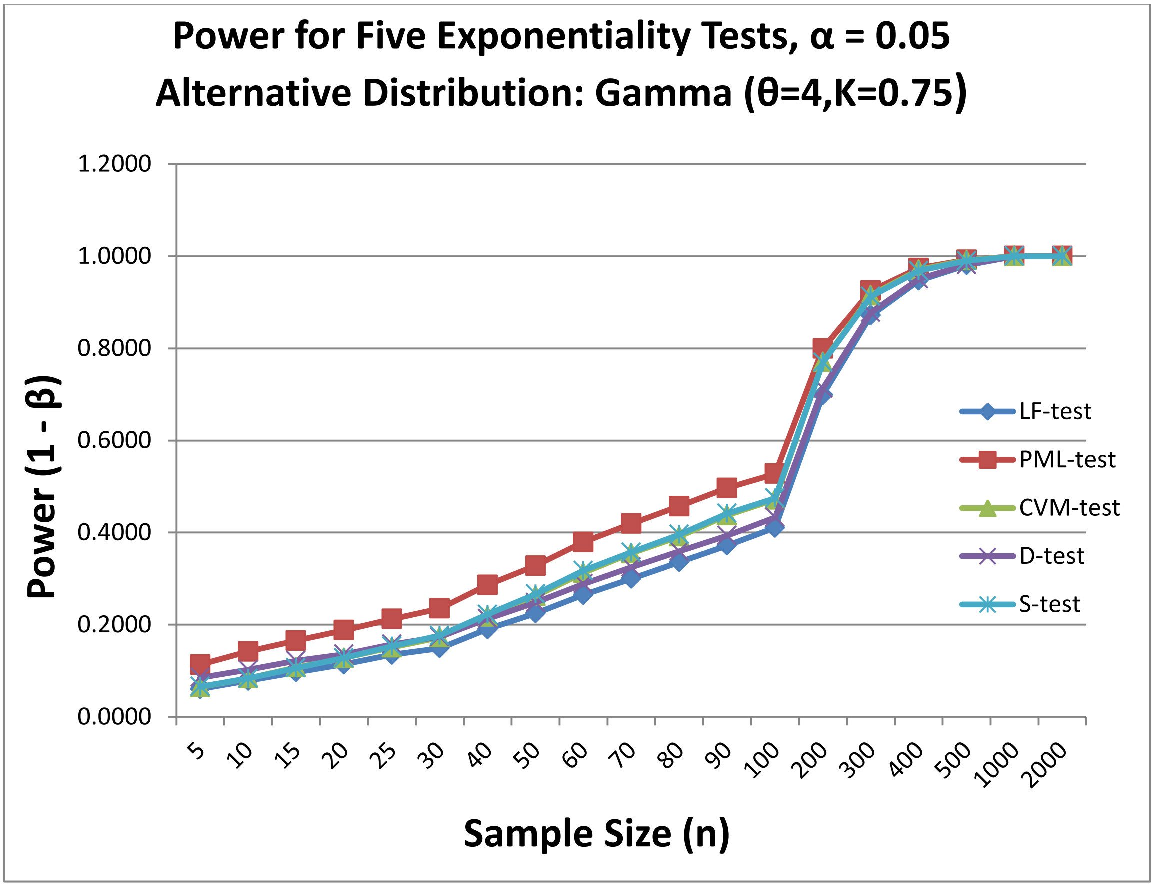

Figure 12. Power for Alternative Distribution: Gamma (4, 0.75)

Before considering the power for next two alternative distributions, it is indispensable to revisit that the Chi-Square distribution is a special case of Gamma distribution (64). This study previously showed that the power for the Gamma distribution depends only on the shape parameter (*k*). Null distributions were generated using the exponential (θ = 5) for power simulation. Using 64, Gamma (4, 1) and Chi-

Square (2) alternative distributions must produce similar powers for the set of parameters (*n* and α). In other words Gamma (4, 1) and Chi-Square (2) alternative distributions can be used for the simulation of significance levels.

Eighth consider the relationship between the alternative distributions, Gamma (4, 1), Chi-Square (2) and the simulated power. Table B.9 and figure 13 summarize the power analysis for the Gamma (4, 1) and Chi-Square (2) alternative distributions. The powers of all five exponentiality tests across all sample sizes and significance levels were too low which were pretty close to their significance levels. It is due to the fact that the power of these five exponentiality tests depends only on the shape parameter (*k*). It appears that the scale parameter (θ) does not have any role on the simulated powers.

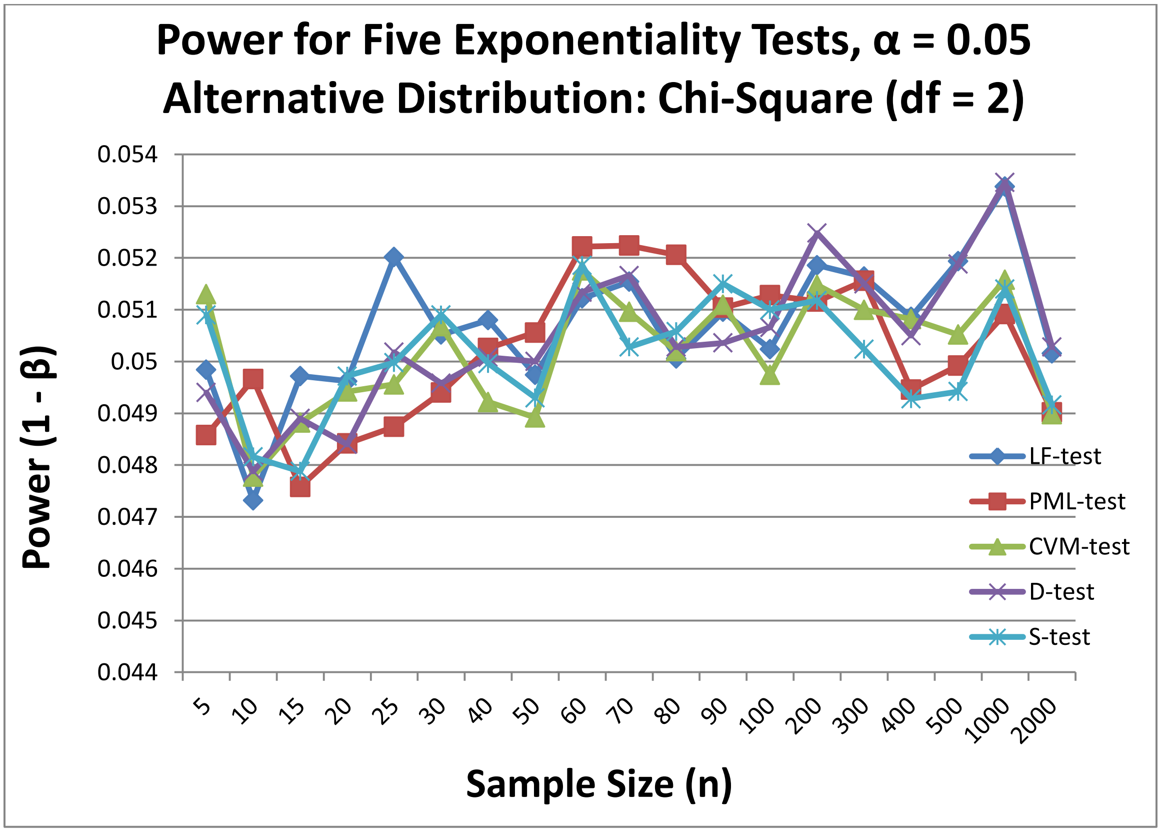

Figure 13. Power for Alternative Distribution: Chi-Square (2)

Ninth consider the relationship between the alternative distribution t (5) and the simulated power. Table B.10 and figure 14 summarize the power analysis for the t (5) alternative distribution. This is the only one symmetric distribution used in the power analyses. All five exponentiality tests quickly detected non-exponentiality. For sample sizes at least 15, the powers for all five tests were almost identical which were close to 1. The range of the powers was found to be very narrow across all sample sizes for a fixed significance level.

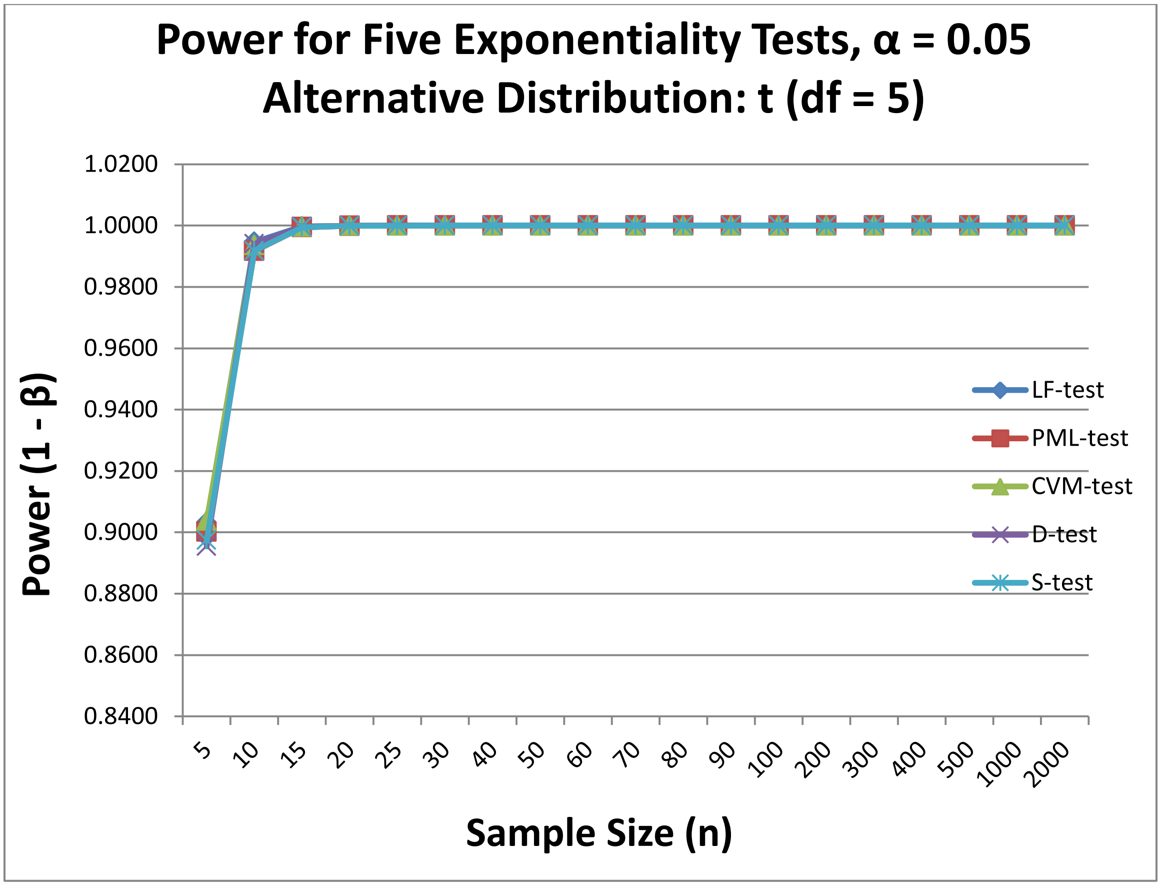

Figure 14. Power for Alternative Distribution: t (5)

Finally consider the relationship between the alternative distribution log-normal (0, 1) and the simulated power. Table B.11 and figure 15 summarize the power analysis for the log-normal (0, 1) alternative distribution. For small sample sizes, all five exponentiality tests demonstrated similar power across all significance levels. For medium to large sample sizes, the PML-test and S-test were in the top, the CVM-test was in the middle and the D-test and LF-test were in the bottom of the power curve. It appears that the PML-test exhibited equal or better power among five exponentiality tests in the set of parameters considered in this study. For sample sizes at least 1000, the powers for all five tests were almost identical which were close to 1.

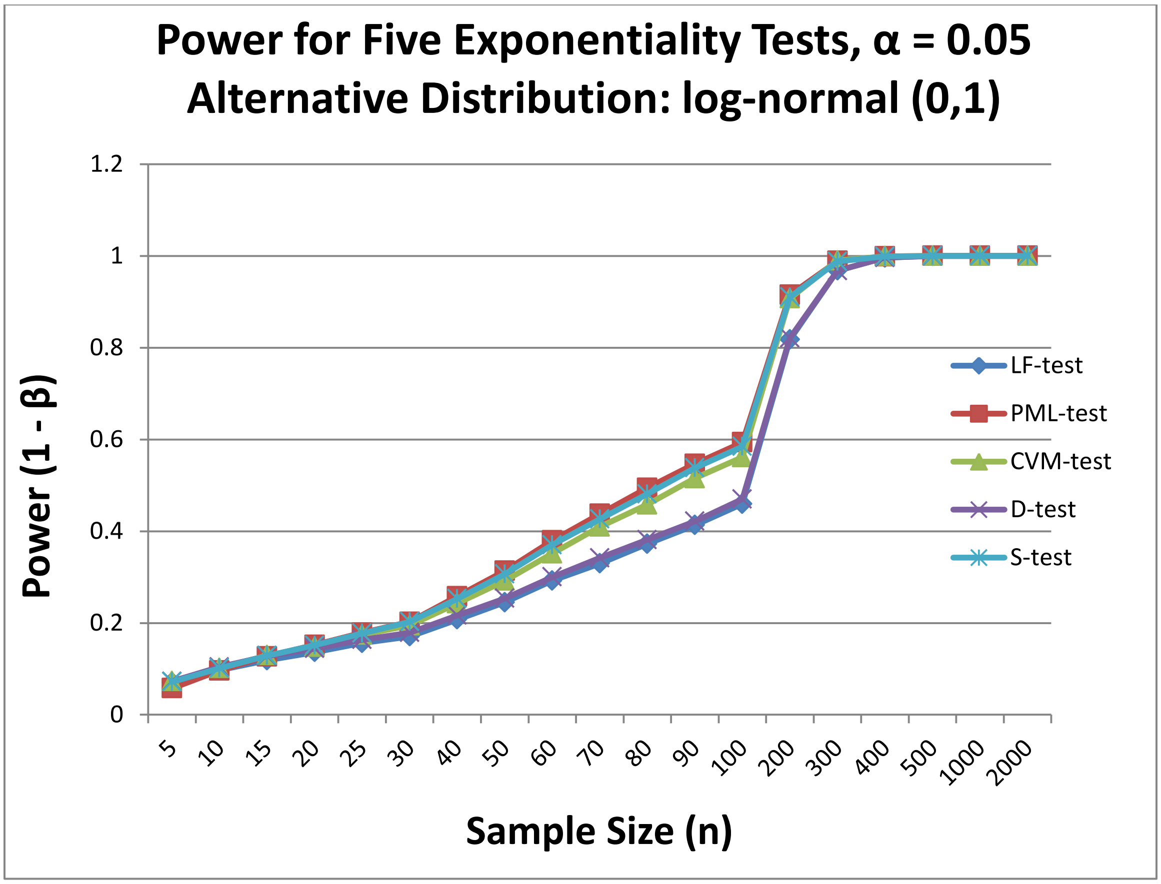

Figure 15. Power for Alternative Distribution: log-normal (0, 1)

From the above discussions, this study claimed that the PML-test demonstrated consistently superior power over the S-test, LF-test, CVM-test, and D-test for most of the alternative distributions presented in this study. The D-test, CVM-test, and S-test exhibited similar power for a fixed sample size and a significance level. The LF-test consistently showed the lowest power among five exponentiality tests. So, practically speaking the proposed test can hope to replace the other four exponentiality tests discussed throughout this study while maintaining a very simple form for computation and easy to understand for those people who have limited knowledge of statistics.

CHAPTER V

DISCUSSION

This chapter summarizes the research findings and the recommendations for future research.

Research Findings

The purpose of this study was to develop a new Goodness-of-Fit Test (GOFT) of exponentiality and compare it with four other existing GOFTs in terms of computation and performance. Using data simulation techniques, critical values for a new test were developed for three specific significance levels. The accuracy of the intended significance levels were verified for the new developed test (PML-test) and compared them with the Lilliefors test (LF-test), Cramer-Von Mises test (CVM-test), Finkelstein & Schafers test (S-test) and $\widetilde{D}_{n}$ test (D-test developed by Schafer et al. (1972)). The power comparisons among these five exponentiality tests were done using 11 right skewed and one symmetric alternative distribution. These results were presented in tables and figures and thoroughly discussed.

The newly developed test was primarily a modification of the original exponentiality test developed by Lilliefors (1969). Lilliefors considered the maximum absolute differences between the sample empirical distribution function (EDF) and the exponential cumulative distribution function (CDF). The proposed test considered the sum of all the absolute differences between the CDF and EDF. By considering the sum of all the absolute differences rather than only a point difference of each observation, the proposed test would expect to be less affected by individual extreme (too low or too high) observations and capable of detecting smaller, but consistent, differences between the distributions. The proposed test statistic is not only easy to understand but also very simple and easy to compute.

The code for critical values was developed in R 3.0.2 which is included in appendix A. To develop critical values, data were generated from an exponential distribution with three different scale parameters (θ). Fifteen different sample sizes and three specific significance levels were considered for the development of critical values. For each sample size and significance level, 50,000 trials were run from an exponential distribution which generated 50,000 test statistics. The 50,000 test statistics were then arranged in the order from smallest to largest. The proposed test is a right tail test. The critical value for various significance levels are $50{,}000*(1-\alpha)^{th}$ ordered value of the simulated test statistics. So, this study used 99^{th}, 95^{th}, and 90^{th} percentile of the test statistics as the critical values for the given sample size for the 0.01, 0.05, and 0.10 significance levels respectively.

To verify the accuracy of the three intended significance levels (α = 0.01, 0.05, and 0.10), data were generated from exponential distributions (null distribution: exponential ($\theta = 5$) and alternative distribution: exponential ($\theta = 10$)). For sample sizes 5, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 1000, and 2000; 50,000 trials were performed and the null hypothesis of data came from an exponential distribution was tested by five exponentiality tests. The results showed that all five tests of exponentiality worked very well in terms of controlling the intended significance levels. The study found that the proposed test performs very closely to other four tests of exponentiality in terms of the accuracy of the intended significance levels (for each sample size and overall averages across the 19 different sample sizes).

To compare the power (1 – β) of the proposed test and the other four exponentiality tests, this study utilized three different significance levels (0.01, 0.05, and 0.10) and 50,000 replications were drawn from each sample size (*n* = 5, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 1000, and 2000). A total of 12 alternative distributions were utilized (11 right skewed distribution and one symmetric distribution). This **s**tudy showed that the PML-test demonstrated consistently superior power over the S-test, LF-test, CVM-test, and D-test for most of the alternative distributions presented in this study. The D-test, CVM-test, and S-test exhibited similar power for a fixed sample size and a significance level. The LF-test consistently showed the lowest power among five exponentiality tests. So, practically speaking the proposed test can hope to replace the other four exponentiality tests discussed throughout this study while maintaining a very simple form for computation and easy to understand for those people who have limited knowledge of statistics. The proposed exponentiality test did successfully improve upon the power of the test it directly modified (i.e. LF-test). The actual method employed in the development of the test statistic in this study achieved its primary goal of improving the power of the LF-test of exponentiality.

Recommendations for Future Research

This study has presented a more powerful test of exponentiality which is not only easy to compute but also easy to understand. This study has shown that using the sum of all the absolute differences between the two functions (CDF and EDF) will have more power than just using the maximum differences between these two functions (like LF-test) or using the sum of squared differences between these two functions (like Cramer-Von Mises type test). The research presented here has the potential to modify many other tests and / or to develop tests for distributional assumption. The concept of sum of all the absolute differences between the EDF and CDF can be used to test if the data came from some distributions such as Beta distribution, Snedecor's *F* distribution, Pareto distribution, Weibull distribution, Gamma distribution, etc.

This study used data simulation techniques to generate random samples of several null and alternative distributions based on some specific parameter settings. Some outliers can be incorporated into these simulated data to see how the proposed test and other exponentiality tests mentioned throughout this study (possibly many more tests) work on this new situation.

This study focused on the comparative powers among five exponentiality tests. Although this study used specific sample sizes and significance levels to study power, this study did not focus what sample size(s) would be appropriate to achieve the desired power for desired significance levels. This research question can be addressed by continuing the present work.

Instead of using the supremum of the EDF and CDF, the average differences between these two functions can be evaluated for testing exponentiality. According to Bluman (2012), the median is affected less than mean by extremely high or extremely low values. So, the median differences between the EDF and CDF can also be used to test exponentiality. The latter would be a better test of exponentiality if there are potential outliers or influential observations in the data.

Multivariate analysis (MVA) is one of the demanding fields in statistics which involves observations and analysis of more than one outcome variables at a time. Many MVA procedures assume multivariate exponentiality (MVE). There are not many effective tests for MVE. Researchers usually see if individual dependent variables have a univariate exponential distribution. Adding extra dimensions on the dependent variables and considering the sum of all the absolute differences between the EDF and CDF with constitutes the natural extension of this current study.

This study used only the continuous distributions for power study. Some discrete distributions can be used as alternative distributions and can see how these five exponentiality tests (may be more) work to detect the departure from exponentiality.

REFERENCES

Acosta, P., & Rojas, G. M. (2009). A Simple IM Test for Exponential Distributions. *Applied Economics Letters*, 16(2), 109-112. doi: 10.1080/13504850601018221

Agresti, A. (1996). *An Introduction to Categorical Data Analysis*. NY.: John Wiley and Sons, Inc.

Anderson, T. W. & Darling, D. A. (1952). Asymptotic Theory of Certain Goodness-of-Fit Criteria Based on Stochastic Processes. *The Annals of Mathematical Statistics,* 23(2), 193-212.

Anderson, T. W. & Darling, D. A. (1954). A Test of Goodness of Fit. *Journal of the American Statistical Association,* 49(268), 765-769.

Arshad, M., Rasol, M. T., & Ahmad, M. I. (2003). Anderson Darling and Modified Anderson Darling Tests for Generalized Pareto Distribution. *Pakistan Journal of Applied Science,* 23(2), 85-88.

Bain, L. & Engelhardt, M. (1992). *Introduction to Probability and Mathematical Statistics.* (2 nd edition). M.A.: PWS-KENT publishing Company.

Baklizi, A. (2005). Weighted Kolmogorov-Smirnov Type Tests for Grouped Rayleigh Data. *Applied Mathematical Modeling*, 30, 437-445. doi: 10.1016/j.apm.2005.05.012

Baratpour, S. & Rad, A. H. (2012). Testing Goodness-of-Fit for Exponential Distribution Based on Cumulative Residual Entropy. *Communication in Statistics – Theory and Methods,* 41(8), 1387-1396. doi: 10.1080/03610926.2010.542857

Bluman, A. G. (2012). *Elementary Statistics: a Step by Step Approach* (8 th edition). N.Y.: McGraw-Hill Companies, Inc.

Breton, M. D., Devore, M. D. & Brown, D. E. (2008). A Tool for Systematically Comparing Power of Tests for Normality. *Journal of Statistical Computing and Simulation,* 78(7), 623-638. doi: 10.1080/00949650701377984

Browne, M. W. & Cudeck, R. (1989). Single Sample Cross-validation Indices in Covariance Structures. *Multivariate Behavioral Research*, 24, 445-455.

Chen, H. (2008). *Analysis of Variance Tests for Exponentiality of two Distributions: Complete and Censored Samples* (Master's Thesis). Retrieved from Library and Archives Canada.

Choi, B., Kim, K. & Song, S. H. (2004). Goodness-of-Fit Test for Exponentiality Based on Kullback-Leibler Information. *Communications in Statistics-Simulation and Computation*, 33(2), 525-536. doi: 10.1081/SAC-120037250

Cochran, W. G. (1952). The χ2 Test of Goodness of Fit. *Annals of Mathematical Statistics*, 23, 315-345.

Coin, D. (2008). Testing Normality in the Presence of Outliers. *Statistical Methods and Applications*, 17, 3-12. doi: 10.1007/s10260-007-0046-8

Conover, W. J. (1972). A Kolmogorov Goodness-of-Fit Test for Discontinuous Distribution. *Journal of the American Statistical Association*, 67(339), 591-596. doi: 10.1080/01621459.1972.10481254

Conover, W. J. (1999). *Practical Nonparametric Statistics* (3 rd edition). N.Y.: John Wiley and Sons, Inc.

Dahiya, R. C. & Gurland, J. (1972). Goodness of Fit Tests for the Gamma and Exponential Distributions. *Technometrics*, 14(3), 791-801.

Dallal, G. E. & Wilkinson, L. (1986). An Analytic Approximation to the Distribution of Lilliefors's Test. *The American Statistician*, 40(4), 294-296. doi: 10.1080/00031305.1986.10475419

D'Agostino, R. B. (1971). An Omnibus Test of Normality for Moderate and Large Size Samples. *Biometrika*, 8(2), 341-348.

Drezner, Z., Tuyrel, O. & Zerom, D. (2009). *A Modified Kolmogorov-Smirnov Test for Normality* (MPRA Report 14385). Retrieved from California State University-Fullerton, Steven G. Mihaylo College of Business and Economics website: http://mpra.ub.uni-muenchen.de/14385/

Dufour, J., Farhat, A., & Gardiol, L. (1998). Empirical Model Selection in Generalized Linear Mixed Effects Models. *Journal of Computational Statistics*, 23, 99-109.

Evans, D. L., Drew, J. H. & Leemis, L. M. (2008). The Distribution of the Kolmogorov-Smirnov, Cramer-Von Mises, and Anderson-Darling Test Statistics for Exponential Populations with Estimated Parameters. *Communications in Statistics-Simulation and Computation*, 31, 1396-1421. doi: 10.1080/03610910801983160

Farrell, P. J. & Rogers-Steward, K. (2006). Comprehensive Study of Tests for Normality and Symmetry. *Journal of Statistical Computing and Simulation,* 76(9), 803-816. doi: 10.1080/10629360500109023

Finkelstein, J. M. & Schafer, R. E. (1971). Improved Goodness-Of-Fit Tests. *Biometrika,* 58(3), 641-645.

Gail, M. H. & Gastwirth, J. L. (1978). A Scale-Free Goodness-of-Fit Test for the Exponential Distribution Based on the Gini Statistic. *Journal of the Royal Statistical Society,* 40(3), 350-357.

Henderson, A. R. (2006). Testing Experimental Data for Univariate Normality. *Clinica Chimica Acta,* 366, 112-129.

Hutchinson, S. R. & Bandalos, D. (1997). A Guide to Monte Carlo Simulation Research for Applied Researchers. *Journal of Vocational Education Research,* 22(4), 233-245.

Islam, T. U. (2011). Normality Testing – A New Direction. *International Journal of Business and Social Science,* 2(3), 115-118.

Klar, B. (2001). Goodness-of-Fit Tests for the Exponential and the Normal Distribution Based on the Integrated Distribution Function. *Annals of the Institute of Statistical Mathematics,* 53(2), 338-353.

Koehler, K. & Larntz, K. (1980). An Empirical Investigation of Goodness-of-Fit Statistics for Sparse Multinomials. *Journal of the American Statistical Association,* 75(370), 336-344.

Kuehl, R. (2000). *Design of Experiments: Statistical Principles of Research Design and Analysis* (2 nd edition). California: Duxbury.

Lilliefors, H. W. (1967). On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown. *Journal of the American Statistical Association,* 62(318), 399-402. doi: 10.1080/01621459.1967.10482916

Lilliefors, H. W. (1969). On the Kolmogorov-Smirnov Test for the Exponential Distribution with Mean Unknown. *Journal of the American Statistical Association,* 64(325), 387-389.

Liu, M. (2012). Power and Sample Size for Some Chi-Square Goodness of Fit Tests (Master’s thesis). Available from ProQuest Dissertations and Theses database. (UMI No. 1514113)

MacCallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model Modifications in Covariance Structure Analysis: The Problem of Capitalization on Chance. *Psychological bulletin*, 111, 490-504.

Margolin, B. H. & Maurer, W. (1976). Tests of the Kolmogorov-Smirnov Type for Exponential Data with Unknown Scale, and Related Problems. *Biometrika*, 63(1), 149-160.

Mendes, M. & Pala, A. (2003). Type I Error Rate and Power of Three Normality Tests. *Pakistan Journal of Information and Technology*, 2(2), 135-139.

Montgomery, D. C., Peck, E. A. & Vining, G. G. (2006). *Introduction to Linear Regression Analysis* (4 th edition). N.J.: John Wiley and Sons, Inc.

Morris, K. W. & Szynal, D. (2013). Goodness-of-Fit Tests Via Characterization Special Cases. *Journal of Mathematical Science*, 191(4), 550-567.

Overholt, B. A. & Schaffer, J. (2013). *A Modified Lilliefors Normality Test* (Doctoral dissertation). University of Northern Colorado, Greeley, Colorado.

Oztuna, D., Elhan, A. H., & Tuccar, E. (2006). Investigation of Four Different Normality Tests in Terms of Type I Error Rate and Power Under Different Distributions. *Turkish Journal of Medical Sciences*, 36(3), 171-176.

Pearson, K. (1900). On the Criterion that a Given System of Deviations from the Probable in the Case of a Correlated System of Variables is Such that It Can be Reasonably Supposed to Have Arisen from Random Sampling. *Philosophical Magazine*, 50(5), 157-175.

Pettitt, A. N. (1977). Tests for the Exponential Distribution with Censored Data Using Cramer-Von Mises Statistics. *Biometrika*, 64(3), 629-632.

Pettitt, A. N. & Stephens, M. A. (1977). The Kolmogorov-Smirnov Goodness-of-Fit Statistics with Discrete and Grouped Data. *Technometrics*, 19(2), 205-210.

Pugh, E. L. (1963). The Best Estimate of Reliability in the Exponential Case. *Operations Research*, 11(1), 57-61.

Razali, N. M. & Wah, Y. B. (2011). Power Comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling Tests. *Journal of Statistical Modeling and Analysis*, 2(1), 21-33.

Robey, R. R. & Barcikowski, R. S. (1992). TType I Error and the Number of Iterations in Monte Carlo Studies of Robustness. *British Journal of mathematical and Statistical Psychology*, 45, 283-288.

Rogozhnikov, A. P. & Lemeshko, B. Y. (2012). A Review of Tests for Exponentiality. *Proceedings of the International Conference,* Russia, 11, 159-166.

Schafer, R. E., Finkelstein, J. M. & Collins, J. (1972). On a Goodness-of-Fit Test for the Exponential Distribution with Mean Unknown. *Biometrika,* 59(1), 222-224.

Schaffer, J. R., & Kim, M. (2007). Number of Replications Required in Control Chart Simulation Studies. *Communication in Statistics-Simulation and Computation,* 36(5), 1075-1087. doi: 10.1080/03610910701539963

Schoder, V., Himmelmann, A. & Sim, K. P. (2006). Preliminary Testing for Normality: Some Statistical Aspects of a Common Concept. *Journal of Clinical and Experimental Dermatology,* 31, 757-761. doi: 10.1111/j.1365-2230.2006.02206.x

Scott, W. F. & Stewart, B. (2011). Tables for the Lilliefors and Modified Cramer_Von Mises Tests of Normality. *Communications in Statistics-Tgheory and methods,* 40(4), 726-730. doi: 10.1080/03610920903453467

Seier, E. (2002). Comparison of Tests for Univariate Normality. *InterStat Statistical Journal*, 1, 1-17.

Seshadri, V., Csorgo, M. & Stephens, M. A. (1969). Tests for the Exponential Distribution Using Kolmogorov-Type Statistics. *Journal of the Royal Statistical Society,* 31(3), 499-509.

Shapiro, S. S., & Wilk, M. B. (1965). An Analysis of Variance Test for Normality (Complete Samples). *Biometrika*, 52(3/4), 591-611.

Shapiro, S. S., Wilk, M. B. & Chen, H. J. (1968). A Comparative Study of Various Tests for Normality. *Journal of the American Statistical Association*, 63(324), 1343-1372. doi: 10.1080/01621459.1968.10480932

Shapiro, S. S. & Francia, R. S. (1972). An Approximate Analysis of Variance Test for Normality. *Journal of the American Statistical Association*, 67(337), 215-216. doi: 10.1080/01621459.1972.10481232

Shaw, A. R. (1994). *The Development of a Nonparametric Horizontal Distance Test of whether Two Independent Samples of Equal Size may have been Drawn from the Same Population* (Doctoral dissertation). University of Northern Colorado, Greeley, Colorado.

Spinelli, J. J. (2001). Testing Fit for the Grouped Exponential Distribution. *The Canadian Journal of Statistics*, 29(3), 451-458.

Spinelli, J. J. & Stephens, M. A. (1987). Tests for Exponentiality When Origin and Scale Parameters are Unknown. *Technometrics,* 29(4), 471-476.

Sprent, P. (1989). Applied Nonparametric Statistical Methods. London: Chapman and Hall.

Stephens, M. A. (1974). EDF Statistics for Goodness of Fit and Some Comparisons. *Journal of the American Statistical Association,* 69(347), 730-737. doi: 10.1080/01621459.1974.10480196

Thongteeraparp, A. & Chodjuntug, K. (2011). A power of Comparison of Goodness of Fit Tests for Exponential Distribution with Grouped Data. *Thailand Statistician,* 9(1), 37-49.

Wang, B. (2008). Goodness-of-fit Test for the Exponential Distribution Based on Pregressively Type-II Censored Sample. *Journal of Statistical Computation and Simulation,* 78(2), 125-132. doi: 10.1080/10629360600944266

Wilcox, R. R. (1988). Simulation as a Research Technique. In J. P. (Ed.), *Education Research, Methodology, and Measurement: An International Handbook* (pp 134-137). New York: Pergamon.

Xiao, Y., Gordon, A., & Yakovlev, A. (2007). A C++ Program for the Cramer-Von Mises Two-sample Test. *Journal of Statistical Software,* 17(8), 1-15.

Yan, X. (2009). *A Robust Likelihood Ratio Test: Testing Equal Means in the Presence of Unequal Variance.* Technical report, College of Economics and International Business, Experimental Statistics Program, New Mexico State University, Las Cruces, New Mexico.

Yap, B. W. & Sim, C. H. (2011). Comparisons of Various Types of Normality Tests. *Journal of Statistical Computing and Simulation,* 81(12), 2141-2155. doi: 10.1080/00949655.2010.520163

Yarnold, J. K. (1970). The Minimum Expectation in χ^2 Goodness of Fit Test and the Accuracy of the Null Distribution. *Journal of the American Statistical Association,* 65(330), 864-886. doi: 10.1080/01621459.1970.10481131

APPENDIX A

Code

A.1: R Code for Critical Values Table

For: Scale Parameter ($\theta = 10$)

```
set.seed (12345)
CV_One_Percent  = numeric(0)
CV_Five_Percent  = numeric(0)
CV_Ten_Percent   = numeric(0)
E = numeric(0)
for (k in 4:50){
n = k
for    (i in 1:50000){
y      = (rexp(n,10))
sn1  = (1:n)/n
Fx1 = sort(1-exp(-y/mean(y)))
KS1= abs(sn1-Fx1)
E[i] = sum(KS1)}
D= sort(E)
CV_One_Percent[k]    = D[49500]
CV_Five_Percent[k]    = D[47500]
CV_Ten_Percent[k]     = D[45000]}
Table_Of_CV = cbind(CV_One_Percent,CV_Five_Percent,CV_Ten_Percent)
```

A.2: R Code for Significance Level Comparisons

For: Sample Size (n = 500)

```
n    = 500
set.seed (12345)
sn1  = (1:n)/n
sn2  = ((1:n)-1)/n
sn3  = ((1:n)-0.5)/n
sn4  = (((1:n)-0.5)-1)/n
En1 = numeric(0)
En2 = numeric(0)
En3 = numeric(0)
En4 = numeric(0)
En5 = numeric(0)
for (i in 1:50000){
y        = (rexp(n,5))
Fx1      = sort(1-exp(-y/mean(y)))
Fx1D  = sort(1-(1-(y/(n*mean(y))))^(n-1))
c1        = 0.08333/n
KS1      = abs(Fx1-sn1)
KS2      = abs(sn2-Fx1)
KS3      = abs(Fx1-sn3)
KS4      = abs(sn4-Fx1)
KS1D  = abs(Fx1D-sn1)
KS2D  = abs(sn2-Fx1D)
ind1      = which(KS1 < KS2)
ind1D  = which(KS1D < KS2D)
ind2      = which(KS3 < KS4)
e1        = KS1
e1D      = KS1D
e2        = KS3
e1[ind1]   <- KS2[ind1]
e1D[ind1D] <- KS2D[ind1D]
e2[ind2]   <- KS4[ind2]
Fct1       = cbind(KS1,KS2)
Fct2       = cbind(Fct1,max=apply(Fct1,1,max))
Fct3       = subset(Fct2,select=c(3))
Fct4       = sum(Fct3)
En1[i]     = max(KS1,KS2)
En2[i]     = sum(KS3*KS3)+c1
En3[i]     = sum(KS1)
En4[i]     = max(e1D)
En5[i]     = Fct4 }
LF     = numeric(0)
CVM = numeric(0)
PML  = numeric(0)
```

```
D       = numeric(0)
S       = numeric(0)
sn11   = (1:n)/n
sn22   = ((1:n)-1)/n
sn33   = ((1:n)-0.5)/n
sn44   = (((1:n)-0.5)-1)/n
for (i in 1:50000){
x           = (rexp(n,10))
Fx2        = sort(1-exp(-x/mean(x)))
Fx2D      = sort(1-(1-(x/(n*mean(x))))^(n-1))
c2          = 0.08333/n
KS11      = abs(Fx2-sn11)
KS22      = abs(sn22-Fx2)
KS33      = abs(Fx2-sn33)
KS44      = abs(sn44-Fx2)
KS11D   = abs(Fx2D-sn11)
KS22D   = abs(sn22-Fx2D)
ind11      = which(KS11 < KS22)
ind11D   = which(KS11D< KS22D)
ind22      = which(KS33 < KS44)
e11         = KS11
e11D      = KS11D
e22         = KS33
e11[ind11]    <- KS22[ind11]
e11D[ind11D]  <- KS22D[ind11D]
e22[ind22]    <- KS44[ind22]
Fct5       = cbind(KS11,KS22)
Fct6       = cbind(Fct5,max=apply(Fct5,1,max))
Fct7       = subset(Fct6,select=c(3))
Fct8       = sum(Fct7)
stat1       = max(KS11,KS22)
stat2       = sum(KS33*KS33)+c2
stat3       = sum(KS11)
stat4       = max(e11D)
stat5       = Fct8
LF[i]      = sum(En1 > stat1)/50000
CVM[i]  = sum(En2 > stat2)/50000
PML[i]   = sum(En3 > stat3)/50000
D[i]        = sum(En4 > stat4)/50000
S[i]         = sum(En5 > stat5)/50000 }
LF1PCT  = 50000 - sum(LF > 0.01,na.rm=TRUE)
LF5PCT  = 50000 - sum(LF > 0.05,na.rm=TRUE)
LF10PCT = 50000 - sum(LF > 0.10,na.rm=TRUE)
LFn = cbind(LF1PCT/50000,LF5PCT/50000,LF10PCT/50000)
CVM1PCT  = 50000 - sum(CVM > 0.01,na.rm=TRUE)
CVM5PCT  = 50000 - sum(CVM > 0.05,na.rm=TRUE)
```

```
CVM10PCT = 50000 - sum(CVM > 0.10,na.rm=TRUE)
CVMn = cbind(CVM1PCT/50000,CVM5PCT/50000,CVM10PCT/50000)
PML1PCT  = 50000 - sum(PML > 0.01,na.rm=TRUE)
PML5PCT  = 50000 - sum(PML > 0.05,na.rm=TRUE)
PML10PCT = 50000 - sum(PML > 0.10,na.rm=TRUE)
PMLn = cbind(PML1PCT/50000,PML5PCT/50000,PML10PCT/50000)
D1PCT  = 50000 - sum(D > 0.01,na.rm=TRUE)
D5PCT  = 50000 - sum(D > 0.05,na.rm=TRUE)
D10PCT = 50000 - sum(D > 0.10,na.rm=TRUE)
Dn = cbind(D1PCT/50000,D5PCT/50000,D10PCT/50000)
S1PCT  = 50000 - sum(S > 0.01,na.rm=TRUE)
S5PCT  = 50000 - sum(S > 0.05,na.rm=TRUE)
S10PCT = 50000 - sum(S > 0.10,na.rm=TRUE)
Sn = cbind(S1PCT/50000,S5PCT/50000,S10PCT/50000)
One_Pct     = cbind (LF1PCT/50000,  D1PCT/50000,  CVM1PCT/50000,  S1PCT/50000,
PML1PCT/50000)
Five_Pct    = cbind (LF5PCT/50000,  D5PCT/50000,  CVM5PCT/50000,  S5PCT/50000,
PML5PCT/50000)
Ten_Pct     = cbind (LF10PCT/50000, D10PCT/50000, CVM10PCT/50000,
S10PCT/50000, PML10PCT/50000)
All_Output  = cbind (One_Pct,Five_Pct,Ten_Pct)
```

A.3: R Code for Power Analysis

For: Sample Size (*n* = 500) and Alternative Distribution: Gamma (0.55, 0.412)

```
n    = 500
set.seed (12345)
sn1  = (1:n)/n
sn2  = ((1:n)-1)/n
sn3  = ((1:n)-0.5)/n
sn4  = (((1:n)-0.5)-1)/n
En1 = numeric(0)
En2 = numeric(0)
En3 = numeric(0)
En4 = numeric(0)
En5 = numeric(0)
for (i in 1:50000){
y     = (rexp(n,5))
Fx1 = sort(1-exp(-y/mean(y)))
Fx1D = sort(1-(1-(y/(n*mean(y))))^(n-1))
c1      = 0.08333/n
KS1     = abs(Fx1-sn1)
KS2     = abs(sn2-Fx1)
KS3     = abs(Fx1-sn3)
KS4     = abs(sn4-Fx1)
KS1D  = abs(Fx1D-sn1)
KS2D  = abs(sn2-Fx1D)
ind1      = which(KS1 < KS2)
ind1D   = which(KS1D < KS2D)
ind2      = which(KS3 < KS4)
e1         = KS1
e1D      = KS1D
e2         = KS3
e1[ind1]   <- KS2[ind1]
e1D[ind1D] <- KS2D[ind1D]
e2[ind2]   <- KS4[ind2]
Fct1       = cbind(KS1,KS2)
Fct2       = cbind(Fct1,max=apply(Fct1,1,max))
Fct3       = subset(Fct2,select=c(3))
Fct4       = sum(Fct3)
En1[i]    = max(KS1,KS2)
En2[i]    = sum(KS3*KS3)+c1
En3[i]    = sum(KS1)
En4[i]    = max(e1D)
En5[i]    = Fct4 }
LF      = numeric(0)
CVM  = numeric(0)
PML   = numeric(0)
```

```
D        = numeric(0)
S        = numeric(0)
sn11   = (1:n)/n
sn22   = ((1:n)-1)/n
sn33   = ((1:n)-0.5)/n
sn44   = (((1:n)-0.5)-1)/n
for (i in 1:50000){
x           = (rgamma(n,0.412,0.55))
Fx2        = sort(1-exp(-x/mean(x)))
Fx2D     = sort(1-(1-(x/(n*mean(x))))^(n-1))
c2          = 0.08333/n
KS11     = abs(Fx2-sn11)
KS22     = abs(sn22-Fx2)
KS33     = abs(Fx2-sn33)
KS44     = abs(sn44-Fx2)
KS11D  = abs(Fx2D-sn11)
KS22D  = abs(sn22-Fx2D)
ind11     = which(KS11 < KS22)
ind11D  = which(KS11D< KS22D)
ind22     = which(KS33 < KS44)
e11        = KS11
e11D     = KS11D
e22        = KS33
e11[ind11]    <- KS22[ind11]
e11D[ind11D]  <- KS22D[ind11D]
e22[ind22]    <- KS44[ind22]
Fct5       = cbind(KS11,KS22)
Fct6       = cbind(Fct5,max=apply(Fct5,1,max))
Fct7       = subset(Fct6,select=c(3))
Fct8       = sum(Fct7)
stat1       = max(KS11,KS22)
stat2       = sum(KS33*KS33)+c2
stat3       = sum(KS11)
stat4       = max(e11D)
stat5       = Fct8
LF[i]      = sum(En1 > stat1)/50000
CVM[i]  = sum(En2 > stat2)/50000
PML[i]   = sum(En3 > stat3)/50000
D[i]        = sum(En4 > stat4)/50000
S[i]        = sum(En5 > stat5)/50000 }
LF1PCT  = 50000 - sum(LF > 0.01,na.rm=TRUE)
LF5PCT  = 50000 - sum(LF > 0.05,na.rm=TRUE)
LF10PCT = 50000 - sum(LF > 0.10,na.rm=TRUE)
LFn = cbind(LF1PCT/50000,LF5PCT/50000,LF10PCT/50000)
CVM1PCT  = 50000 - sum(CVM > 0.01,na.rm=TRUE)
CVM5PCT  = 50000 - sum(CVM > 0.05,na.rm=TRUE)
```

```
CVM10PCT = 50000 - sum(CVM > 0.10,na.rm=TRUE)
CVMn = cbind(CVM1PCT/50000,CVM5PCT/50000,CVM10PCT/50000)
PML1PCT  = 50000 - sum(PML > 0.01,na.rm=TRUE)
PML5PCT  = 50000 - sum(PML > 0.05,na.rm=TRUE)
PML10PCT = 50000 - sum(PML > 0.10,na.rm=TRUE)
PMLn = cbind(PML1PCT/50000,PML5PCT/50000,PML10PCT/50000)
D1PCT  = 50000 - sum(D > 0.01,na.rm=TRUE)
D5PCT  = 50000 - sum(D > 0.05,na.rm=TRUE)
D10PCT = 50000 - sum(D > 0.10,na.rm=TRUE)
Dn = cbind(D1PCT/50000,D5PCT/50000,D10PCT/50000)
S1PCT  = 50000 - sum(S > 0.01,na.rm=TRUE)
S5PCT  = 50000 - sum(S > 0.05,na.rm=TRUE)
S10PCT = 50000 - sum(S > 0.10,na.rm=TRUE)
Sn = cbind(S1PCT/50000,S5PCT/50000,S10PCT/50000)
One_Pct     = cbind (LF1PCT/50000,  D1PCT/50000,  CVM1PCT/50000,  S1PCT/50000,
PML1PCT/50000)
Five_Pct    = cbind (LF5PCT/50000,  D5PCT/50000,  CVM5PCT/50000,  S5PCT/50000,
PML5PCT/50000)
Ten_Pct     = cbind (LF10PCT/50000, D10PCT/50000, CVM10PCT/50000,
S10PCT/50000, PML10PCT/50000)
All_Output  = cbind (One_Pct,Five_Pct,Ten_Pct)
```

APPENDIX B

Power and Significance Level Analyses Tables and Figures

Table B.1
Simulated Significance Levels

	$\alpha = 0.01$					$\alpha = 0.05$					$\alpha = 0.10$				
n	LF	D	CVM	S	PML	LF	D	CVM	S	PML	LF	D	CVM	S	PML
5	0.010	0.010	0.010	0.010	0.011	0.049	0.051	0.051	0.051	0.052	0.098	0.100	0.101	0.101	0.099
10	0.010	0.011	0.011	0.010	0.010	0.051	0.051	0.051	0.051	0.053	0.101	0.101	0.102	0.103	0.104
15	0.010	0.010	0.010	0.010	0.011	0.050	0.050	0.050	0.050	0.051	0.101	0.100	0.099	0.100	0.100
20	0.010	0.010	0.010	0.010	0.010	0.050	0.049	0.050	0.050	0.050	0.099	0.100	0.100	0.099	0.100
25	0.010	0.010	0.010	0.010	0.010	0.051	0.051	0.050	0.050	0.049	0.100	0.099	0.098	0.098	0.101
30	0.009	0.009	0.009	0.008	0.009	0.049	0.049	0.049	0.050	0.049	0.097	0.098	0.099	0.101	0.099
40	0.010	0.011	0.010	0.010	0.009	0.050	0.049	0.049	0.049	0.050	0.099	0.101	0.098	0.099	0.098
50	0.010	0.010	0.011	0.011	0.011	0.050	0.050	0.051	0.051	0.050	0.100	0.099	0.099	0.098	0.098
60	0.010	0.010	0.011	0.010	0.010	0.052	0.052	0.052	0.053	0.052	0.102	0.103	0.103	0.103	0.103
70	0.011	0.010	0.010	0.010	0.010	0.050	0.050	0.051	0.051	0.050	0.100	0.100	0.101	0.101	0.101
80	0.010	0.010	0.010	0.010	0.010	0.051	0.051	0.051	0.051	0.051	0.100	0.101	0.100	0.102	0.103
90	0.011	0.011	0.011	0.012	0.011	0.051	0.050	0.052	0.052	0.052	0.100	0.099	0.102	0.102	0.104
100	0.010	0.010	0.011	0.011	0.011	0.050	0.050	0.050	0.052	0.052	0.098	0.098	0.101	0.102	0.101
200	0.010	0.010	0.010	0.010	0.010	0.050	0.051	0.051	0.052	0.052	0.099	0.100	0.102	0.101	0.101
300	0.011	0.011	0.011	0.011	0.011	0.052	0.052	0.052	0.051	0.052	0.100	0.100	0.102	0.102	0.102
400	0.011	0.011	0.011	0.011	0.010	0.050	0.050	0.052	0.051	0.051	0.102	0.101	0.100	0.099	0.100
500	0.010	0.010	0.010	0.010	0.010	0.050	0.050	0.049	0.049	0.049	0.100	0.100	0.100	0.100	0.100
1000	0.011	0.011	0.010	0.010	0.010	0.053	0.053	0.051	0.052	0.052	0.102	0.101	0.103	0.101	0.102
2000	0.010	0.010	0.011	0.011	0.011	0.051	0.051	0.049	0.049	0.049	0.102	0.102	0.101	0.101	0.101

Table B.2
Simulated Power for Alternative Distribution: Weibull (1, 0.5)

	α = 0.01					α = 0.05					α = 0.10				
n	LF	D	CVM	S	PML	LF	D	CVM	S	PML	LF	D	CVM	S	PML
5	0.171	0.221	0.190	0.189	0.279	0.276	0.354	0.310	0.317	0.466	0.338	0.452	0.383	0.396	0.566
10	0.386	0.433	0.452	0.449	0.539	0.568	0.626	0.613	0.616	0.731	0.657	0.713	0.697	0.700	0.806
15	0.585	0.626	0.654	0.651	0.733	0.750	0.786	0.797	0.796	0.863	0.822	0.852	0.852	0.854	0.909
20	0.734	0.759	0.800	0.799	0.848	0.865	0.885	0.901	0.900	0.936	0.913	0.929	0.934	0.934	0.961
25	0.835	0.858	0.891	0.889	0.919	0.931	0.942	0.953	0.952	0.971	0.958	0.966	0.972	0.972	0.983
30	0.901	0.913	0.941	0.939	0.957	0.965	0.971	0.979	0.978	0.987	0.981	0.985	0.988	0.988	0.993
40	0.969	0.974	0.985	0.985	0.990	0.992	0.994	0.996	0.996	0.998	0.996	0.997	0.998	0.998	0.999
50	0.991	0.992	0.997	0.996	0.998	0.998	0.998	0.999	0.999	0.999	0.999	0.999	1.000	1.000	1.000
60	0.998	0.998	0.999	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
70	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
80	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
90	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
100	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
200	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
300	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
400	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
1000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
2000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table B.3
Simulated Power for Alternative Distribution: Weibull (1, 0.75)

	α = 0.01					α = 0.05					α = 0.10				
n	LF	D	CVM	S	PML	LF	D	CVM	S	PML	LF	D	CVM	S	PML
5	0.033	0.051	0.036	0.036	0.067	0.087	0.129	0.094	0.095	0.176	0.137	0.192	0.146	0.149	0.257
10	0.058	0.077	0.072	0.072	0.115	0.146	0.190	0.163	0.164	0.261	0.221	0.275	0.238	0.241	0.360
15	0.089	0.111	0.109	0.109	0.167	0.201	0.243	0.232	0.231	0.330	0.289	0.340	0.317	0.319	0.434
20	0.120	0.141	0.153	0.155	0.216	0.257	0.299	0.299	0.302	0.401	0.357	0.404	0.394	0.396	0.506
25	0.152	0.181	0.198	0.201	0.264	0.320	0.358	0.364	0.368	0.463	0.424	0.467	0.465	0.469	0.572
30	0.186	0.210	0.240	0.243	0.310	0.366	0.405	0.425	0.431	0.518	0.475	0.519	0.526	0.529	0.624
40	0.259	0.289	0.332	0.339	0.405	0.461	0.494	0.528	0.536	0.615	0.574	0.612	0.632	0.635	0.711
50	0.335	0.364	0.427	0.439	0.501	0.550	0.579	0.625	0.630	0.698	0.664	0.693	0.719	0.722	0.785
60	0.411	0.438	0.513	0.522	0.582	0.634	0.660	0.705	0.711	0.766	0.735	0.759	0.788	0.789	0.840
70	0.482	0.505	0.591	0.598	0.654	0.699	0.724	0.774	0.778	0.824	0.792	0.811	0.844	0.846	0.883
80	0.550	0.575	0.653	0.659	0.706	0.753	0.772	0.820	0.825	0.864	0.834	0.850	0.881	0.885	0.913
90	0.613	0.636	0.717	0.725	0.769	0.801	0.817	0.863	0.866	0.895	0.871	0.884	0.912	0.913	0.935
100	0.669	0.690	0.772	0.779	0.815	0.842	0.855	0.894	0.897	0.920	0.901	0.911	0.935	0.936	0.951
200	0.950	0.954	0.981	0.981	0.984	0.989	0.990	0.995	0.995	0.996	0.995	0.996	0.998	0.998	0.998
300	0.995	0.996	0.999	0.999	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
400	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
1000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
2000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table B.4
Simulated Power for Alternative Distribution: Gamma (4, 0.25)

	α = 0.01					α = 0.05					α = 0.10				
n	LF	D	CVM	S	PML	LF	D	CVM	S	PML	LF	D	CVM	S	PML
5	0.286	0.340	0.312	0.303	0.412	0.395	0.469	0.445	0.457	0.604	0.453	0.605	0.521	0.537	0.699
10	0.575	0.615	0.630	0.618	0.696	0.741	0.781	0.777	0.771	0.854	0.813	0.849	0.841	0.838	0.907
15	0.788	0.813	0.829	0.818	0.868	0.897	0.915	0.921	0.916	0.949	0.940	0.951	0.951	0.948	0.972
20	0.902	0.914	0.929	0.921	0.945	0.964	0.970	0.974	0.971	0.983	0.980	0.984	0.985	0.984	0.992
25	0.959	0.965	0.973	0.969	0.979	0.988	0.990	0.992	0.990	0.995	0.994	0.995	0.996	0.995	0.998
30	0.983	0.985	0.990	0.988	0.992	0.996	0.996	0.998	0.997	0.998	0.998	0.999	0.999	0.999	0.999
40	0.998	0.998	0.999	0.999	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
50	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
60	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
70	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
80	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
90	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
100	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
200	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
300	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
400	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
1000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
2000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table B.5
Simulated Power for Alternative Distribution: Gamma (0.55, 0.275)

	α = 0.01					α = 0.05					α = 0.10				
n	LF	D	CVM	S	PML	LF	D	CVM	S	PML	LF	D	CVM	S	PML
5	0.246	0.299	0.268	0.262	0.366	0.352	0.429	0.400	0.412	0.560	0.410	0.558	0.475	0.492	0.660
10	0.503	0.547	0.559	0.549	0.633	0.681	0.727	0.721	0.717	0.812	0.761	0.805	0.795	0.793	0.875
15	0.716	0.746	0.765	0.752	0.817	0.852	0.874	0.883	0.877	0.924	0.908	0.926	0.925	0.921	0.956
20	0.852	0.867	0.891	0.882	0.914	0.939	0.950	0.956	0.952	0.971	0.965	0.971	0.974	0.972	0.984
25	0.926	0.937	0.951	0.945	0.963	0.976	0.980	0.985	0.982	0.989	0.988	0.990	0.992	0.991	0.995
30	0.964	0.969	0.978	0.974	0.982	0.990	0.992	0.994	0.993	0.996	0.995	0.996	0.997	0.997	0.998
40	0.994	0.995	0.997	0.996	0.998	0.999	0.999	0.999	0.999	1.000	1.000	1.000	1.000	1.000	1.000
50	0.999	0.999	1.000	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
60	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
70	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
80	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
90	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
100	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
200	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
300	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
400	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
1000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
2000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table B.6
Simulated Power for Alternative Distribution: Gamma (0.55, 0.412)

	α = 0.01					α = 0.05					α = 0.10				
n	LF	D	CVM	S	PML	LF	D	CVM	S	PML	LF	D	CVM	S	PML
5	0.106	0.144	0.117	0.115	0.186	0.190	0.257	0.217	0.223	0.362	0.246	0.354	0.287	0.299	0.465
10	0.215	0.255	0.262	0.259	0.339	0.386	0.445	0.425	0.427	0.555	0.483	0.548	0.524	0.525	0.662
15	0.353	0.393	0.409	0.402	0.498	0.546	0.593	0.598	0.596	0.698	0.653	0.699	0.691	0.690	0.787
20	0.472	0.506	0.545	0.538	0.616	0.675	0.711	0.727	0.723	0.801	0.769	0.800	0.805	0.801	0.869
25	0.583	0.620	0.662	0.658	0.722	0.779	0.804	0.822	0.818	0.870	0.852	0.872	0.881	0.878	0.921
30	0.682	0.706	0.755	0.745	0.797	0.845	0.863	0.883	0.879	0.916	0.901	0.916	0.926	0.924	0.951
40	0.828	0.844	0.881	0.875	0.903	0.933	0.941	0.954	0.952	0.968	0.963	0.969	0.975	0.973	0.982
50	0.912	0.922	0.947	0.944	0.957	0.972	0.975	0.984	0.982	0.988	0.986	0.988	0.991	0.991	0.994
60	0.958	0.963	0.978	0.974	0.981	0.990	0.991	0.995	0.994	0.996	0.996	0.996	0.998	0.997	0.998
70	0.981	0.983	0.991	0.989	0.992	0.995	0.996	0.998	0.998	0.999	0.998	0.999	0.999	0.999	0.999
80	0.992	0.993	0.996	0.995	0.996	0.999	0.999	0.999	0.999	0.999	1.000	1.000	1.000	1.000	1.000
90	0.997	0.997	0.999	0.998	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
100	0.999	0.999	1.000	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
200	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
300	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
400	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
1000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
2000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table B.7
Simulated Power for Alternative Distribution: Chi-Square (1)

	α = 0.01					α = 0.05					α = 0.10				
n	LF	D	CVM	S	PML	LF	D	CVM	S	PML	LF	D	CVM	S	PML
5	0.063	0.092	0.070	0.070	0.120	0.132	0.187	0.149	0.153	0.270	0.185	0.267	0.212	0.220	0.363
10	0.119	0.150	0.149	0.148	0.211	0.254	0.306	0.282	0.283	0.408	0.342	0.407	0.375	0.378	0.519
15	0.196	0.232	0.235	0.233	0.316	0.366	0.414	0.411	0.409	0.522	0.472	0.525	0.510	0.512	0.631
20	0.270	0.299	0.326	0.324	0.402	0.466	0.508	0.521	0.523	0.620	0.576	0.621	0.622	0.622	0.720
25	0.345	0.383	0.418	0.416	0.494	0.563	0.599	0.619	0.619	0.702	0.669	0.703	0.712	0.713	0.792
30	0.422	0.452	0.503	0.497	0.570	0.642	0.675	0.702	0.700	0.767	0.740	0.770	0.784	0.783	0.844
40	0.572	0.600	0.654	0.651	0.705	0.766	0.788	0.819	0.817	0.861	0.845	0.864	0.878	0.876	0.912
50	0.689	0.713	0.769	0.768	0.809	0.852	0.867	0.895	0.891	0.921	0.911	0.922	0.935	0.933	0.955
60	0.785	0.801	0.853	0.849	0.878	0.914	0.922	0.944	0.942	0.959	0.952	0.958	0.968	0.967	0.978
70	0.857	0.867	0.910	0.904	0.925	0.948	0.954	0.969	0.967	0.976	0.973	0.977	0.983	0.982	0.988
80	0.907	0.915	0.944	0.939	0.951	0.971	0.974	0.983	0.982	0.988	0.985	0.987	0.991	0.991	0.994
90	0.939	0.945	0.966	0.964	0.973	0.983	0.985	0.991	0.990	0.993	0.993	0.993	0.996	0.995	0.997
100	0.963	0.966	0.981	0.979	0.984	0.990	0.992	0.995	0.994	0.996	0.996	0.997	0.998	0.997	0.998
200	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
300	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
400	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
1000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
2000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table B.8
Simulated Power for Alternative Distribution: Gamma (4, 0.75)

	α = 0.01					α = 0.05					α = 0.10				
n	LF	D	CVM	S	PML	LF	D	CVM	S	PML	LF	D	CVM	S	PML
5	0.019	0.028	0.020	0.020	0.036	0.061	0.085	0.065	0.066	0.113	0.105	0.142	0.112	0.114	0.183
10	0.024	0.034	0.028	0.028	0.046	0.078	0.102	0.084	0.084	0.142	0.136	0.168	0.143	0.144	0.221
15	0.033	0.043	0.036	0.036	0.062	0.097	0.122	0.107	0.106	0.166	0.163	0.194	0.172	0.173	0.251
20	0.039	0.048	0.046	0.046	0.073	0.114	0.136	0.128	0.128	0.188	0.185	0.218	0.199	0.200	0.280
25	0.045	0.057	0.056	0.057	0.085	0.135	0.157	0.151	0.153	0.212	0.210	0.240	0.229	0.232	0.309
30	0.052	0.062	0.067	0.066	0.096	0.149	0.174	0.172	0.176	0.235	0.234	0.268	0.256	0.259	0.338
40	0.072	0.085	0.091	0.093	0.125	0.190	0.213	0.217	0.222	0.286	0.281	0.315	0.313	0.316	0.389
50	0.090	0.106	0.120	0.125	0.160	0.225	0.248	0.262	0.267	0.328	0.325	0.354	0.363	0.366	0.437
60	0.112	0.126	0.147	0.150	0.190	0.265	0.289	0.313	0.317	0.379	0.375	0.403	0.420	0.422	0.491
70	0.136	0.148	0.176	0.180	0.219	0.300	0.325	0.355	0.357	0.419	0.414	0.442	0.466	0.470	0.533
80	0.160	0.175	0.203	0.205	0.242	0.336	0.359	0.391	0.396	0.457	0.453	0.480	0.506	0.511	0.574
90	0.184	0.202	0.236	0.242	0.285	0.372	0.393	0.438	0.442	0.497	0.493	0.517	0.547	0.551	0.608
100	0.207	0.223	0.267	0.273	0.315	0.411	0.433	0.472	0.475	0.528	0.529	0.551	0.587	0.589	0.643
200	0.465	0.481	0.564	0.561	0.594	0.699	0.713	0.770	0.769	0.800	0.797	0.809	0.849	0.849	0.870
300	0.691	0.704	0.786	0.783	0.805	0.872	0.878	0.915	0.912	0.925	0.924	0.928	0.951	0.950	0.959
400	0.839	0.844	0.905	0.902	0.911	0.947	0.950	0.970	0.968	0.974	0.975	0.976	0.986	0.984	0.987
500	0.922	0.927	0.961	0.957	0.962	0.981	0.982	0.991	0.990	0.992	0.992	0.992	0.996	0.996	0.997
1000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
2000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table B.9
Simulated Power for Alternative Distribution: Chi-Square (2)

	α = 0.01					α = 0.05					α = 0.10				
n	LF	D	CVM	S	PML	LF	D	CVM	S	PML	LF	D	CVM	S	PML
5	0.011	0.011	0.011	0.011	0.011	0.050	0.049	0.051	0.051	0.049	0.098	0.097	0.098	0.098	0.097
10	0.009	0.009	0.010	0.009	0.009	0.047	0.048	0.048	0.048	0.050	0.098	0.098	0.097	0.097	0.101
15	0.011	0.011	0.009	0.009	0.011	0.050	0.049	0.049	0.048	0.048	0.098	0.098	0.097	0.096	0.096
20	0.009	0.009	0.009	0.009	0.009	0.050	0.048	0.049	0.050	0.048	0.099	0.100	0.099	0.097	0.099
25	0.010	0.010	0.009	0.009	0.010	0.052	0.050	0.050	0.050	0.049	0.101	0.100	0.101	0.100	0.100
30	0.009	0.009	0.009	0.009	0.009	0.051	0.050	0.051	0.051	0.049	0.099	0.101	0.101	0.101	0.101
40	0.010	0.010	0.010	0.010	0.010	0.051	0.050	0.049	0.050	0.050	0.099	0.100	0.099	0.100	0.099
50	0.010	0.010	0.010	0.011	0.010	0.050	0.050	0.049	0.049	0.051	0.100	0.099	0.099	0.099	0.098
60	0.010	0.010	0.010	0.010	0.011	0.051	0.051	0.052	0.052	0.052	0.103	0.103	0.103	0.102	0.102
70	0.010	0.010	0.011	0.010	0.010	0.052	0.052	0.051	0.050	0.052	0.100	0.102	0.101	0.100	0.101
80	0.011	0.011	0.010	0.010	0.010	0.050	0.050	0.050	0.051	0.052	0.100	0.102	0.100	0.101	0.101
90	0.011	0.011	0.010	0.011	0.011	0.051	0.050	0.051	0.052	0.051	0.101	0.101	0.100	0.099	0.102
100	0.010	0.011	0.011	0.011	0.010	0.050	0.051	0.050	0.051	0.051	0.099	0.100	0.102	0.103	0.102
200	0.010	0.010	0.011	0.010	0.010	0.052	0.052	0.051	0.051	0.051	0.101	0.101	0.101	0.101	0.100
300	0.010	0.010	0.010	0.010	0.010	0.052	0.052	0.051	0.050	0.052	0.098	0.099	0.100	0.102	0.101
400	0.010	0.010	0.010	0.010	0.010	0.051	0.051	0.051	0.049	0.049	0.102	0.102	0.100	0.099	0.100
500	0.010	0.010	0.010	0.010	0.010	0.052	0.052	0.051	0.049	0.050	0.101	0.101	0.101	0.102	0.102
1000	0.011	0.011	0.011	0.010	0.011	0.053	0.053	0.052	0.051	0.051	0.102	0.101	0.102	0.100	0.101
2000	0.010	0.010	0.010	0.010	0.010	0.050	0.050	0.049	0.049	0.049	0.101	0.101	0.100	0.100	0.100

Table B.10
Simulated Power for Alternative Distribution: t (5)

	α = 0.01					α = 0.05					α = 0.10				
n	LF	D	CVM	S	PML	LF	D	CVM	S	PML	LF	D	CVM	S	PML
5	0.887	0.874	0.884	0.874	0.869	0.903	0.895	0.904	0.897	0.900	0.915	0.909	0.916	0.912	0.915
10	0.993	0.992	0.991	0.988	0.987	0.995	0.994	0.994	0.992	0.992	0.996	0.995	0.995	0.994	0.994
15	1.000	1.000	0.999	0.999	0.999	1.000	1.000	1.000	0.999	0.999	1.000	1.000	1.000	1.000	1.000
20	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
25	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
30	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
40	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
50	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
60	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
70	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
80	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
90	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
100	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
200	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
300	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
400	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
1000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
2000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table B.11
Simulated Power for Alternative Distribution: log-normal (0, 1)

	α = 0.01					α = 0.05					α = 0.10				
n	LF	D	CVM	S	PML	LF	D	CVM	S	PML	LF	D	CVM	S	PML
5	0.019	0.020	0.019	0.018	0.018	0.070	0.073	0.073	0.072	0.058	0.131	0.132	0.134	0.134	0.106
10	0.031	0.035	0.034	0.033	0.033	0.097	0.104	0.102	0.101	0.096	0.169	0.175	0.177	0.177	0.163
15	0.042	0.048	0.046	0.043	0.049	0.119	0.128	0.129	0.129	0.127	0.197	0.205	0.210	0.212	0.204
20	0.049	0.053	0.053	0.051	0.056	0.137	0.144	0.148	0.151	0.152	0.221	0.231	0.240	0.244	0.241
25	0.053	0.061	0.062	0.062	0.067	0.156	0.164	0.174	0.178	0.179	0.249	0.258	0.275	0.280	0.279
30	0.060	0.066	0.070	0.069	0.075	0.171	0.179	0.195	0.203	0.202	0.267	0.281	0.305	0.314	0.314
40	0.080	0.087	0.093	0.095	0.100	0.208	0.216	0.243	0.253	0.258	0.322	0.334	0.369	0.382	0.385
50	0.096	0.105	0.117	0.125	0.131	0.245	0.254	0.292	0.307	0.313	0.375	0.384	0.428	0.443	0.448
60	0.116	0.125	0.145	0.154	0.161	0.292	0.300	0.352	0.370	0.380	0.428	0.439	0.503	0.518	0.525
70	0.137	0.143	0.176	0.186	0.195	0.330	0.342	0.410	0.426	0.438	0.480	0.489	0.564	0.580	0.587
80	0.162	0.170	0.203	0.217	0.223	0.372	0.381	0.458	0.481	0.495	0.528	0.536	0.622	0.637	0.648
90	0.183	0.194	0.238	0.261	0.274	0.413	0.421	0.515	0.538	0.547	0.579	0.582	0.673	0.685	0.694
100	0.208	0.216	0.276	0.305	0.313	0.460	0.470	0.561	0.585	0.594	0.621	0.628	0.723	0.735	0.740
200	0.498	0.507	0.671	0.696	0.700	0.818	0.821	0.908	0.910	0.915	0.921	0.923	0.965	0.961	0.963
300	0.784	0.789	0.910	0.918	0.921	0.968	0.968	0.990	0.988	0.989	0.991	0.991	0.998	0.996	0.997
400	0.941	0.939	0.985	0.984	0.985	0.996	0.996	0.999	0.999	0.999	0.999	0.999	1.000	1.000	1.000
500	0.989	0.990	0.998	0.998	0.998	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
1000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
2000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000